ESTIMATING THE CAUSAL EFFECT OF SMOKING CESSATION IN THE PRESENCE OF CONFOUNDING FACTORS USING A RANK PRESERVING STRUCTURAL FAILURE TIME MODEL

STEVEN D. MARK
Epidemiologic Methods Section, National Cancer Institute, 6130 Executive Blvd., EPN/403, Rockville, MD 20892, U.S.A.

AND

JAMES M. ROBINS
Departments of Epidemiology and Biostatistics, Harvard School of Public Health, U.S.A.

SUMMARY
Estimating the causal effect of quitting smoking on time to death or first myocardial infarction requires that one control for the differences in risk factors between individuals who elect to quit at each time t versus those who elect to continue smoking at time t. In this paper we examine the limitations of standard time varying Cox proportional hazards models to yield tests and estimates of this effect. Implementing the method of G-estimation proposed by Robins, we perform an observational analysis of data from the Multiple Risk Factor Intervention Trial (MRFIT) and estimate the causal effect of cigarette cessation while controlling for such time varying confounders as angina. We reject the null hypothesis of no effect of quitting on time to failure, and estimate that by quitting smoking, an individual increases by 50 per cent his time to death or first myocardial infarction (MI).

1. INTRODUCTION

Though most health professionals would not hesitate to advise their patients who smoke cigarettes to quit, the only large randomized trial1 that looked exclusively at the effect of quitting cigarette smoking on mortality and morbidity was inconclusive. This study, as well as other randomized trials in which cigarette cessation was one of several interventions,2-4 was hampered by poor compliance in the group assigned to quit. Such poor compliance drastically reduces the power of the usual intent to treat logrank test. The empirical evidence, therefore, that cigarette cessation is efficacious, derives from the results of observational analyses,5-10 these generally yield a relative risk in quitters to smokers of approximately 0.5, regardless of whether the endpoint is coronary heart disease (CHD) deaths, CHD morbidity and mortality, or all deaths and first myocardial infarction. Estimates from these studies have been utilized to make recommendations regarding the economic consequences and cost-effectiveness of smoking cessation techniques, and to project the impact of such programme on the occurrence of coronary artery disease in the United States.11,12

Some of the strongest evidence for the effectiveness of quitting smoking comes from observational analyses of data from the Multiple Risk Factor Intervention Trial (MRFIT). This study,
described in more detail below, possesses several unique attributes: smoking status was assessed annually by interview and biochemical measurements; information on potential confounders was gathered annually; endpoint assessment was comprehensive and performed without knowledge of an individual's smoking status. One recent MRFIT analysis compared subsequent CHD deaths among people who had quit smoking for all of their first three annual visits, with those who had not. Using a non-time-varying Cox proportional hazards model (CPH) to measure the exposure effect while controlling for baseline covariates, these investigators found the relative risk of dying of CHD was significantly lower amongst the quitters (relative risk = 0.35).

There have, however, been objections raised to studies that control only for baseline risk factors. It is known that persons who electively decide to quit smoking differ on risk factors for CHD from those who continue to smoke. If we account only for baseline risk factors, but there are time varying factors that predict both risk and smoking status, the above analysis will not be comparing comparable groups, and may yield biased estimates of the true effect of quitting. These limitations are frequently acknowledged by the authors of these observational analyses who offer such cautionary statements as 'bias may be inherent . . . if quitters are in some way are psychologically or physiologically different from those who continue to smoke.'

One response to this problem might be to use CPH models with time varying covariates. An earlier MRFIT analysis using CPH models that controlled for the time varying covariates diastolic blood pressure and serum cholesterol, as well as for baseline covariates, found that persons who had quit at time t had a relative risk of 0.54 compared to individuals who smoked at time t. However, what is not generally realized, is that these time varying CPH models may also fail to yield valid tests of the null hypothesis of no causal effect of quitting.

In this paper we use a method for analyzing observational data proposed by Robins and estimate the causal effect of quitting smoking while controlling for time varying confounders. In Sections 2 and 3 we describe the MRFIT study and our notation for the observable data. In Section 4 we introduce some latent failure time variables and formally state the assumption which, if true, allows us to draw causal inferences form observational data. In Sections 5 and 6 and we define the causal null hypothesis, and describe the conditions that must be met for tests of the null hypothesis of the standard CPH models to be valid tests of the causal null hypothesis. We demonstrate in Section 7 that these conditions are not met in the MRFIT data. In particular, we show that when compared to individuals without angina (symptomatic cardiac chest pain) by time t, individuals with angina have higher subsequent morbidity and mortality; are more likely to quit smoking at t; and are more likely to have been smokers prior to t. In Section 8 we introduce the rank preserving structural failure time model (RPSFTM). In Section 9 we consider how to obtain estimates of the parameter of the RPSFTM we propose. The MRFIT data are analyzed in Section 10. We contrast the results of our approach with that of a standard CPH model. Section 11 discusses how our estimating procedure handles censoring by end of follow-up. In Sections 12 and 13 we examine the robustness of our methods to data and model assumptions. We conclude with a discussion.

2. MRFIT

The design, conduct and analysis of MRFIT have been elaborately described. Briefly, MRFIT was a multicenter primary prevention trial designed to test the effect of a multifactorial intervention on coronary heart disease morbidity and mortality. The participants were 12,866 men aged 35-57 who, though free of overt cardiac disease, were at increased risk of coronary heart disease. The men were randomly assigned to a special intervention group (SI) or a control group called the usual care group (UC). The SI group received a stepped care protocol for the
Table I. Baseline and time varying covariates

<table>
<thead>
<tr>
<th>Indicator variables</th>
<th>Continuous variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>abnormal resting EKG; abnormal exercise EKG; randomization group; on medicine for HTN at $k = 0$</td>
<td>cholesterol; diastolic pressure; age; number of cigarettes smoked at $k = 0$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Indicator variables</th>
<th>Continuous variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>angina in interval $(k - 1, k)$; on medicine for HTN at $k$</td>
<td>cholesterol, diastolic pressure</td>
</tr>
</tbody>
</table>

Treatment of hypertension (HTN), participated in smoking intervention programmes, and received dietary counselling aimed at lowering serum cholesterol. Men in the UC group were referred to their community physicians for treatment. Baseline data and data from annual visits collected from both groups included repeat medical histories and examination, information on past smoking, information on current smoking, baseline exercise electrocardiograms (EKG's), annual rest EKG's, and annual serum cholesterol. Smoking status was ascertained by interview and by measurement of serum thiocyanate (SCN), a biochemical marker of current cigarette smoking. For the purpose of our analysis, a man is classified as a quitter at visit $k$ if he is a self-declared non-smoker and his SCN is $< 100$ μg. The primary endpoints of MRFIT were all deaths, all CHD deaths, and CHD mortality (combined CHD deaths and MI).

We are restricting our analysis to the 7686 men who were smokers at time 0 and who had complete baseline data including an exercise EKG. We refer to the minimum of time to first MI or death as individual's failure time. By end of follow-up a total of 740 failures had occurred. All the surviving men had been seen for at least 6 annual visits; approximately 40 per cent had been followed to their seventh annual visit or slightly beyond.

3. THE OBSERVABLE DATA

We use the continuous random variable $T_i$ to designate the failure time for the $i$th study subject with time as time since randomization. Data on time varying exposure (smoking status) and covariate status were gathered at each of $k$ yearly visits (0, 1, 2, $\ldots$, 7). We assume an individual's smoking status at visit $k$ remains unchanged for the interval $(k, k + 1]$, and denote by $Q_{i,k}$ subject $i$'s exposure status at $k$: $Q_{i,k} = 1$ if $i$ has quit at $k$; 0 otherwise. For example, if at visit 3 an individual's questionnaire and SCN level confirm he is currently a non-smoker, we assign him $Q_{i,3} = 1$ and assume he is a non-smoker from years [3, 4]. We let $L_{i,k}$ denote the vector consisting of all time dependent covariate measurements made on subject $i$ at visit $k$. $L_{i,k}$ contains data on baseline covariates. The time varying and baseline covariates that we use appear in Table I. If at visit 3 a man is found to have had angina between his second and third visit, his indicator variable for angina equals 1 at time 3. In this manner $L_{i,k}$ contains information that occurs prior to assessment of smoking status. We discuss the sensitivity of our analysis to this temporal ordering of smoking status and covariates in Section 12. For any time dependent variable $Z_{i,k}$ we use overbars to denote the history of the time dependent process. For example $\bar{Q}_{i,k} = (Q_{i,0}, \ldots, Q_{i,k})$. We use $\bar{Q}(t), \bar{L}(t)$ when we refer to the smoking or covariate history at some time $t$ between yearly visits, and $\bar{Q}(t), \bar{L}(t)$ when we refer to the history of the time dependent process up to some time $t$. In the absence of censoring, the observable data consist of
independent, identically distributed random variables of the form \((T_i, \bar{Q}_i(T_i), \bar{L}_i(T_i))\). We designate the number of the last visit before failure for person \(i\) as \(\text{int}(T_i)\). Though all the actual analyses we present account for the end of follow-up censoring that occurs in MRFIT, for expository reasons, we postpone our discussion of censoring until Section 11. Until then we present our model as if the data were uncensored.

4. THE LATENT DATA

To define explicitly the causal hypothesis we wish to test, and the assumptions about the observable data that we require, we assume the existence of latent failure time variables. Following Rubin’s \(^{23}\) and Holland’s \(^{26}\) formulation in point exposure studies and Robins’s \(^{27-25}\) in time-varying exposure studies, we assume that for each individual there exists a set of failure times \(\{U_{i,g=h}\}\), where \(g = h\) indicates that individual \(i\) followed the treatment history specified by \(h\) until his failure. \(h\) might be the treatment history never quit smoking; the treatment history always quit smoking; the treatment history smoke every other year. For each individual we observe only \(T_i\), which in terms of our latent random variables, we can write as \(T_i = U_{i,g=\delta_{T_i,T_c}}\).

In particular, we have interest in the latent failure times if individual \(i\) follows his observed smoking history up to visit \(k\), and then becomes a smoker \((Q_{ikn} = 0\) for \(m > k\)) until his failure. We designate these latent failure times as \(U_{ik}\), so that \(U_{i0}\) is the time to failure if individual \(i\) always smokes; \(U_{ik1}\) is the time to failure if individual \(i\) follows his actual smoking history up to visit 1 and then becomes a smoker again until his first MI or death, etc. We refer to these \(U_{ik}\) as baseline risk at \(k\) since they are the time to failure if individual \(i\) reverts to his ‘baseline habit of smoking’ from \(k\) onward. By convention, \(U_{i,k}\) equals \(T_i\) for subjects whose observed failure time is less than \(k\).

Our fundamental assumption is that for any history \(h\)

\[
U_{i,g=a} \perp Q_{i,k} | \bar{Q}_{i,k-1}, \bar{L}_{i,k}
\]

(1)

where \(A \perp B | C\) means \(A\) is independent of \(B\) given \(C\). One would believe (1) to be true provided he/she believed that for individuals with identical past smoking histories up to time \(k\) \((\bar{Q}_{i,k-1})\), and identical values of some known set of baseline and time dependent covariates through time \(k\) \((\bar{L}_{i,k})\), the decision to quit at time \(k\) \((Q_{ik})\) did not depend on the subject’s latest outcome \((U_{i,g=\delta k})\) regardless of his smoking history subsequent to \(k\). One could have made (1) true in the MRFIT data by adhering to the following study protocol:

1. At each time \(k\), stratify individuals on their past smoking \((\bar{Q}_{i,k-1})\) and covariate \((\bar{L}_{i,k})\) histories.
2. Within each stratum randomly assign, for instance by a flip of a coin, every individual to one of two groups. Randomization guarantees that the groups are comparable in the sense that if one could force every individual in both of the two groups to adhere to the same treatment regime \(h\), the distributions of failure times, \(U_{i,g=a}\), would (in expectation) be identical in the two groups.
3. In each stratum force one of the two groups in step 2 to quit smoking, and one to continue smoking, for the next time interval.

Since physical randomization in step 2 guarantees (1), even if the probability that the coin lands heads up at time \(k\) depends on past treatment and covariate history, most people accept that one can draw valid causal inferences from randomized studies. As with a randomized study, (1), if true, allows one to draw causal inference from observational data. In point exposure studies Rosenbaum and Rubin have called this the assumption of strong treatment ignorability.\(^{23}\) We refer to
this analogous assumption in longitudinal studies as the condition of maintained comparability (Robins\textsuperscript{17,23,24} has previously referred to (1) as the assumption of no unmeasured confounders, and as the assumption of randomization with respect to survival\textsuperscript{18,19,21,25}) In the hypothetical study where we randomize at each time $k$, we can make (1) true for any set of covariates $L_{i,k}$ simply by stratifying on those particular covariates in step 1. In contrast, in an observational study, (1) will be true provided we have contained in $L_{i,k}$ data on all those covariates that are risk factors for the latest outcomes and that an individual uses to decide whether or not to smoke at $k$.

For instance, if we had no information on the development of angina, but believed that conditional on the other available information, angina was both a predictor of outcome and a predictor of whether one smoked at $k$, then we would expect (1) to be false. In practice, one primary goal of investigators who conduct an observational study is to collect data on a sufficient number of covariates to assure that the condition of maintained comparability is at least approximately true.

5. THE CAUSAL NULL AND THE BIAS OF STANDARD METHODS

We define the null hypothesis of no causal effect of treatment on survival time as

$$T_i = U_i = U_i$$ for all $i$, and all $h = 1, 2$. \hspace{1cm} (2)

If this causal null hypothesis (2) is true, then individual $i$'s observed failure time would be unchanged regardless of what smoking history he were to follow. In conjunction with our assumption of maintained comparability (1), the causal null hypothesis implies that conditional on past smoking and covariate history, the observable failure times, $T_i$, do not depend on current smoking status $T_i \cup Q_{i,1}, Q_{i,1-1}, L_{i,1}$. \hspace{1cm} (3)

Thus if (1) is true, an $x$-level test of the observable causal null hypothesis (3) is an $x$-level test of the causal null hypothesis (2). One of the principal aims of our paper is to show (Section 9) that we can use a rank preserving structural failure time model to test (3).

Rather than testing (3), the standard approach in observational studies that assess the effect of smoking history on survival, is to test whether, conditional on past covariate history, the hazard for failure at time $t$ depends on smoking history up to time $t$.

For instance, one standard approach is to ignore the time varying covariates other than the smoking history and test the hypothesis

$$\lambda(t | \tilde{Q}(t), L_i(0)) = \hat{\lambda}(t | L_i(0))$$ \hspace{1cm} (4)

where $\hat{\lambda}(t | \cdot)$ is the hazard at $t$ of $T_i$ conditional on $\cdot$. One might test equation (4) by testing $\theta = 0$ in a CPH model such as

$$\lambda(t | \tilde{Q}(t), L_i(0)) = \lambda_0(t) \exp[\theta W_i(t) + \sum_{\gamma=1}^{\mu} \gamma Z_i(t)]$$ \hspace{1cm} (5)

where $W_i(t)$ is some real valued function of smoking history such as cumulative exposure ($W_i(t) = \int_0^t Q_i(u)du$ or current exposure ($W_i(t) = Q_i(t)$); $Z_i = Z_i, \ldots, Z_i$) are real valued functions of the baseline covariates in Table 1; and $\theta$ and $\gamma$ are unknown parameters.

A second standard approach is to utilize the time dependent covariates up to time $t$ ($\tilde{E}_i(t)$) and test the hypothesis

$$\lambda(t | \tilde{Q}(t), L_i(t)) = \hat{\lambda}(t | L_i(t))$$ \hspace{1cm} (6)

by testing $\theta = 0$ in a correctly specified model such as

$$\lambda(t | \tilde{Q}(t), L_i(t)) = \lambda_0(t) \exp[\theta W_i(t) + \sum_{\gamma=1}^{\mu} \gamma Z_i(t)]$$ \hspace{1cm} (7)
where \([Z_{t-n}(t), \ldots, Z_n(t)]\) are real valued functions of the baseline and time varying covariates listed in Table I. If we let \(W(t) = Q(t)\), and let \(L_n(t)\) contain the three baseline covariates age, serum cholesterol, and diastolic blood pressure, and the two time varying covariates serum cholesterol and diastolic blood pressure, then we have the time varying model used for estimating the effect of quitting cited in Reference 16 in the Introduction.

Given maintained comparability (1) is true, the test of \(\theta = 0\) in model (5) is a valid \(\alpha\)-level test of the causal null hypothesis (2) only if (2) implies (4). Similarly the test of \(\theta = 0\) in model (7) can be a valid \(\alpha\)-level test of (2) only if (2) implies (6). However, (2) implies neither (4) nor (6). The following hypothetical example provides some intuition for this mathematical fact.

**Example 1:** Suppose that we have covariate information only on smoking history and angina and that equation (1) (maintained comparability) is true conditional on these two covariates. Further suppose that although the causal null hypothesis is true, the following relationships exist between angina and failure time, and angina and smoking:

1. Conditional on past smoking history, the probability of failure at \(t\) is greater in individuals who have angina by time \(t\). (Angina indicates clinically significant CHD.)
2. Conditional on past smoking history, the probability of quitting smoking at \(t\) is increased in individuals who have angina by time \(t\). (Individuals who develop angina become concerned about their health and will more likely quit smoking.)
3. Smokers who develop angina have more severe CHD than quitters who develop angina. (Smoking has no direct causal effect on CHD, angina or death. Smokers, however, are more sedentary than non-smokers so that angina, an exertionally induced symptom, does not develop until the smoker has reached a more advanced stage of CHD.)

Owing to condition 2 above, the group of non-smokers at \(t\) contains a larger proportion of people with angina than the group of current smokers. Since angina predicts failure (condition 1 above), if one ignored the time dependent covariate angina and used model (5), one would find \(\theta > 0\) and might wrongly conclude that quitting smoking increased the hazard of the time to first MI or death. Suppose one 'controlled' for the effect of angina by entering it as a covariate in a model such as (7)? Among subjects who develop angina at the same \(t\), condition 3 implies that smokers will have more severe coronary artery disease than quitters. Thus smokers are more likely to fail. Then, \(\theta < 0\) in model (7) and one might conclude wrongly that quitting cigarettes improved survival.

It is important to emphasize that the failure of the two CPH models to provide valid \(\alpha\)-level tests of the causal null hypothesis was not due to the misspecification of any of the hazards in models (5) or (7). Rather they failed because, as the above example shows, if there is an association of quitting with a risk factor such as angina, the left and right hand sides of both (4) and (6) may be unequal despite the causal null hypothesis being true. In Section 6 we provide a formal definition of time varying confounders and describe the associations that covariates such as angina must have with either smoking or outcome, if models (5) and (7) are to be useful for inference regarding the causal null hypothesis.

### 6. TIME VARYING CONFOUNDERS

We say that \(L\) history is not a time varying confounder in our data set if either of the following two conditions are true for all times \(t > 0\):

\[
\lambda(t)(\bar{Q}(t), L_n(0), E_n(t)) = \lambda(t)(\bar{Q}(t), L_0(0))
\]

\[
P[Q_{i,k} = 1 | Q_{i,k-1}, L_n(0), E_n, T_i > k] = P[Q_{i,k} = 1 | Q_{i,k-1}, L_0(0), T_i > k]
\]
where \( P[Q_{i,k} = 1 \mid T_i > k] \) is the probability that person \( i \) is a quitter at time \( k \) conditional on \( L_i(0) \) and being at risk.

Equation (8) is true if, conditional on past smoking history and the baseline covariates \( L_i(0) \), the time varying covariates \( L_i(t) \) do not predict survival. In Example 1 there were no baseline covariates and equation (8) was false since after matching on past smoking history, angina was a risk factor for death. Equation (9) is true if, given past smoking history and baseline covariates, time varying covariates do not predict future smoking. In Example 1 this also was false since conditional on past smoking history, angina history did predict smoking status at \( k \).

When either (8) or (9) is true we say there are no time varying confounders and then maintained comparability (1) and the causal null (2) imply (4), and we can test the causal null hypothesis (2) using the CPH model (5) (see Appendix I). In Section 7 we examine our data for time varying confounders and reject (8) and (9).

We say that smoking is not a predictor of subsequent \( L \) history if

\[
P[Q_{i,k} \mid L_{i,k-1}, L_{i,k-1}, T > k] = P[L_i(t) = L_i(k) \mid L_{i,k-1}, T > k]
\]

where \( P[Q_{i,k} = 1 \mid L_{i,k-1}, L_{i,k-1}, T > k] \) is the probability of \( L_i \) conditional on past smoking and \( L_i(k) \) and history, and being at risk at \( k \).

In Example 1 where the only time varying covariate is angina, (10) is not true since the probability of developing angina at \( k \) is lower in smokers than in non-smokers.

A simple calculation shows that when (10) is true, (i) \( \theta = 0 \) in CPH model (6) implies the observable causal null hypothesis (3) is true, although (ii) if \( \theta \neq 0 \), we cannot conclude that (3) is false. In Section 7 we reject (10) and show that individuals who smoke are more likely to develop angina than those who quit.

7. TESTING FOR TIME VARYING CONFOUNDERS

In this section we test whether the time varying covariates in Table I (\( L_i \) history) are time varying confounders by modelling (8) and (9). First we model the probability of failure using a discrete failure time version of the Cox proportional hazards model. That is, we assume that the probability of failure in the time interval \([k, k + 1]\) conditional on past smoking and covariate history and having survived to time \( k \) is of the form

\[
\text{logit} \ P[T_i \leq k + 1 \mid T_i > k, \bar{Q}_{i,k}, \bar{L}_{i,k}) = \alpha_i V_i + \beta W_i(k) + \gamma Z_i(k) + \delta D_i
\]

where \( V_i \) is a vector of indicator variables for visit number and \( W_i(k) \), \( Z_i(k) \), \( D_i \) are real (possibly vector) valued functions of \( \bar{Q}_{i,k}, \bar{L}_{i,k} > 0 \) and \( L_{i,0} \), respectively, and \( \alpha_i, \beta, \gamma \) and \( \delta \) are unknown parameter vectors. If \( \gamma = 0 \) then (8) is true and we have no time varying confounders. Since the probability of failing in any year is small, inferences based on the discrete time Cox model are essentially the same as those based on a continuous time Cox model such as (7).

We modelled (11) as a function of current smoking status at \( k \), and the variables listed in Table I. Table II contains the hazard ratios and Wald \( p \)-values for the most important baseline and time varying predictors of failure. As one would anticipate, increasing age, cholesterol and number of baseline cigarettes smoked all increase the probability of an MI or death. Most importantly, we find that several time varying covariates, angina in particular, predict failure. Thus we would not conclude that (8) is true since, even with control of smoking status and other baseline covariates, persons with angina at \( k \) have nearly twice the risk of failing in the ensuing interval than persons without angina at \( k \).

We also used logistic models to test (9). That is, we used models of the form

\[
\text{logit} \ P[Q_{i,k} = 1 \mid \bar{Q}_{i,k-1}, \bar{L}_{i,k}, T_i > k] = \alpha_i V_i + \beta W_i(k) + \gamma Z_i(k) + \delta D_i
\]
If $\gamma = 0$ then time varying $L$ history does not predict smoking, (9) is true, and we have no time varying confounders.

Rather than run a single logistic regression such as (12), we divided the data by visit numbers, matched on prior smoking history at $k-1$, entered as covariates the smoking history prior to $K-1$, and ran five separate logistic regressions on the subsets defined in column 1 of Table III. Separating the data in this fashion allowed us to control for the relationship of past smoking to current smoking with fewer modelling assumptions.
Table III records the odds ratios for the most important predictors in the history of quitting smoking at \( k \). We find that individuals with angina in the period \( k - 1 \) to \( k \) are more likely to quit smoking at \( k \) than are angina-free individuals. We therefore reject (9).

Since angina is a significant predictor of both survival and smoking history, we cannot obtain valid \( z \)-level tests of the causal null hypothesis (2) with use of the CPH model (5) since it fails to account for time varying covariates. In Section 6 we noted that if (10) were true and past smoking does not influence current angina while controlling for past angina, then CPH models that do include angina (7), or the discrete model (11) would at least allow one to accept the observable causal null hypothesis (3) if one accepts the hypothesis that \( \theta = 0 \) in CPH model (7). To test (10) we again use logistic regression. In this regression the dependent variable is an indicator variable that equals 1 if an individual develops angina in the interval \( (k - 1, k) \). The independent variable is quit status at \( k - 1 \). We find that people who quit at \( k - 1 \) are less likely to develop angina at \( k \) than are people who smoke at \( k - 1 \) (odds ratio = 0.83, \( p = 0.08 \)). This is consistent with the biological fact that smoking decreases blood oxygen, increases blood carbon monoxide, and thus may be expected to precipitate angina.

8. RANK PRESERVING STRUCTURAL FAILURE TIME MODEL

We have shown in the MRFIT data that the nature of the associations between angina and smoking, and angina and outcome, precludes construction of valid \( z \)-level tests of the causal null hypothesis (2) by testing whether the hazard at \( t \) is a function of past treatment history (4) or (6).

In this section we introduce a model for treatment effect proposed by Robins, the rank preserving structural failure time model (RPSFTM). In Section 9 we show that the test we propose of the null hypothesis of the RPSFTM is a valid \( z \)-level test of the observable causal null hypothesis (3), and hence a valid test of the causal null hypothesis (2).

In particular, our RPSFTM assumes that the latent variables baseline failure time at \( k(U_{i,k}) \) relate to the observable data \( \{T_i, \hat{Q}(T_i)\} \) by the relationship

\[
U_{i,k} = k + \int_k^{T_i} \exp \{\psi_0 \hat{Q}(u)\} \, du \quad \text{if } T_i > k
\]

\[
U_{i,k} = T_i \quad \text{if } T_i \leq k
\]

where \( \psi_0 \in \mathbb{R}^d \) is an unknown parameter. (Equations (13) are essentially the strong version of the accelerated failure time model of Cox and Oakes.) If \( \psi_0 = -0.1 \) and the observed data for individual \( i \) were \( (T_i = 2, \hat{Q}_{i,2} = \{1,0,1\}) \) then

\[
U_{i,2} = 2 + (2 \times 2 - 2)\exp(1 \times -0.1) = 2.18
\]

\[
U_{i,1} = 1 + (2 \times 2 - 2)\exp(1 \times -0.1) + (2 - 1)\exp(0 \times 0.1) = 2.18
\]

\[
U_{i,0} = 0 + (2 \times 2 - 2)\exp(1 \times -0.1) + (2 - 1)\exp(0 \times 0.1) + (1 - 0)\exp(1 \times -0.1) = 2.09
\]

When \( \psi_0 = 0 \) the function returns the identity:

\[
U_{i,k} = T_i \quad \text{for all } i, \text{ and all } k.
\]

Thus the causal null hypothesis (2) implies \( \psi_0 = 0 \) in (13). Therefore, given (2), model (13) is specified correctly and a test of \( \psi_0 = 0 \) is a valid \( z \)-level test of the causal null.

To understand the implications of the model when \( \psi_0 \neq 0 \), we consider another latent failure time, \( V_i \), where \( V_i \) is a person’s failure time if he permanently quits smoking at time 0 (that is,
\( V_i = U_{i,k-1} \), where \( h \) corresponds to the smoking history always quit. If, as suggested by model (13), we assume \( V_i \) relates to \( U_{i,0} \) by

\[
U_{i,0} = \exp[\psi_0 \times 1] \int_0^\infty \exp(\psi_0) \, du = \exp(\psi_0) V_i
\]

\[
\exp(-\psi_0) U_{i,0} = V_i
\]

then, \( \psi_0 < 0 \) implies that sustained quitting extends life by a factor \( \exp(-\psi_0) \), \( \psi_0 > 0 \) implies that sustained quitting decreases life by a factor \( \exp(-\psi_0) \). The expansion factor \( \exp(-\psi_0) \) relates to a common parameter of public health interest. Specifically

\[
\frac{V_i - U_{i,0}}{U_{i,0}} = \exp(-\psi_0) - 1
\]

is the fractional increase in survival for individual \( i \) if he permanently quits smoking.

We note that, in direct contrast to CPH models such as (5) or (7), the RPSFTM does not include the time independent or time varying confounders (L. history). One feature that distinguishes our estimating procedure from estimation using CPH models is that we do not utilize the model for treatment effect to control for confounding. In the next section we demonstrate how comparability conditions (that is, confounders) do enter the estimating procedure.

9. ESTIMATION OF \( \psi_0 \)

In this section we discuss how to construct asymptotic \( z \)-level tests of \( \psi = \psi_0 \). The tests we propose are semi-parametric in the sense that one need specify only a portion of the full likelihood of the observable data \( (T_i, \hat{Q}_i(T_i), L_i(T_i)) \). In particular, they require correctly specified models for the probability that an individual quits smoking at \( k \) given his past smoking and covariate history:

\[
P[Q_{i,k} \mid \hat{L}_{i,k}, \hat{Q}_{i,k-1}] \tag{17}
\]

In the setting of time independent treatments, Rosenbaum and Rubin,\(^{13}\) Rosenbaum\(^{40}\) and Robins \( et \ al.\) have previously considered estimating treatment effects in observational studies by modelling the probability of treatment given covariates. In References 23 and 24 Robins shows that under maintained comparability (1) and the RPSFTM (13), the likelihood factors into the product of (17) and models for the two probability densities \( P[U_{i,0}] \) and \( P[L_{i,k} \mid L_{i,k-1}, \hat{Q}_{i,k-1}, U_{i,0}] \). Since the \( U_{i,0} \) are not directly observed, one would expect it difficult, if not impossible, to specify a parametric model for \( P[L_{i,k} \mid L_{i,k-1}, \hat{Q}_{i,k-1}, U_{i,0}] \) which an investigator believes to approximate the truth. In contrast, one might be more secure in specifying parametric models for (17). Such models require only an understanding of the factors that influence an individual’s decision to quit smoking at \( k \) given his past smoking and covariate history. Thus, for reasons of robustness we prefer this semi-parametric method that allows the densities for \( P[U_{i,0}] \) and \( P[L_{i,k} \mid L_{i,k-1}, \hat{Q}_{i,k-1}, U_{i,0}] \) to remain completely unrestricted.\(^{23,24}\)

In Section 7 we encountered models for (17) when we tested to see whether any time varying covariates predicted quitting. In this section, for ease of explication, we assume (17) is modelled using a single logistic model of the form

\[
P[Q_{i,k} \mid L_{i,k}, \hat{Q}_{i,k-1}] = \frac{\exp(z_0 W_{i,k})}{1 + \exp(z_0 W_{i,k})} \tag{18}
\]
Here $k \in \{1, 2, \ldots, \text{int}(T; \theta)\}$, $a_0 \in \mathbb{R}^p$ is an unknown parameter, and $W_{i,k}$ is a $p$-dimensional function of $(\vec{Z}_{i,k}, \hat{Q}_{i,k-1})$. For example, $W_{i,k}$ might be the vector consisting of smoking at $(k-1)$, smoking at $(k-2)$, average smoking prior to $(k-2)$, diastolic blood pressure at $k$, age, and indicator variables for each of the $k$ occasions, for the presence of angina at $k$, for randomization group, etc.

The key to the testing procedure is to note that the condition of maintained comparability (1) imposes a restriction on the joint distribution of the conditional probability of quitting at $k$ and the baseline risk at $k$: when (1) is true, the baseline risk at $k$, $U_{i,k}$, is conditionally independent of the decision to quit at $k$. We can write this as

$$P[Q_{i,k} | \vec{Z}_{i,k}, \hat{Q}_{i,k-1}] = P[Q_{i,k} | \vec{Z}_{i,k}, \hat{Q}_{i,k-1}, U_{i,k}].$$ (19)

Let $U_{i,k}(\psi)$ denote the right hand side of equation (13a) when $\psi_0$ is replaced by $\psi$. (Recall that $U_{i,k}(\psi_0)$ is the baseline risk at $k$, $U_{i,k}$.) Equation (19) implies that the true value of $\theta$ in the extended logistic model (20) equals 0:

$$P[Q_{i,k} | \vec{Z}_{i,k}, \hat{Q}_{i,k-1}, U_{i,k}(\psi_0)] = \frac{\exp(aW_{i,k} + \theta U_{i,k}(\psi_0))^{Q_{i,k}}}{1 + \exp(aW_{i,k} + \theta U_{i,k}(\psi_0))}.$$ (20)

For any hypothesized value $\psi$ we can calculate $U_{i,k}(\psi)$ from the data and model (13). Not knowing the true $\psi_0$, we test the hypothesis that a particular value of $\psi$ equals $\psi_0$ by seeing whether a score test accepts or rejects the hypothesis that the true value of $\theta$ is 0 in model (20).

Notice that under the null hypothesis of the RPSFTM ($\psi_0 = 0$), $U_{i,k}(\psi) = T_k$. Substituting $T_k$ for $U_{i,k}$ equation (19) becomes

$$P[Q_{i,k} | \vec{Z}_{i,k}, \hat{Q}_{i,k-1}] = P[Q_{i,k} | \vec{Z}_{i,k}, \hat{Q}_{i,k-1}, T_k].$$

Thus our test of the hypothesis $\psi_0 = 0$ is a test of the observable causal null hypothesis (3).

The score test, which we call the G-test statistic, $Z(\psi)$, consists of a statistic, $S(\hat{z}, \theta = 0, \psi)$, divided by a consistent estimate of its standard deviation, $\tilde{Q}^{1/2}(\psi)$

$$Z(\psi) = S(\hat{z}, \theta = 0, \psi)/\tilde{Q}^{1/2}(\psi).$$ (21)

The computational formula for $S(\hat{z}, \theta = 0, \psi)$, is

$$S(\hat{z}, \theta = 0, \psi) = \sum_{i=1}^{n} \sum_{k=0}^{\text{int}(T; \theta)} U_{i,k}(\psi) \{Q_{i,k} - P_{i,k}(\hat{z}, \theta = 0)\}$$ (22)

where the inner sum is over visits and the outer sum is over individuals. Here $P_{i,k}(\hat{z}, \theta = 0)$ is the predicted value of $P[Q_{i,k} = 1 | \vec{Z}_{i,k}, \hat{Q}_{i,k-1}, U_{i,k}(\psi)]$, and $\hat{z}$ is the restricted MLE of $z_0$ when $\theta$ is set to 0 in model (20). Note that model (20) with $\theta$ set to 0 is just model (18). Hence we can obtain these predicted values from a standard logistic regression package by arranging our data as if the visits from each individual were independent (that is, we have a total of $N = \sum_{i=1}^{n} \text{int}(T; \theta)$ observations) and by outputting the predicted values from a MLE fit of model (18). Under the hypothesis $\psi_0 = \psi$, $S(\hat{z}, \theta = 0, \psi)$ has asymptotic distribution $N(0, \Omega(\psi))$. In Appendix II we provide a formula for a consistent estimate, $\tilde{Q}(\psi)$, of $\Omega(\psi)$ and demonstrate the relationship of our G-test to the usual score test for logistic regression.

The set of $\psi$ for which $|Z(\psi)|$ is less than the upper $a/2$ percentile of a $N(0, 1)$ random variable is an asymptotic $(1 - a)$ confidence interval for $\psi_0$.

Our point estimate is the $\hat{\psi}$ solving $Z(\hat{\psi}) = 0$; $n^{1/2}(\hat{\psi} - \psi)$ is asymptotically a mean 0 normal random variable. Following References 23 and 24, we will refer to $\psi$ as a G-estimate. We can
obtain a consistent estimate of the asymptotic variance of \( \hat{\psi} \) by squaring the inverse of the numerical estimate of the slope of our statistic \( Z(\psi) \) evaluated at \( \hat{\psi} \).

10. DATA ANALYSIS

In this section we present the results of three different analyses of the MRFIT data. The analyses differ from one another by the covariates entered into the logistic regression models for predicting smoking at \( k \) (models such as (18)). For the past smoking only analysis we enter only smoking history through \( k - 1 \) as predictors of smoking at \( k \). For the baseline analysis we enter past smoking history and all the baseline covariates in Table I. For the time varying analysis we enter past smoking history and all the baseline and time varying covariates in Table I. Assuming that we have specified correctly the KPSFTM (13), and that for maintained comparability (1) to be true we need to condition on all the baseline and time varying covariates recorded in \( L \) history, it is the time varying analyses that will give an unconfounded (that is, consistent) estimate of \( \psi_0 \).

The other two analyses permit us to judge the degree of confounding that occurs when we disregard important covariates such as the development of angina. For reasons explained in the section on censoring, all logistic regressions in each analyses also included the baseline covariate potential censoring time which we define for each individual as the time between his randomization date and the common end of follow-up date of the study. We denote this observable covariate as \( C_i \).

To estimate \( \psi_0 \) for any given analysis we implement the following procedure. First, adopting the modelling strategy of Section 7, we divide our data into the five distinct subsets defined in column 1 of Table III, and fit in each subset separate logistic regression models of the form (18) for the probability of quitting at \( k \). From each model we calculate the predicted probability that individual \( i \) quits at \( k \), \( \hat{\psi}_i(\hat{\psi}, \hat{\theta} = 0) \). We need only calculate these predicted values once for each analysis; they do not change as a function of \( \hat{\psi} \). Next we choose a test value of \( \psi \) and calculate the \( U_i(\psi) \) using the right hand side of (13a) with \( \psi \) replacing \( \psi_0 \). For each of the five logistic models we calculate the numerator of our G-test and an estimate of its variance. Our overall G-test statistic (21) for a particular \( \psi \) is the sum of these five numerators divided by the square root of the sum of their estimated variances. Since all our participants were smokers at time 0, the numerator of our statistic (21) is modified to be a sum over visits \( k = 1 \) to \( int(T_i) \) rather than from \( k = 0 \) to \( int(T_i) \).

Figure 1 is a graph of the G-test statistic versus \( \psi \) for the time varying analysis. Since the G-test statistic, \( Z(\psi) \), equals 0 when \( \psi = -0.43 \), our point estimate of \( \psi_0 \) is \( \hat{\psi} = -0.43 \). The variance estimate, 0.011, is the square of the inverse of the slope at \( \hat{\psi} \) (the steeper the slope, the lower the variance), the 95 per cent confidence interval extends from \(-0.64\) to \(-0.22\) and excludes the null value of 0. Hence, if our assumptions are true we can reject the causal null hypothesis and conclude that quitting smoking has a beneficial effect on time to failure. Using (16) we estimate that an individual increases his time to failure by 54 per cent if he quits smoking permanently at time 0.

In Section 7 we had determined that some of the covariates were time varying confounders, and that an analysis adjusting for those confounders was not mathematically identical to one that ignored those confounders. Yet when we compare the results of the three analyses in Table IV, we find that the estimate of \( \psi_0 \) does not change with the model. The magnitude of confounding, however, depends not only upon the strength of a confounder’s independent association with exposure and outcome, but also upon its prevalence. Our analysis in Section 7 revealed that only angina had an independent effect on both exposure and outcome. The development of angina,
however, was rare, occurring in just 3 per cent of the visits. As a consequence, ignoring the time varying covariate angina has no detectable effect on our estimate of $\psi_0$.

If removing the time varying confounders from our model does not produce bias, we may use CPH models such as (5) to estimate the causal hazard ratio, $\lambda_0(t)/\lambda_{0,0}(t)$. $\lambda_0(t)/\lambda_{0,0}(t)$ is the ratio of the hazard if everyone quit smoking at time 0 (the hazard of the latent variable $V$) to the hazard if no one ever quit smoking (the hazard of the latent variable $U_{i,0}$). When the underlying distribution of $U_{i,0}$ is Weibull (that is, $P[U_{i,0} > t] = \exp(-\lambda t^k)$), there is a one-to-one correspondence between the $\psi_0$ of model (13) and the causal hazard ratio.\(^3\) The causal hazard ratio equals $\exp(k\psi_0)$. Figure 2 is a graph of the log of the Kaplan–Meier estimate of the $P(U_{i,0} > t)$ versus time (to generate the $U_{i,0}$ we assumed $\psi_0 = -0.43$). The linearity of the graph suggests that the $U_{i,0}$ arise from an exponential survival distribution: that is, a Weibull with $k = 1$. Using parametric likelihood methods for the Weibull model (see Appendix III), we formally tested the hypothesis that $k = 1$ (2-sided $p$-value = 0.78, point estimate of $k = 0.987$) and estimated the
hazard as $\hat{\lambda} = 0.12$. For $U_{t,0}$ that are exponentially distributed, $\psi_0 = -0.43$ in the RPSFTM corresponds to a causal hazard ratio of 0.651. Estimating the causal hazard ratio directly from the discrete time version of CPH model (5) (that is, model (11) with the time varying covariates excluded) we found $\hat{\lambda}_t(t)/\hat{\lambda}_{t,0}(t) = 0.647$. Thus in these data where one can ignore time varying confounders and one can use the CPH model (5) to estimate the causal hazard ratio, these two distinct methods of estimation yield nearly identical estimates of $\hat{\lambda}_t(t)/\hat{\lambda}_{t,0}(t)$ (Although we did not do so, it would have actually been more appropriate to determine that the distribution of $U_{t,0}$ was exponential within levels of time independent covariates $L_{t,0}$).

11. CENSORING

Only 740 of the 7686 individuals in our cohort experienced a failure. Therefore, our observable data, rather than of the form $(T_i, \tilde{Q}(T_i), \tilde{L}(T_i))$, are of the form $(X_i = \min(T_i, C_i), \tilde{Q}(X_i), \tilde{L}(X_i), C_i)$ where $C_i$ is defined as in the first paragraph of Section 10. It might be natural to replace the now unobservable $U_{i,0}(\psi_0)$ by a new random variable, $X_{i,0}^*(\psi_0)$ generated by substituting $X_i$ for $T_i$ in (13). Unfortunately, if $\psi_0 \neq 0$ then $X_{i,0}^*(\psi_0)$ is not independent of $Q_{i,k}$ given $(\tilde{Q}_{i,k-1}, \tilde{L}_{i,k})$ thus an alternative approach is necessary. The key to understanding the modification necessary for analysing censored data is to realize that our assumption of maintained comparability (1) implies that any function of $(U_{i,0}(\psi_0), \tilde{Q}_{i,k-1}, \tilde{L}_{i,k})$ is independent of $Q_{i,k}$ given $(\tilde{Q}_{i,k-1}, \tilde{L}_{i,k})$. (Recall that $C_i$ is a component of $L_{t,0}$ since $C_i$ is known at the start of the study.) Thus we define an observable random variable that is a function of $(U_{i,0}(\psi_0), \tilde{Q}_{i,k-1}, \tilde{L}_{i,k})$ and use it as a basis for inference concerning $\psi_0$. 
Specifically,

\[ X_{i,k}(\psi) = \min \{ U_{i,k}(\psi), C_{i,k}(\psi) \} \]

where

\[ C_{i,k}(\psi) = C_i - k \quad \text{if} \quad \psi \geq 0 \]

\[ C_{i,k}(\psi) = k + (C_i - k)\exp(\psi) \quad \text{if} \quad \psi < 0. \]

\(X_{i,k}(\psi)\) is a function of \(\psi\), \(C_i\), \(U_{i,k}(\psi)\) and \(k\). If an individual is censored \((C_i < T_i)\) then \(C_{i,k}(\psi) < U_{i,k}(\psi)\). Thus \(X_{i,k}(\psi)\) is observable for all \(\psi\). We let

\[ \Delta_{i,k}(\psi) = I[C_{i,k}(\psi) > U_{i,k}(\psi)] \]

where \(I[ \cdot ]\) is the indicator function. When \(\Delta_{i,k}(\psi) = 0\) we say individual \(i\) is \(\psi\)-censored at \(k\). Since

\[ \left[ \{ \Delta_{i,k}(\psi_0), X_{i,k}(\psi_0) \} \right] \left[ \bar{Q}_{i,k} | \bar{Q}_{i,k-1}, L_{i,k} \right] \]

we know that for any real valued function \(g[\Delta_{i,k}(\psi_0), X_{i,k}(\psi_0), \bar{Q}_{i,k-1}, L_{i,k}]\),

\[ g[\Delta_{i,k}(\psi_0), X_{i,k}(\psi_0), \bar{Q}_{i,k-1}, L_{i,k}] \mid \bar{Q}_{i,k} | \bar{Q}_{i,k-1}, L_{i,k}. \tag{23} \]

Equation (23) implies that the results of Section 9 concerning the estimation of \(\psi_0\) remain true when \(g[\Delta_{i,k}(\psi), X_{i,k}(\psi), \bar{Q}_{i,k-1}, L_{i,k}]\) replaces \(U_{i,k}(\psi)\) in (22). In Section 10, we generated our estimates, including Figure 1, using \(g[\Delta_{i,k}(\psi), X_{i,k}(\psi), \bar{Q}_{i,k-1}, L_{i,k}] = \Delta_{i,k}(\psi)X_{i,k}(\psi)\); that is, we used the \(U_{i,k}(\psi)\) for individuals not \(\psi\)-censored at \(k\), and set \(U_{i,k}(\psi) = 0\) for those \(\psi\)-censored at \(k\).

We could also form a point estimate and confidence interval for \(\psi_0\) using the \(C_{i,k}(\psi)\) for people \(\psi\)-censored at \(k\), and setting the others equal to \(0\); that is, using \(g[\Delta_{i,k}(\psi), X_{i,k}(\psi), \bar{Q}_{i,k-1}, L_{i,k}] = |1 - \Delta_{i,k}(\psi)|X_{i,k}(\psi)\). Figure 3 is a graph of the G-test statistic versus \(\psi\) in the time
Table V. Estimates of $\hat{\psi}_0$ using the $\psi$-censored

<table>
<thead>
<tr>
<th>Model</th>
<th>Point estimate and 95 per cent CI of $\hat{\psi}_0$</th>
<th>Variance estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Past smoking only</td>
<td>$-0.39 (-0.56, -0.21)$</td>
<td>0.008</td>
</tr>
<tr>
<td>Baseline</td>
<td>$-0.39 (-0.56, -0.21)$</td>
<td>0.008</td>
</tr>
<tr>
<td>Time varying</td>
<td>$-0.39 (-0.56, -0.21)$</td>
<td>0.008</td>
</tr>
</tbody>
</table>

varying analysis using this function. Though, as we should expect, in the neighbourhood of $\hat{\psi}$, the slope of Figure 3 differs in sign from that of Figure 1, a comparison of the results in Table V to those in Table IV reveals that our point estimates and variances are similar.

We have used the two ‘simplest’ $g\{1, 1, 1, 1, 1\}$, $X, X_0, Q, L, L_0$. We could, however, use other functions. To minimize the variance there is an optimal $g$ that Robins\textsuperscript{23, 24, 25} has shown is a function of the unknown joint distribution of $\{T, L(T), Q(T), L_0(T)\}$.

12. ROBUSTNESS TO DATA ASSUMPTIONS

To analyse the MRFIT data we had to make several assumptions. For the logistic regressions we assumed that missing data was missing completely at random. We performed the logistic regressions for all three of our analyses on the subset of visits that had complete data under the model with the largest number of covariates (the time varying model). We eliminated only a total of $6.5\%$ of eligible visits due to either a missed visit or a missing covariate. Five per cent of the visits had missing information only on smoking status at $k$. For those visits we replaced the missing information with last known smoking status.

Since information on smoking and other time dependent covariates was gathered only at discrete times, we needed to make assumptions regarding the onset and duration of changes in smoking and covariate status. For smoking we assumed that one’s smoking status as measured at visit $k$ was the smoking status in the time interval $(k - 1, k]$. Performing the analysis under the extreme alternative assumption that smoking status at $k$ defines actual exposure at time $(k - 1, k]$ did not change our results.

Implicit in our logistic regression models for the probability of quitting at $k$ are assumptions about the temporal relationship between smoking status recorded at $k$ and the covariates recorded at $k$. For some covariates the relationship is obvious. For example, we know that blood pressure measurements actually made at visit $k$ occurred subsequent to the assessment of smoking status at $k$. Therefore in the logistic regressions for the probability of smoking at $k$, we only included blood pressure measurements made through $k - 1$. For other covariates, however, notably angina, there is no such obvious temporal ordering. With angina, for instance, there are two possible extremes. Either the development of angina in the interval $(k - 1, k]$ always precedes the decision to smoke at $k$, or vice versa. If the first is true, then angina can influence smoking status and we should include angina recorded at $k$ as a predictor of smoking at $k$; if the latter is true and the decision to smoke always occurs before the development of angina in the interval, we should include only angina up to $k - 1$ in the regression. Our analyses were insensitive to assumptions regarding these temporal relations.

13. ROBUSTNESS TO MODEL ASSUMPTIONS

The inference from our model rests upon three sets of assumptions:

1. We have all the covariates necessary for maintained comparability (1) to be true.
2. The RPSFTM (13) is correctly specified.
3. The probability of exposure at time \( k \) (18) is correctly specified.

If assumption 1 is false, then the G-test statistic is not a valid test of the causal null hypothesis. For instance, suppose even conditional on all measured \( L \) history, the amount of regular cardiovascular exercise one does at \( k \) is an independent predictor of both outcome and smoking.

In the absence of any data on cardiovascular exercise, we would expect (1) to be false. When (1) is false we have no valid \( z \)-level test of the causal null hypothesis. Since (1) is inherently about latent survival times, there is no way to test its veracity from the observable data. As always in observational analyses, causal inference requires that we believe there are no unmeasured confounders.

Model (13) also relates observable data to latent survival times. As discussed in Section 8, the causal null hypothesis implies \( \phi_0 = 0 \). The \( z \)-level tests of \( \phi_0 = 0 \) are \( z \)-level tests of the causal null hypothesis. Under a non-null alternative, however, we must specify the model correctly to have the interpretation we give in Section 8. As presented, (13) makes a strong non-interaction assumption. For example, suppose two individuals, \( m \) and \( n \), both of whom permanently quit smoking at time \( 0 \), die of an MI at the same time. According to (13), if they had both never quit smoking, their failure times, \( U_{m,0} \) and \( U_{n,0} \), are equal. This equality of \( U_{m,0} \) and \( U_{n,0} \), however, may be biologically implausible. For instance, let us suppose further that at time \( 0 \), individual \( m \) had more advanced coronary artery disease than individual \( n \) and, in contrast to the predictions of the model, had they both continued to smoke, \( m \) would have died before \( n \). When they both did quit smoking they died at the same time because \( n \) experienced a detrimental consequence of quitting that \( m \) did not; \( n \) gained a considerable amount of weight. This increase in weight had a negative effect on \( n \)'s cardiovascular health and counteracted some of the positive effect of quitting.

To allow for this interaction between the effect of quitting smoking and weight gain, we could have proposed a two parameter RPSFTM of the form

\[
U_{i,k} = k + \int_{k}^{T_i} \exp \left[ \psi_{10} Q(t) + \psi_{20} Q(t) I(t) \right] \, dt
\]

(24)

where \( I(t) \) is an indicator variable for a 5 per cent or greater increase in weight by time \( u \). \( \psi_{20} \) is positive if the beneficial effect of cigarette smoking decreases in people with weight gain. (Note that we have not entered the indicator variable into the model to control for the effect of weight gain as a confounder, but rather to account for its role as an effect modifier of quitting.) Even though model (24) includes an interaction between smoking and weight gain, the parameters \( \psi_{10} \) and \( \psi_{20} \) continue to represent the overall effect of smoking on time to failure. Methods for estimation of the direct biological effect of smoking cessation are considered in References 17–19, 21, 24 and 42. Generalization of the procedures of Section 9 to estimation of a multiparameter model such as (24) is straightforward.

Though allowing interactions, model (24) still assumes that the magnitude of treatment effect does not depend on unmeasured factors; that is, it assumes that the baseline failure times at \( k \), \( U_{i,k} \), are a deterministic function of the observables. Robins\(^{21,22,24}\) has proposed a class of failure time models, the structural nested failure time models, that includes the RPSFTM as a sub-class, and that allows the magnitude of treatment effect to depend on unmeasured factors. The method we presented here is valid for this less restrictive class of models. Since under the comparability assumption (1), the distribution of \( U_{i,k} \) is non-parametrically identified,\(^{17,22,24}\) in principle we are able to test the specification of model (13) or its generalization (24).
Finally, the consistency of our estimator also depends upon our ability to specify models correctly for the probability of quitting at $k$ given the past (18). To increase robustness relating to the specification of these models, we could add more covariates to the model. For example, in our time varying models we could add terms for angina and DBP history at $k - 1$ and $k - 2$; interactions between hypertensive drug at $k$ and group; etc. Adding more covariates usually results in less bias, and, provided we specify the more parsimonious model correctly, the efficiency with which we estimate $\psi_0$ never decreases and often increases.$^3, 4$ In practice, there is a limit to the number of parameters we can add, since the asymptotic normality of our tests and estimation of $\psi_0$ require that we estimate the parameters $\lambda_k$ in (18) at a rate of convergence of $n^{-1}$ or better.$^2$

14. DISCUSSION

Ideally, to answer the question of whether quitting smoking changes an individual's time to MI or death, we would possess two sets of failure times for everyone: a failure time when an individual smokes and a failure time when he quits. Lacking such unobtainable information, we are left with comparison of the outcomes of individuals who quit smoking with comparable individuals who remain smokers. In this paper we have formally defined what we mean by comparability in time varying exposure studies (1), and have asserted that it is unlikely one can achieve this comparability by conditioning only on baseline (non-time-varying) covariates. In particular, we undertook this analysis with the strong suspicion that individuals who developed the cardiac symptom angina are more likely to quit smoking and are at higher risk of failure than are asymptomatic individuals. Our examination of the data confirmed the existence of these relationships, and convinced us that we must account for angina when we assemble comparable groups for comparisons.

We have also considered the usefulness of the time-varying Cox proportional hazards model (7) given that angina is indeed a predictor of both smoking and risk. We have shown that we can use (7) to make some limited inference regarding the observable causal null hypothesis (3) provided that smoking does not affect the development of angina (10). Our data analysis, however, indicated that smoking did influence the development of angina. Furthermore, if the alternative were true, and the effect of smoking on outcome was at least partially mediated through an antecedent effect on angina, then entering angina into the CPH model would bias toward the null since we would be controlling for a variable on the causal pathway. This is a well recognized danger in time varying CPH models,$^{14}$ and, we suspect, the reason previous investigators have avoided placing angina in a CPH model.

In contrast, under our assumptions the G-test procedure used in this paper allows us to control for all measured covariates, angina included, and still give valid tests of the causal null hypothesis and consistent estimates of the parameter in the RPSFTM (13) that quantifies the overall effect of quitting on outcome. The key to the G-test procedure is its ability to yield consistent estimates of the parameter of the RPSFTM subject to the restrictions placed upon the observable data by maintained comparability (1). Specifically, when the causal null hypothesis (2) is true, these restrictions are the observable causal null hypothesis (3). Since (3) does not imply that the coefficient for the exposure effect in CPH models (5) or (7) is zero, those models (except under the conditions stated in Section 7) are not useful for inference regarding the causal null (2). In Appendix IV we express a restriction implied by (3) in terms of conditional hazards, and specify a CPH model different from model (5) or (7) that will produce valid a-level tests of (3) and thus of the causal null hypothesis (2). As we discuss in that appendix, however, in contrast to our G-test procedure, the test of (3) based on this CPH model can have very poor power. Indeed when the
beneficial effects of quitting at time \( k \) on survival do not manifest themselves until \( k + 1 \) (that is, there is a biological latency period of at least one year), this test will have power only equal to its \( \alpha \)-level!

Using the G-test procedure, we performed three separate analyses of the MRFIT data. We offer the time varying analysis as the correct one. Though our \textit{a priori} beliefs and the associations we observed in the data would have us regard the baseline analysis and smoking only analysis as confounded, we saw that ignoring these confounders had a negligible effect upon our estimates of \( \psi_0 \). Although a sensitivity analysis might have alerted us to the fact that the potential magnitude of confounding was small, we feel that the best way to examine the consequence of ignoring confounders is to use a method of analysis, such as the one we present, that is valid when adjusting for them.

From our analysis we reject the null hypothesis of no causal effect of quitting on time to first MI or death. We conclude that quitting smoking permanently prolongs by approximately 50 per cent the time to failure. Given that during the 6–7 years of follow-up of the study the underlying distribution of the time to failure if no one gave up smoking \((U_{ij0})\) is exponential, this corresponds to a causal hazard ratio of 0.65.

We have previously analysed these data\(^{46}\) using an RPSFTM without making any assumptions about comparability beyond those implied by randomization at time 0. The point estimate from that analysis is consistent with what we have found here.

The associations we found between smoking and angina, and angina and outcome are not unusual. Whenever exposure status at time \( k \) is subject to the conscious decision making of an individual or his doctor, one can anticipate that exposure will relate to symptoms of disease or disease markers. These symptoms or markers are themselves likely influenced by past exposure, and likely are predictors of future outcome. Even in the presence of such associations, provided that maintained comparability \((1)\) is true and that we have specified correctly the models for the probability of exposure at time \( k \) \((18)\), the G-test procedure gives valid \( \alpha \)-level tests of the causal null hypothesis \((2)\). If we specify correctly the model for the effect of exposure on outcome \( \text{the RPSFTM: models (13) or (24), the G-test procedure produces consistent estimates of its parameter(s).} \)

In this paper we have only considered censoring by end of follow-up and have ignored the problems created by missing data. References 21–25 discuss extensions of G-estimation that allow for censoring by competing risks and for missing treatment and covariate data. References 21, 24, 25 and 45 consider G-estimation of the effect of a time-varying exposure on the evolution of a non-failure time variable such as blood pressure. G-estimation is important for estimating the effect of a time-varying treatment on a repeated measures outcome since standard covariance adjustment, as for example, by generalized estimating equations,\(^{46}\) cannot consistently estimate the treatment effect\(^{25}\) when time-dependent confounders are present.

**APPENDIX I**

Robins\(^{18,20,22}\) has shown that given our fundamental assumption of maintained comparability \((1)\) the survival curve for any latent random variable \( U_{g \times k} \) relates to the observable data by the computation algorithm formulae

\[
P[U_{g \times k} > t|I_0 = 0] = \prod_{m=1}^{k} P[T > m|I_{m-1}, \tilde{q}_{m-1}, T > m - 1] \prod_{m=1}^{k} dF[I_{m-1}, \tilde{q}_{m-1}, T > m] \quad \text{for } k < t \leq k + 1. \tag{25}
\]
The $Q$ history in the conditioning events in (25) is the $Q$ history mandated by whatever hypothesized treatment, $h$, one considers. He has further shown** that when $L$ history is not a confounder (either (8) or (9) is true), (25) is equivalent to

$$P[U_{s+1} > t|L_0 = l_0] = P[T > t|T > k, l_0, q_k] \prod_{m=1}^k P[T > m|l_0, q_{m-1}, T > m-1]$$

(26)

for $k < t < k + 1$.

We can rewrite (26) as

$$P[U_{s+1} > t|L_0 = l_0] = \exp \left\{ - \int_0^t \lambda(u, q(u), l_0) \, du \right\}.$$  

(27)

Under the causal null hypothesis (2) the survival curves (27) are identical for all $h$. This implies (4) is true, and thus $\theta = 0$ in model (5). Therefore, as we state in Section 6, when there are no time varying confounders, a test of $\theta = 0$ in model (5) is a valid $z$-level test of the causal null.

**APPENDIX II**

In this appendix we provide a computational formulae for $\hat{\theta}(\psi)$ (the variance of $S(\hat{\theta}, \theta = 0, \psi)$), and demonstrate the relationship of our $G$-test to the usual score test for logistic regression.

Assume we have $n$ individuals who at time 0 either quit or do not quit smoking for the duration of their lives. In this data structure equation (20) simplifies to

$$P[Q_{i,0}|L_{i,0}, U_{i,0}(\psi)] = \frac{\exp(zW_i + \theta U_{i,0}(\psi))^{Q_{i,0}}}{1 + \exp(zW_i + \theta U_{i,0}(\psi))}.$$  

(28)

For fixed $\psi$, we write the likelihood of (28) as

$$L(z, \theta, \psi) = \prod_{i=1}^n L_i(z, \theta, \psi)$$  

(29)

where

$$L_i(z, \theta, \psi) = \{P_i(z, \theta, \psi)^{Q_{i,0}}[1 - P_i(z, \theta, \psi)]^{1 - Q_{i,0}}\}.$$  

(30)

The numerator of the score test* statistic of the hypothesis $\theta = 0$ is

$$S(\hat{\theta}, \theta = 0, \psi) = \sum_{i=1}^n U_{i,0}(\psi)[Q_{i,0} - P_i(\hat{\theta}, \theta = 0)]$$  

(31)

where $\hat{\theta}$ is the restricted MLE solution of model (29) when $\theta = 0$ and $\psi$ is fixed. Notice that in (31) we have written $P_i(\hat{\theta}, \theta = 0)$ rather than $P_i(z, \theta = 0, \psi)$ since under the hypothesis, $\theta = 0$, the predicted values are not a function of $\psi$. We write the observed information matrix for the likelihood of (29) as

$$I(\psi) = \begin{bmatrix} I_{\theta\theta}(\psi) & I_{\theta\psi}(\psi) \\ I_{\psi\theta}(\psi) & I_{\psi\psi}(\psi) \end{bmatrix}.$$  

$I(\psi)$ is the negative of the second derivative with respect to $\theta$, of the log of (29). We define $\hat{I}(\psi)$ as $I(\psi)$ evaluated at $\psi = \hat{\psi}, \theta = 0$. For example, the computational formulae for $I_{\theta\theta}(\psi)$ is

$$\hat{I}_{\theta\theta}(\psi) = \sum_{i=1}^n \{P_i(\hat{\theta}, \theta = 0)(1 - P_i(\hat{\theta}, \theta = 0))U_{i,0}(\psi)U_{i,0}(\psi)\}.$$  

(32)
When (28) is correctly specified and \( \psi = \psi_0 \), we know, by standard likelihood results,\(^{23,44}\) that (31) is asymptotically normal with asymptotic variance that we can estimate consistently by
\[
\hat{\Theta}(\psi_0) = \{ \hat{I}_{\lambda,\psi}(\psi_0) - \hat{I}_{\psi}(\psi_0) \hat{T}_{\alpha}(\psi_0) \hat{I}_{\psi}(\psi_0) \}. \quad (33)
\]
Given our actual data structure in which we have repeated observations for each individual, the numerator of our statistic becomes equation (22). In terms of asymptotic distribution,\(^{20}\) Robins\(^{23,44}\) has shown that we may treat the data as if we have \( N \) independent observations where \( N = \sum_{i=1}^{n_i} \text{int}(T_i) \). Equation (22) is asymptotically normal and has asymptotic variance that we can estimate consistently by (33) where now, for example,
\[
\hat{I}_{\lambda,\psi}(\psi) = \sum_{i=1}^{n} \sum_{k=0}^{m(T_i)} \{ P_{i,k}(\hat{x}, \theta = 0) (1 - P_{i,k}(\hat{x}, \theta = 0)) U_{i,k}(\psi) U_{i,k}(\psi) \}. \quad (34)
\]

**APPENDIX III**

In Section 10 we assumed the latent failure times \( U_{i,\psi} \) followed a Weibull distribution, and estimated the parameters of that distribution using maximum likelihood under an assumption of independent censoring. In this appendix we describe how we handled the end of follow-up censoring so as to assure that the censoring times were stochastically independent of the failure times, \( U_{i,\psi} \).

The minimum \( C_i \) in our population was \( C_i = 6.2 \). Thus, given that \( \psi_0 = 0.43 \), all those individuals whose \( U_{i,\psi} \leq 6.2 \) have an observable \( U_{i,\psi} = \) that is independent of what smoking history an individual followed, if \( U_{i,\psi} \leq 4.2 \), then \( T_i < C_i \). Therefore, for the purpose of estimating the Weibull parameters for the \( U_{i,\psi} \), we define a censoring time, \( S_i \), as \( S_i = 4.2 \) for all \( i \). Thus, we consider the data as \( X_i = \min(U_{i,\psi}, S_i) \). The observable data then become \((X_i, \gamma_i)\), where \( \gamma_i = I(S_i > U_{i,\psi}) \). Since \( S_i \) and \( U_{i,\psi} \) are independent, we can now write out and estimate the likelihood of the \((X_i, \gamma_i)\) using the standard approach for right censored data under an independent censoring mechanism. Note, that in contrast to the above, had we defined censoring as we do in Section 11 (that is \( S_i = C_i \text{exp}(0.43) \)), then independence of \( U_{i,\psi} \) from \( S_i \) would require the additional assumption that \( U_{i,\psi} \) was independent of \( C_i \); that is, that baseline risk did not change during the enrolment period of the study.

**APPENDIX IV**

Maintained comparability (1) and the causal null hypothesis (2) imply (3) which in turn implies
\[
\lambda(t; \bar{Q}_{i,k}, \bar{L}_{i,k}, T_i \geq t) = \lambda(t; \bar{Q}_{i,k} = 0, \bar{L}_{i,k}, T_i \geq t) \quad \text{for } k < t \leq k + 1. \quad (35)
\]
Therefore, we can produce valid \( \alpha \)-level tests of (2) by testing \( \theta = 0 \) in a correctly specified CPH model such as
\[
\lambda(t; \bar{Q}_{i,k}, \bar{L}_{i,k}) = \lambda_0(t) \exp(\theta \bar{Q}_{i,k} + s \bar{W}(\bar{L}_{i,k}) + \gamma Z(\bar{L}_{i,k})) \quad \text{for } k < t \leq k + 1 \quad (36)
\]
where \( \bar{W}(\bar{L}_{i,k}) \), \( Z(\bar{L}_{i,k}) \) are real (possibly vector) valued functions of \( \bar{Q}_{i,k}, \bar{L}_{i,k} \) and \( \theta, s, \gamma \) are unknown parameter vectors. A test of \( \theta = 0 \) is in model (36), however, though a valid test of the causal null hypothesis, has poor power to detect effects of quitting at time \( k \) that are manifested by improved survival only after time \( k + 1 \). In such a case the hazard of failure for \( T \) in the interval \( k \) to \( k + 1 \) will not depend on the treatment received at \( k \) given the past: that is, equation (35) will be true. Hence our test of \( \theta = 0 \) in (36) has power \( \alpha \) regardless of how great an effect quitting at \( k \) has on the hazard subsequent to \( k + 1 \). As a specific example, suppose that the effect of quitting
can be correctly described by the RPSFTM (37):

\[ U_{i,k} = T_i \text{ for } T_i \leq k + 1 \]

\[ U_{i,k} = k + 1 + \int_{k+1}^{T_i} \exp \left[ \phi_0Q_1(u-1) \right] du \text{ for } T_i > k + 1. \]

In this model \( \psi_0 < 0 \) implies that quitting at \( k \) is beneficial for those who survive past \( k + 1 \). A \( G \)-test of \( \psi_0 = \psi \) can be constructed as before with \( U_{i,k}(\psi) \) equal to the right hand side of equation (37) when \( \psi_0 \) is replaced by \( \psi \). Since according to (37) \( U_{i,k}(\psi = 0) = T_i \), a test of the null hypothesis \( \psi_0 = 0 \) proceeds as in Section 9. However, unlike the test of \( \theta = 0 \) in CPH model (36), the power of the \( G \)-test of the hypothesis \( \psi_0 = 0 \) will increase as the magnitude of the treatment effect \( \psi_0 \) increases.

ACKNOWLEDGEMENTS

This work was supported in part by NIH grants NC1ST32CA90001, 5R01ES03405, R01AI132475, 5KO4ES01810 and 5P01ES00002. We would like to extend our warmest thanks to Dr. James Neaton and Ms. Marjorie Ireland, who provided us with the MRFIT data. We are also indebted to the reviewers and editor whose careful readings and suggestions improved the paper’s clarity.

REFERENCES


