Principal stratification designs to estimate input data missing due to death

Constantine E. Frangakis[1], Donald B. Rubin[2], Ming-Wen An[1], and Ellen MacKenzie[3]

[1]Department of Biostatistics, Johns Hopkins University
[2]Department of Statistics, Harvard University
[3]Department of Health Policy and Management, Johns Hopkins University

SUMMARY. We consider studies of cohorts of individuals after a critical event, such as an injury, with the following characteristics. First, the studies are designed to measure "input" variables, which describe the period before the critical event, and to characterize the distribution of the input variables in the cohort. Second, the studies are designed to measure "output" variables, primarily mortality after the critical event, and to characterize the predictive (conditional) distribution of mortality given the input variables in the cohort. Such studies often possess the complication that the input data are missing for those who die shortly after the critical event because the data collection takes place after the event. Standard methods of dealing with the missing inputs, such as imputation or weighting methods based on an assumption of ignorable missingness, are known to be generally invalid when the missingness of inputs is nonignorable, that is, when the distribution of the inputs is different between those who die and those who live. To address this issue, we propose a novel design that obtains and uses information on an additional key variable – a treatment or externally controlled variable, which if set at its "effective" level, could have prevented the death of those who died. We show that the new design can be used to draw valid inferences for the marginal distribution of inputs in the entire cohort, and for the conditional distribution of mortality given the inputs, also in the entire cohort, even under nonignorable missingness. The crucial framework that we use is principal stratification based on the potential outcomes, here mortality under both levels of treatment. We also show using illustrative preliminary injury data, that our approach can reveal results that are more reasonable than the results of standard methods, in relatively dramatic ways. Thus, our approach suggests that the routine collection of data on variables that could be used as possible treatments in such studies of inputs and mortality should become common.

KEY WORDS: Causal inference; Censoring by death; Missing data; Potential Outcomes; Principal Stratification; Quantum mechanics.

1. Introduction.

We consider studies that interview cohorts of individuals after a critical event, such as injury or stroke, with the following two characteristics. First, the studies are designed to measure "input" variables, which describe the period before the critical event, and to characterize the distribution of the input variables in the cohort. Second, the studies are designed to measure "output" variables, primarily mortality after the critical event, and to characterize the predictive (or conditional) distribution of mortality given the input variables in the cohort. Such studies, however, are often complicated by the fact that the input data are missing for those who die shortly after the critical event because the data collection takes place after the event.

This problem, input data missing due to death, occurs commonly, for example, in studies of elders (Cornoni et al., 1993; Reuben, 1995; Cohen, 2002), or victims of injuries (e.g., MacKenzie et al., 2006). The goals we address for such studies are how to estimate the inputs missing due to death, and how to characterize the predictive (or conditional) distribution of mortality given the input variables in the cohort. Answers to these goals are important because, first, they can be used to better alert the individuals and their physicians about increases in risks, and second, they inform about the pathways of such risks.

As a motivating example, consider the National Study on the Costs and Outcomes of Trauma Centers (NSCOT, MacKenzie et al., 2006). That study used hospital discharge records to identify and enroll individuals who received care for injuries. The first follow-up visit was scheduled at three months. During this visit, patients were interviewed about their pre-injury disability, as measured by "activities of daily living (ADL)". It is of interest to evaluate the relation that prior disability has to the risk of death following an injury. However, some patients died as a result of injury, before this first follow-up visit. Thus, the ADL values are missing for these patients. If these missing past ADL values have a different distribution than the observed

past ADL values among survivors, standard methods cannot estimate that relation.

Another class of examples arises in the evaluation of the effect that a periodic exposure (e.g., to drug) has on the risk of a critical event using a case-crossover design (Maclure, 1991). *In its basic form, this design aims to measure, for each one of a group of injury cases, the gap time between the last exposure and the critical event, and a measure of that person's typical frequency of past exposure. A measure of association between exposure and the critical event is then defined by comparing the observed gap times to their distribution that would be expected if the critical event had been unrelated to the exposure process defined by the past frequencies.* In this design, even if we know the victims' most recent exposure to drugs (e.g., by blood measurement), the frequency of past exposure becomes missing for those who die as a result of severe injuries, and this missingness is usually ignored (e.g., Vinson et al., 1995). As discussed below, such missingness needs to be addressed by new and more appropriate methods. Such examples are summarized in Table 1.

Table 1 here.

Standard methods confronted with missing data from death, as also noted by Zhang and Rubin (2003), can be classified into three types. The first type is concerned only with the observed data (e.g., cause-specific hazards, dating to Prentice et al. 1978; and partly conditional on being alive, Kurland and Heagerty, 2005); these methods are not relevant to our problem because they do not attempt to estimate the missing data. The second type of method assumes ignorability (Rubin, 1976) of missing data and essentially replaces them with data matched from fully observed strata, either across time from the same person, or across people for the same time (e.g., McMahon and Harrell, 2001; Lin, McCullough and Mayne, 2002) or both; these methods are known to be inappropriate when the distribution of data missing data due to death differs from that in observed strata (Rubin, 1978). The third type posits non-ignorable assumptions relying simply on the parametric structure of models (e.g., Fairclough, Peterson

and Chang, 1998); these methods are *sensitive to the parametric assumptions because, without such assumptions, the distributions of interest are not identifiable unless additional design structure is introduced.*

We address the problem's goals from a combination of design and analyses perspectives. First, we recognize that the problem is related to, but differs from, the problem of censoring by death discussed in Rubin (2000), Frangakis and Rubin (2002), and developed by Zhang and Rubin (2003). The goal of the latter problem is to compare treatments on potential outcomes (Neyman 1923; Rubin, 1974, 1978) when some patients in either treatment die. In that problem, the *future* outcome of a person who dies is "missing", not because it exists and is unobserved, but because it is not defined. Because the patients who die may not be comparable between the two treatments, death creates the need to define meaningful treatment effects on the outcomes. Such effects are well defined if we restrict attention to patients who would survive no matter which treatment they would receive (Rubin, 2000) rather than to the larger group of patients who are observed to survive. This group of patients, who would survive no matter the treatment, is a special case of a "principal stratum" (Frangakis and Rubin, 2002), that is, here, a stratum defined by a patient's joint potential outcomes of death under the two treatments. Thus, in that case, the principal strata are critical for defining treatment effects. In the present problem, the variable of interest is a well defined input preceding death, and is missing because the attempt to record it takes place after death. The key, from the design perspective, then, is to recognize that the missing data of an individual who dies, would be observed "under explicit alternative conditions for which the same individual would have survived". Formalizing this, we show that it is also important here for the goal of estimating the missing information, that: (1) the *design* finds data on factors (e.g., treatments) that (1a) could have prevented deaths and (1b) were assigned to the individuals after the time when the inputs of interest became defined but before the time of death; and (2) these data be analyzed

using principal stratification.

In the next section, we formulate more explicitly the problem and its goals, and formalize the proposed design with data on externally controllable factors, such as treatments, that can prevent deaths. In Section 3, we describe a method can address our goals using the data from the proposed design and the framework of principal stratification. We show that the proposed method allows the distribution of missing inputs to differ systematically from the distribution of the observed inputs, yet this method is able to estimate the distribution of the missing inputs. In Section 4, we demonstrate using preliminary data from NSCOT including transport time to hospital as the externally controllable factor, that our design and analysis method can uncover results that are dramatically different and more plausible than those of standard methods. Section 5 provides extensions of the proposed methods in more general situations. Section 6 discusses the commonalities and differences between this and other related uses of principal stratification. Section 7 concludes with remarks, including connections between this new, interventional approach to missing data and the principles of quantum mechanics.

2. Design using principal stratification.

2.1 Initial design and goals.

Consider a cohort of individuals who had a critical event (at time say $t = 1$), such as an injury (e.g., crash). We are interested in learning about a variable A that takes its value at a time, say $t = 0$, before the critical event and so is called an input. For example, A can be activities of daily living that the person cannot perform, or exposure to drugs. To record A, we schedule an interview at a time, say $t = 2$, after the critical event, e.g., an interview at discharge from the hospital. However, a subset of individuals die before the interview, as a result of the critical event; for those individuals, the value of A still exists, since it occurred before death, but becomes missing because there is no interview.

Throughout, we use i to index an individual. Let A_i be the value of A for individuals at $t = 0$; and let $S_i^{obs} = 1$ for surviving individuals at $t = 2$, and 0 otherwise. This initial setting is shown in Fig. 1(a).

Goals. We wish to address the following: (a) Estimate the distribution of the past input A_i for the people who died without reporting them; and (b) Estimate quantities such as predictive distributions and associations that are defined based on the distribution of all values A_i, missing and observed, for example, the prediction of death based on A_i. The first goal is important for characterizing the distribution of the inputs for all individuals. The second goal differs from predicting death from the *observed* inputs in this study, $\text{pr}(S_i^{obs} = 0 \mid \{A_i : A_i \text{ is observed }\})$, which is by definition deterministically 0 and is of no interest. Goals of type (b) are important because they inform us about the degree to which the past inputs A_i in the original cohort are actually related to death (or to the critical event using additional data from people without that event). Because of the deaths, the inputs A are not all reported *in this* study, so these relations need to be estimated indirectly. These relations should suggest better monitoring methods in *subsequent* studies, which would alert physicians and individuals about sudden increases in the risk of death. Also, goals (b) contribute by helping medical research understand the pathways through which those inputs relate to critical events and death.

2.2 New design elements and principal strata.

Consider the following additional design elements:

(i) For all individuals, we find and record a factor or treatment (labeled Z_i) that was assigned externally (that is, by a person or process other than the individual), and a level of which could have prevented death for those who died. For this factor, let $z = 0$ denote a standard level, and $z = 1$ denote the more effective level. For example, for injuries, such a "treatment factor" can be the *transport time* (long or short) from the time of injury to

arrival at the hospital or to surgery, whereas for strokes or myocardial infarctions, such a factor can be the prompt administration of a thrombolytic drug.

(ii) We also record covariates X_i that were used to decide the level Z_i of the factor for the individual. The variables X_i may correlate with the input A_i.

The level of factor z to which a particular individual is assigned can affect the future of that individual, although we assume it cannot affect the future of a different individual (no interference, Rubin, 1978; Cox, 1992). For an individual i, denote by $S_i(z)$ the potential survival outcome (Rubin, 1978) that indicates the survival status if the individual is assigned level z of the factor. It is then important, as in Rubin (2000), Frangakis and Rubin (2002) and Zhang and Rubin (2003), to consider the principal strata of survival, that is, the strata of the individuals with respect to the joint values of $(S_i(0), S_i(1))$. These are generally the following: (1) individuals who would survive no matter the level of z, that is, $S_i(0) = S_i(1) = 1$; (2) individuals who would die under the standard level but would live under the effective one, that is, $S_i(0) = 0$ and $S_i(1) = 1$; (3) individuals who would die no matter the level, that is, $S_i(0) = S_i(1) = 0$; and (4) individuals who would survive under the standard level but would die under the effective treatment, that is, $S_i(0) = 1$ and $S_i(1) = 0$. We denote the principal stratum of individual i by P_i and label the above four possible strata as "always survivors", "protectable", "never survivors", and "defiers", respectively, combining terminology of Angrist, Imbens and Rubin (1996), and Gilbert, Bosch and Hudgens (2003) for vaccines.

Our main argument is that addressing the goals (a) and (b) can be helped by recording and using data on such a factor z (there can be more than one choice) that can justify plausible assumptions about the assignment of Z_i and about the principal strata.

Figure 1 here.

A simple example reveals how our structure can help us achieve our goals. Consider a factor z that can justify the following two assumptions (for extensions see Sec. 5):

Assumption 1. *Ignorable assignment of external factor* : The levels Z_i are independent of (A_i, P_i) conditionally on the variables X_i that were used for assignment.

Assumption 2. *Preventability of deaths from external factor* : Individuals are either $P_i =$ "protectable" by the effective level $(z = 1)$ of the factor, or else "always survivors" .

Assumption 1 is plausible when we choose z and X_i so that conditionally on X_i the reasons for the remaining variability of Z_i are independent of the individuals' health prior to the critical event. For example, we can ask physicians to tell us all the variables they used to decide assignment of a treatment z. So, the external assignment of z makes its ignorability achievable, whereas this is not true for an assumption of "ignorability of death", which is typically made by the standard methods (Sec. 1). Note that, by definition, the values of A_i and P_i are not affected by the actual treatment that is assigned (Frangakis and Rubin, 2002). The second assumption excludes "never survivor" and "defier" patients, and is related to the monotonicity assumption in other settings (e.g., Angrist, Imbens and Rubin, 1996). Preventability, when combined with ignorability, is testable from the observed data, since under these assumptions we must observe that among individuals within levels of X_i and assigned the "effective" treatment, all survive, whereas among those assigned the standard treatment some die and some survive, as in Fig. 2(b). *More generally, some notion of both, the ignorability of the controllable factor, and a type of monotonic effect of that factor on the reason of missing outcomes (here mortality) are critical for using this design. Nevertheless, the preventability assumption is more flexible than it originally appears when made within levels of the covariate strata X_i. The preventability Assumption 2 can also be relaxed to allow for "never survivors" as discussed in Section 5.1.* We now show how the above design addresses our goals.

3. Estimability of input data missing due to death.

Distribution of missing input data.

For the observed data, we assume without loss of generality that we are already within covariate strata $X_i = x$, so, for brevity we omit the explicit conditioning on X_i in the notation of the distributions below. The possibly missing input A_i is taken as an indicator for poor functional ability (e.g, dichotomized activities of daily living (ADL) =1 for poor status).

Consider first the goal of estimating the distribution of the missing functional inputs, $\mathrm{pr}(A_i = 1 \mid S_i^{obs} = 0, Z_i = 0)$. The above ignorability of the assignment of the prevention factor levels Z_i reflects that, conditionally on the variables X_i, and on which we have already stratified, assignment of Z_i balances all other covariates, including the input A_i, which is a covariate that took its value before the prevention factor Z_i was assigned, even though assignment of Z_i preceded the time when A_i was to be measured. In other words, because A_i is a covariate and Z_i is effectively randomized (given X_i), the proportion $\mathrm{pr}(A_i = 1 \mid Z_i = 0)$ of poor inputs among individuals assigned the standard prevention level of z equals the proportion $\mathrm{pr}(A_i = 1 \mid Z_i = 1)$ among those assigned the effective prevention level. Since, the former group includes both individuals with observed and missing values, we have that:

$$\begin{aligned}\mathrm{pr}(A_i = 1 \mid Z_i = 1) &= \mathrm{pr}(A_i = 1 \mid Z_i = 0) \\ &= \sum_{s=0,1} \mathrm{pr}(A_i = 1 \mid S_i^{obs} = s, Z_i = 0)\mathrm{pr}(S_i^{obs} = s \mid Z_i = 0). \qquad (1)\end{aligned}$$

From the observed data, as Fig. 2(b) shows, we can estimate directly the proportion $\mathrm{pr}(A_i = 1 \mid Z_i = 1)$ of people who had had poor function among those assigned the effective level of z. The equality in (1) then implies that we can also estimate the proportion $\mathrm{pr}(A_i = 1 \mid Z_i = 0)$ of people who had had poor function among those assigned the standard level of z. Moreover, Fig. 2(b) shows that we can also directly estimate from the observed data: the

proportion $\text{pr}(S_i^{obs} = 1 \mid Z_i = 0)$ of survivors among individuals assigned the standard z; and the proportion $\text{pr}(A_i = 1 \mid S_i^{obs} = 1, Z_i = 0)$ who had poor function among those who survived after being assigned the standard level of factor z. It follows then, from (1), that the distribution of missing past inputs can be expressed as

$$\begin{aligned} &\text{pr}(A_i = 1 \mid S_i^{obs} = 0, Z_i = 0) \\ &\quad = \frac{\text{pr}(A_i = 1 \mid Z_i = 1) - \text{pr}(A_i = 1 \mid S_i^{obs} = 1, Z_i = 0)\,\text{pr}(S_i^{obs} = 1 \mid Z_i = 0)}{\text{pr}(S_i^{obs} = 0 \mid Z_i = 0)}. \end{aligned} \tag{2}$$

Therefore, we have reduced the unknown distribution of missing input data to an expression, the RHS of (2), that involves quantities that can be directly estimated as discussed above. This calculation is related to the instrumental variables equations of the effect of a treatment on post-treatment outcomes in a trial with non-compliance (Imbens and Rubin, 1997). However, the context and goal of the problem here are different, and this parallel arises from the more fundamental commonality of "principal stratification" shared between the two types of problems (see Sec. 6).

Relation between input and mortality.

The ability to estimate better the missing data allows us to also examine better relations between those data and clinical variables. As an example, we show here how we can estimate the degree to which the input A_i predicts death. Because death depends on the principal strata P_i and the level of the prevention factor, it is important to examine if the input A_i predicts the principal strata of death. This would indicate that A_i predicts the underlying predisposition of a person to die.

Specifically, we wish to estimate:

$$\text{pr}(S_i(0) = 0 \mid A_i = a) = \frac{\text{pr}(S_i(0) = 0)\text{pr}(A_i = a|S_i(0) = 0)}{\text{pr}(A_i = a)}, \tag{3}$$

and compare (3) with $a = 0$ and 1. From the top of (1), we have that $\text{pr}(A_i = 1)$ equals the directly estimable proportion $\text{pr}(A_i = 1 \mid Z_i = 1)$ under the effective prevention level. Moreover, from ignorability of treatment assignment with respect to the principal strata, we have that the protectable patients $\{i : S_i(0) = 0\}$ are balanced between the levels of z (all probabilities are implicitly given X_i), and so $\text{pr}(S_i(0) = 0)$ in the RHS of (3) equals the directly estimable proportion $\text{pr}(S_i^{obs} = 0 \mid Z_i = 0)$ of patients who die under the standard prevention level, where the principal strata are observed (see Fig. 2(b)). Also by ignorability, the proportion $\text{pr}(A_i = a | S_i(0) = 0)$ of protectable patients who have input a, involved in the RHS of (3), is also balanced between the levels of z and so equals the proportion of patients with input a among those who die in the standard prevention level, i.e., $\text{pr}(A_i = a \mid S_i^{obs} = 0, Z_i = 0)$, where the latter is estimable from (2). These arguments show estimability of the proportions in (3). Using these arguments to substitute the RHS of (3) with estimable quantities based on (2), we can express the relative risk of being a protectable (not always survivor) patient when having poor versus good input A_i as

$$\begin{aligned}\frac{\text{pr}(S_i(0) = 0 \mid A_i = 1)}{\text{pr}(S_i(0) = 0 \mid A_i = 0)} &= \frac{\text{pr}(A_i = 0 \mid Z_i = 1)}{\text{pr}(A_i = 1 \mid Z_i = 1)} \\ &\times \frac{\text{pr}(A_i = 1 \mid Z_i = 1) - \text{pr}(A_i = 1 \mid S_i^{obs} = 1, Z_i = 0)\,\text{pr}(S_i^{obs} = 1 \mid Z_i = 0)}{\text{pr}(S_i^{obs} = 0 \mid Z_i = 0) - \text{pr}(A_i = 1 \mid Z_i = 1) + \text{pr}(A_i = 1 \mid S_i^{obs} = 1, Z_i = 0)\,\text{pr}(S_i^{obs} = 1 \mid Z_i = 0)},\end{aligned} \tag{4}$$

where the quantities in the RHS of equation (4) are all directly estimable as described in the paragraph following (1).

4. Demonstration.

We return to the NSCOT study (MacKenzie et al., 2006) on injuries described in Sec. 1. To illustrate the contrast between our approach to missing data and standard approaches, we consider patients who have sustained injuries with a relatively low ($X_i = 0$) or high ($X_i = 1$)

severity ($n = 354, 135$ respectively). The follow-up interview is scheduled three months after the injury to measure by questionnaire the functional status ($A_i = 1$ for poor ADL) that existed before injury, and this is missing if injured person i dies before the interview as a result of the injuries. The prevention factor z we use here is based on the time it took to transport the injured person to the hospital.

Regarding the assumption of ignorability of the assignment mechanism of the transport time to hospital, the two main reasons for variability of this time are (a) the severity of the injury as judged by medical personnel - more severe injuries are attempted to be transported faster; and (b) external reasons such as time of day, distance, traffic, or weather, that prevent fast transport, but that are themselves in principle not directly related to the person's health before injury. It is therefore plausible to assume ignorable assignment of Z_i after conditioning on the measured severity of injury X_i used to decide Z_i: among individuals of the same injury severity X_i (high, or low, see Table 2), those transported slowly are assumed to have the same distributions of past ADL A_i and principal strata P_i as the individuals transported quickly. Of course, one may wish to adjust for additional levels of covariates to remove possible remaining confounding, for example, using the approach of Sec. 5.2, but the principles for those analyses remain the same. The assumption that quick transportation to hospital can prevent an important proportion of deaths is supported both by literature for other critical events (e.g., GISSI 1986), and empirically by our data: within either of our strata (high, or low) of injury severity X_i, there were no deaths for injuries delivered to the hospital within 10 minutes, although there were 19% deaths for patients with a high injury severity delivered later than 10 minutes and 5% deaths for patients with a low injury severity delivered later than 10 minutes. Based on the above, Table 2 gives relevant summary proportions, directly computed from the data. We treat these summaries here as population proportions, because they indicate plausible results for each method. Inferential statements were not planned for and so do not

achieve statistical significance, since the study had not been planned to use the new design.

Table 2 here.

Focusing first on high injury severity, there were $\text{pr}(Z_i = 1) = 8\%$ of patients transported quickly; among the patients who were transported slowly, 81% survived, i.e., $\text{pr}(S_i^{obs} = 1 \mid Z_i = 0)$=81%; among those transported quickly, all survived, i.e., $\text{pr}(S_i^{obs} = 1 \mid Z_i = 1)$=100% (not shown); of those, there were 9% who had poor ADL before injury, i.e., $\text{pr}(A_i = 1 \mid Z_i = 1)$=9%; and among those who survived after being transported slowly, 5% had poor past A_i, i.e., $\text{pr}(A_i = 1 \mid S_i^{obs} = 1, Z_i = 0)$=5%. Then, the approach that would estimate the protectable patients' missing data distribution $\text{pr}(A_i = 1 \mid S_i^{obs} = 0, Z_i = 0)$ with the distribution of observed data after matching on slow time $Z_i = 0$ would give 5% poor function. On the other hand, an approach that would estimate the missing data distribution with the observed data without matching on time would give $\text{pr}(A_i = 1 \mid S_i^{obs} = 1)$ which equals $\sum_z \text{pr}(A_i = 1 \mid S_i^{obs} = 1, Z_i = z) \frac{\text{pr}(S_i^{obs}=1|Z_i=z)\text{pr}(Z_i=z)}{\sum_{z'} \text{pr}(S_i^{obs}=1|Z_i=z')\text{pr}(Z_i=z')}$, and which, using the information given in Table 2, gives 5.4%. More generally, the result of the standard methods is bounded to be between the directly observed $\text{pr}(A_i = 1 \mid S_i^{obs} = 1, Z_i = z)$, for $z = 0, 1$ (here, between 5% and 9%), as a convex combination of the two.

With the new method however, the missing proportion of poor past function for protectable patients is allowed to be different from the observed strata. In particular, from (1), the missing proportion $\text{pr}(A_i = 1 \mid S_i^{obs} = 0, Z_i = 0)$ must be such that when mixed with the proportion of $\text{pr}(A_i = 1 \mid S_i^{obs} = 1, Z_i = 0) = 5\%$ of poor past function for always survivors, the result should be the proportion of $\text{pr}(A_i = 1 \mid Z_i = 0) = \text{pr}(A_i = 1 \mid Z_i = 1) = 9\%$ observed for all patients transported quickly to the hospital (Fig. 1(b)). The fact that, by (1), this is a convex mixing based on the probabilities $\text{pr}(S_i^{obs} = s \mid Z_i = 0)$, for $s = 0, 1$, implies that the missing proportion $\text{pr}(A_i = 1 \mid S_i^{obs} = 0, Z_i = 0)$ of poor past function for the protectable patients must be *higher* than the mixture, $\text{pr}(A_i = 1 \mid Z_i = 1) = 9\%$. Using (2), the missing

proportion $\text{pr}(A_i = 1 \mid S_i^{obs} = 0, Z_i = 0)$ is $\{9\% - (5\%)(81\%)\}/(100\% - 81\%) = 26\%$. This shows that the actual result can be estimable and substantially different from those of the standard methods. Note that this proportion is in line with a hypothesis that those who died had generally poorer past ADL than the survivors. Analogous comparisons are obtained for injuries with low severity. Finally, the larger proportions of poor ADL for low versus high injury severity is in accordance with the hypothesis that individuals who sustain injuries of light severity *and* who, nevertheless, *need* hospitalization, were more frail before the injury than individuals who get hospitalized after sustaining a severe injury.

The relative risk in (4) is implicitly assumed to equal 1 by the standard method that replaces the missing data distribution $\text{pr}(A_i = 1 \mid S_i^{obs} = 0, Z_i = 0)$ with that of the observed data after matching on the prevention level, that is, with $\text{pr}(A_i = 1 \mid S_i^{obs} = 1, Z_i = 0)$. With the new method, however, and the empirical proportions of Table 2, the relative risk in (4) is estimated to be 13.7 and 3.6, for low and high injury severity, respectively. This means that, even after conditioning on observed strata, the possibly missing functional ability is an important predictor of the underlying ability of a patient to survive the injury when transportation takes a standard time to the hospital. The first implication is that follow-up e.g., of individuals with history of poor functionality, should use new designs (e.g., based on automated reporting devices) to make sure that some dimensions of functional ability be measured at higher frequency. This would give better prediction for which patients transition to high risk for death from a critical event. The second implication is that sudden changes to low functional ability inputs should be examined physiologically to understand and ultimately address the pathways through which these inputs predict death from injury even in the short term.

5. More general role of new methods.

5.1 *Partial preventability.*

The new methods are important also for more general input data, designs and assumptions. A plausible prevention factor may partly, but not fully, prevent death. For example, prompt delivery of thrombolytic drugs prevents death after stroke in some but not all cases (GISSI, 1986). More specifically for such settings, we consider an external factor z that satisfies no interference and Assumption 1, as in Section 2.2, and a generalization of Assumption 2:

Assumption 2'. Partial preventability of deaths from external factor *: Individuals are either* P_i= *"never survivors", "protectable", or "always survivors".*

For never survivors – those who would not survive no matter the factor's level – the observation of outcomes then remains essentially undefined just based on this factor, and so is not estimable without further assumptions. So the goal in this setting is limited to the estimation of the distribution of missing inputs for protectable patients under the standard level of assignment, which equals $pr(A_i \mid P_i =$ *protectable*$)$ *by Assumption 1. Standard methods cannot estimate correctly this distribution, as they cannot do so in the setting given in Sections 2 and 3. Yet we show below that this distribution is still estimable without further assumptions.*

To see this, note that the distribution of observed inputs under the effective factor level, as in Section 2.2, is still a mixture of the distribution among protectables and always survivors. Letting p, a *stand for protectables and always survivors, respectively, we then have*

$$
\begin{aligned}
pr(A_i = 1 \mid S_i^{obs} = 1, Z_i = 1) = \\
&= \sum_{q=p,a} pr(A_i = 1 \mid P_i = q, S_i^{obs} = 1, Z_i = 1) \times pr(P_i = q \mid S_i^{obs} = 1, Z_i = 1) \\
&= \sum_{q=p,a} pr(A_i = 1 \mid P_i = q) \times pr(P_i = q)/pr(P_i \in \{p, a\}) \qquad (5)
\end{aligned}
$$

where the last equality for the first summand arises first, because S^{obs} is a function of P and Z, and then because A, P is independent of Z, by Assumption 1.

To recover the target of interest, $pr(A_i = 1 \mid P_i = p)$, from (5), note that, since among the patients assigned the effective level $Z_i = 1$, those who survive are the protectables and always survivors, the proportions $pr(S_i^{obs} = 1 \mid Z_i = 1)$ and $pr(P_i \in \{p, a\})$ are equal. Moreover, since among those assigned the standard level, $Z_i = 0$, those who survive are always survivors, it follows that the proportions $pr(S_i^{obs} = 1 \mid Z_i = 0)$ and $pr(P_i = a)$ are equal, and the distribution of input data $pr(A_i = 1 \mid P_i = a)$ equals the directly estimable distribution $pr(A_i = 1 \mid S^{obs} = 1, Z_i = 0)$. By substituting these in (5) and after some rearrangement of terms we find that the target distribution satisfies

$$pr(A_i = 1 \mid P_i = p) =$$
$$\frac{pr(A_i = 1 \mid S_i^{obs} = 1, Z_i = 1)\, pr(S_i^{obs} = 1 \mid Z_i = 1) - pr(A_i = 1 \mid S_i^{obs} = 1, Z_i = 0)\, pr(S_i^{obs} = 1 \mid Z_i = 0)}{pr(S_i^{obs} = 1 \mid Z_i = 1) - pr(S_i^{obs} = 1 \mid Z_i = 0)}$$

Therefore for the subset of patients that are protectable or always survivors we can still assess the ignorability of missingess of data, and also find the direction along which its violation occurs (e.g., if such input data for those who died were higher on average than the observed ones). Thus in such more general settings, the importance of the new methods is essentially intact for addressing the scientific goals.

5.2 Modeling covariates.

Suppose we still make Assumptions 1 and 2', but we first wish to condition on multiple, and possibly continuous, covariates X_i, and that to do so, we model the distribution of the principal strata of survival and of a continuous input given principal strata by parametric functions

$$\begin{aligned} l^{(\mathrm{P})}(q, x, \beta^{(\mathrm{P})}) &:= \mathrm{pr}(P_i = q \mid X_i = x, \beta^{(\mathrm{P})}), \text{ and} \\ l^{(\mathrm{A})}(a, q, x, \beta^{(\mathrm{A})}) &:= \mathrm{pr}(A_i = a \mid P_i = q, X_i = x, \beta^{(\mathrm{A})}), \end{aligned} \tag{6}$$

where the last function is defined only for q = protectable, or always survivor. Denote by $\mathcal{P}(Z_i, S_i^{obs})$ the set of possible principal strata as a function of the observed level Z_i and survival status S_i^{obs}: if $Z_i = 0$ (standard) and $S^{obs} = 1$ (alive), then $\mathcal{P}(Z_i, A_i^{obs}) = \{\text{always survivor}\}$; if $Z_i = 0$ and $S^{obs} = 0$ (dead), then $\mathcal{P}(Z_i, S_i^{obs}) = \{\text{protectable, never survivor}\}$, if $Z_i = 1$ (effective) and $S^{obs} = 0$ (dead), then $\mathcal{P}(Z_i, S_i^{obs}) = \{\text{never survivor}\}$, and if $Z_i = 1$ and $S^{obs} = 1$ (alive), then $\mathcal{P}(Z_i, S_i^{obs}) = \{\text{protectable, always survivor}\}$. Then the likelihood of the collection of data

$$X_i, Z_i, S_i^{obs}, \text{ and } A_i \text{ if } S_i^{obs} = 1$$

over independent individuals, conditional on the covariates and the observed factor levels, is

$$\text{Likd}(\beta^{(\text{P})}, \beta^{(\text{A})}) = \prod_i \sum_{q \in \mathcal{P}(Z_i, S_i^{obs})} l^{(\text{P})}(q, X_i, \beta^{(\text{P})}) \cdot \{l^{(\text{A})}(A_i, q, X_i, \beta^{(\text{A})})\}^{S_i^{obs}} \tag{7}$$

Under this setting, we can more generally express a quantity of interest as a function $Q(\beta^{(\text{P})}, \beta^{(\text{A})})$ of the parameters, which can then be estimated by using likelihood or Bayesian methods to estimate the parameters from (7). Semiparametric methods, as discussed by Scharfstein, Rotnitzky and Robins (1999) in general, and by Gilbert et al (2003) for an application of principal stratification to vaccine trials, are also of interest. The fact that these quantities would be identifiable by our method even without the models in (6) if samples were large enough means that the results should not be sensitive to the particular parametric models, as long as they are flexible. Moreover, we can also show better estimation of general quantities of importance in Table 1, such as for associations using case-crossover designs.

6. Related Problems.

The design and structure of principal stratification we proposed for this problem, "inputs missing due to death", has commonalities and also differences with the structure of two other

problems where studies assign a treatment to examine its effect on an outcome. The first problem, "treatment noncompliance", deals with subjects who do not comply with the assigned treatment, and its structure with principal strata was discussed by Imbens and Rubin (1997). The second problem, "outcomes censored by death" (see Sec. 1), deals with subjects who die before *the intended future outcome is measured, and its structure with principal strata has been discussed by Rubin (2000), Frangakis and Rubin (2002), Zhang and Rubin (2003), and, with adaptation to HIV vaccines, by Gilbert et al. (2003).*

The common structure across these problems is centered around a factor that can be thought of as controllable, in the sense that its assignment is assumed ignorable. All three problems also have factors whose values are measurable after the controllable factor is assigned, namely post-controllable (or endogenous) factors; and factors whose values are defined (but not necessarily measurable) before the controllable factor is assigned, namely pre-controllable factors. The latter include all potential *outcomes of the post-controllable factors, and, therefore, include principal strata, that is, cross classifications of subjects by some subset of potential outcomes. The three problems also have differences, in their structure, their goals, and in the role that principal stratification plays in addressing these goals.*

In the problem with "treament noncompliance", the controllable factor is the treatment assignment; the post-controllable factors are the observed treatment received and the outcome; and the pre-controllable factors are the potential values of the treatment received and of the outcome. Of particular importance is the principal stratum of "compliers", that is, the subjects for whom the potential values of treatment received are the same as the treatment assigned, for all assignment levels (Imbens and Rubin, 1997). In this problem, principal stratification helps to formulate and, under assumptions, estimate the effect of treatment assignment (or intention to treat, ITT) on the outcome among the compliers. This goal is important because for compliers, the experimental comparison of outcomes among the levels of the controlled assignment is also

a comparison among the different levels of treatment received.

In the problem with "outcomes censored by death", the controllable factor is again the treatment assignment; the post-controllable factors are the observed survival status, and the observed outcome if the person survives; and the pre-controllable factors are the potential values of the survival and of the outcome. Here, a principal stratum of particular importance is that of "always survivors", defined as in Section 2. Principal stratification helps formulate and estimate the effect of treatment assignment on the outcome among always survivors. This goal is important because always survivors are the only subjects for whom potential outcomes are well defined for all assignment levels.

In the problem with "inputs missing due to death", the controllable factor is one that affects survival after the critical event; the post controllable factors are the observed survival status of the person, which determines measurement (if alive) or no measurement (if dead) of the input that occurred before the critical event; and the pre-controllable factors are the inputs of interest and the principal strata of survival. Here, a principal stratum of particular importance is that of "protectables". In this problem, principal stratification provides the framework for appropriately positing assumptions, such as those of Section 2 or 3, that allow estimation of the distribution of the missing inputs for protectable patients. As discussed, this goal is important because it better characterizes the differences between observed and missing inputs, and helps better understand the role that the inputs have for predicting mortality from the critical event.

7. Discussion.

We proposed a framework for addressing data missing due to death by obtaining and using data and explicit assumptions about a treatment assignment mechanism that could cause missing values to become observed if different levels of the treatment had been assigned. Thus, although a relation between causal inference and missing data has been obvious since Neyman

(1923) and Rubin (1974, 1976, 1978), the proposed framework for data missing due to death emphasizes a particular order for understanding these concepts: causal inference with potential outcomes is not just a special case of missing data, but is *more* fundamental than missing data (see also Rubin, 1987; 2005). Specifically, in the framework we proposed, data can only be regarded as having a missing value if an explicit intervention can be proposed that would provide us with that value. This principle for missing data, therefore, follows the principle of quantum mechanics, by which a measurable value of a physical quantity is only defined in terms of an explicit intervention that can be applied in order to provide that value. This parallel of principles is also reflected in the parallel of primary elements of the two frameworks – the complex wave function in quantum mechanics, and the principal strata of potential outcomes in the proposed framework for missing data: these primary elements give rise to the observed data by specific rules, but the primary elements are not themselves directly observable, providing an additional dimension that empowers the frameworks to better explain observations.

The use of an intervention factor to address missing data has the limitation that there can be settings where such a factor can exist, but still not available in the design. This can be so especially since such factors are not, at present systematically recorded for the purposes of addressing missing data, because their role in this problem had not previously been demonstrated. For such cases were the missing values are well defined but where design features do not allow their identifiability, sensitivity analyses can be implemented (e.g., Rubin, 1977; Manski 2003). Our results and illustration, though, demonstrate that using such intervention factors can improve the evaluation of and utility of studies with missing data due to death, and so can be the first step to a more systematic recording of such factors.

It will also be of interest to combine the setting discussed here, where possible deaths of patients can imply that their unobserved past is different from pasts that are observed, with the settings considered by Rubin (2000) and Zhang and Rubin (2003). In those settings, patients

who die could have had also a different *future* outcome trajectory from observed trajectories, under conditions that would have prevented their death. Developing methods to answer such combined questions is important for evaluating, for example, not only the potential benefit of prevention programs for saving lives, but also the programs' effects on the quality of patients' lives, and the relation of these effects to past input variables.

References.

[1] Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables (with Discussion). *Journal of the American Statistical Association* **91**, 444–472.

[2] Cohen HJ et al. (2002). A controlled trial of inpatient and outpatient geriatric evaluation and management. *New England Journal of Medicine* **346**, 905-912.

[3] Cornoni-Huntley J. et al. (1993). Established populations for epidemiologic studies of the Elderly: study design and methodology. *Aging (Milano)* **5**, 27–37.

[4] Cox, D. R. (1992). Causality: Some Statistical Aspects. *Journal of the Royal Statistical Society*, A **155**, part 2, 291–301.

[5] Fairclough, DL, Peterson HF, and Chag, V (1998). Why are missing quality of life data a problem in clinical trials of cancer therapy ? *Statistics in Medicine* **17**, 667–677.

[6] Frangakis, CE, and Rubin, DB (2002). Principal stratification in causal inference. *Biometrics*, **58**, 21–29.

[7] Gilbert, P. B., Bosch, R. J., and M. G. Hudgens (2003). Sensitivity analysis for the assessment of causal vaccine effects on viral load in AIDS vaccine trials. *Biometrics*, **59**, 531–541.

[8] GISSI. Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico. *The Lancet* **1**(8478), 397–402.

[9] Imbens, G. W. and Rubin, D. B. (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics* **25**, 305–327.

[10] Kurland, BF and Heagerty PJ. (2005). Directly parametrized regression conditioning on being alive: analysis of longitudinal data truncated by deaths. *Biostatistics* **6**, 241-258.

[11] Lin H, MCCulloch, CE and Mayne ST (2002). Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables *Statistics in Medicine* **21**, 2369-2382.

[12] MacKenzie, EJ, Rivara, FP, Jurkovich, GJ, Nathens, AB, Frey, KP, Egleston, BL, Salkever, DS, and Scharfstein DO. (2006). A national evaluation of the effect of Trauma-Center Care of Mortality. *New England Journal of Medicine* **354**, 366–378.

[13] Maclure, M. (1991). The case-crossover design: a method for studying transient effects on the risk of acute events. *American Journal of Epidemiology* **133**, 144–153.

[14] MacMahon RP. and Harrell Jr FE (2001). Joint testing of mortality and a non-fatal outcome in clinical trials. *Statistics in Medicine* **20**, 1165-1172.

[15] Manski, C. F. (2003). *Partial identification of probability distributions.* New York: Springer.

[16] Neyman, J. (1923). On the application of probability theory to agricultural experiments: essay on principles, Section 9. Translated in *Statistical Science* **5**, 465–80, 1990.

[17] Prentice et al. (1978). The analysis of failure times in the presence of competing risks. *Biometrics* **34**, 541-554.

[18] Reuben D., Borok G., Wolde–Tsadik G., Ershoff D., Fishman L., Ambrosini V., Liu Y., Rubenstein L., and Beck J. (1995). Randomized trial of comprehensive geriatric assessment in the care of hospitalized patients. *New England Journal of Medicine* **332**, 1345–1350.

[19] Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66**, 688–701.

[20] Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–92.

[21] Rubin, D. B. (1977). Formalizing subjective notions about the effects of nonrespondents in sample surveys. *Journal of the American Statistical Association* **72**, 538–543.

[22] Rubin, D. B. (1978). Bayesian inference for causal effects. *Annals of Statistics* **6**, 34–58.

[23] Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

[24] Rubin, D. B. (2000). Comment on "Causal inference without counterfactuals", by AP Dawid, *Journal of the American Statistical Association* **95**, 435–437.

[25] Rubin, DB. (2005). Causal Inference Using Potential Outcomes: Design, Modeling, Decisions. *Journal of the American Statistical Association* **469**, 322–331.

[26] Scharfstein, DO, Rotnitzky, A, and Robins, JM. (1999). Adjusting for Nonignorable Drop-out Using Semiparametric Nonresponse Models (with discussion). *Journal of the American Statistical Association* **94**, 1096–1146.

[27] Vinson DC, Mabe N, Leonard LL, Alexander J, Becker J, Boyer J, Moll J. (1995). Alcohol and injury. A case-crossover study. *Archives of family medicine* **4**, 505–511.

[28] Zhang, JL and Rubin DB (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *Journal of Educational and Behavioral Statistics* **28**, 353-368.

Table 1: Examples of studies with input data missing due to death.

population; original goal	measures of interest (time 0)	critical event (time 1)
elders or sick; relate functional measures to mortality	activities of daily living (ADL), intense emotional stress, intense physical activity,	stroke, falls, myocardial infarction, opportunistic infections
youths; relate exposure measures to severe injury/mortality	controlled substance use (e.g, alcohol, drug abuse)	injuries (e.g., crash)

Table 2: Demonstration of design using injury data from NSCOT.

	Data				Results based on new design and method	
injury severity X_i	% transported quickly $\text{pr}(Z_i = 1)$	survival (%) when transported slowly $\text{pr}(S_i^{obs} = 1 \mid Z_i = 0)$	input ADL when transported quickly $\text{pr}(A_i = 1 \mid Z_i = 1)$	input ADL of survivors when transported slowly $\text{pr}(A_i = 1 \mid S_i^{obs} = 1, Z_i = 0)$	**input ADL of those who died when transported slowly** $\text{pr}(A_i = 1 \mid S_i^{obs} = 0, Z_i = 0)$	**Relative Risk of death as in expr (4)**
low	2%	95%	25%	22%	82%	13.7
high	8%	81%	9%	5%	26%	3.6

Standard methods would assume these the *same*

Standard methods would assume these = 1

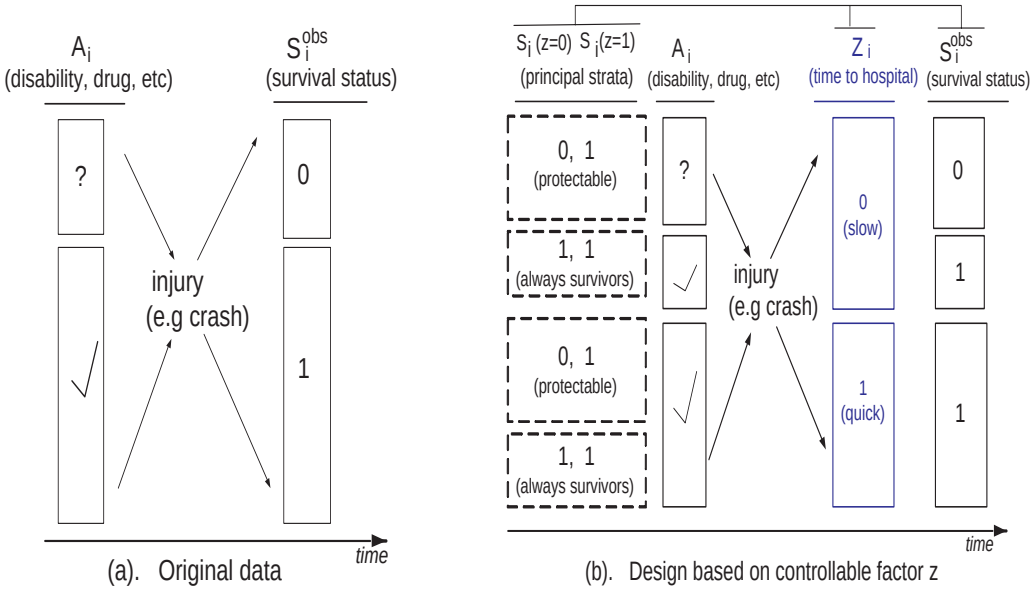

Figure 1.

(a): Initial design on input variable A and survival status S^{obs}, matched for past covariates;

(b): New design based on a controllable factor. Dashed boxes indicate principal strata with respect to survival. The presentational order from left to right of (principal strata $(S_i(0), S_i(1))$ and input A_i), controllable factor Z_i, and observed survival S_i^{obs}, which determines measurement or no measurement of the input A_i, is also the time order of definition from earliest to latest variable. Other covariates defined before the controllable factor can be used as in Sec. 5.2.