doi: 10.1111/j.1467-9469.2006.00542.x
 Published by Blackwell Publishing Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA Vol 34: 137–154, 2007

# Cohort Sampling Schemes for the Mantel–Haenszel Estimator

LARRY GOLDSTEIN

*Department of Mathematics, University of Southern California*

BRYAN LANGHOLZ

*Department of Preventive Medicine, University of Southern California*

**ABSTRACT. In many epidemiological studies, disease occurrences and their rates are naturally modelled by counting processes and their intensities, allowing an analysis based on martingale methods. Applied to the Mantel–Haenszel estimator, these methods lend themselves to the analysis of general control selection sampling designs and the accommodation of time-varying exposures.**

*Key words:* baseline hazard estimation, case-control studies, counter matching, counting process, epidemiology, survival analysis

## 1. Introduction

Mantel–Haenszel estimators (Mantel & Haenszel, 1959; Breslow & Day, 1980) have long been used in medical research to quantify the relative risk of disease between two groups. Historically, the long-standing appeal of the classical Mantel–Haenszel estimators is that they have a simple closed form formula which does not require the solution of an estimating equation. Although maximum likelihood estimation has become the dominant approach to the analysis of epidemiological data, the Mantel–Haenszel estimator continues to be popular. A medline search of papers in the years 2000–2005 gives a total of 420 references where Mantel–Haenszel is cited in the abstract as the method applied. An excellent review of the development of the Mantel–Haenszel estimator for analysis of epidemiological case-control studies, as well as the prominent role it has played in epidemiological research generally, is given in Breslow (1996).

In a cohort $\mathcal{R}=\{1,\ldots,n\}$ of individuals followed over the time interval $[0,\tau], 0<\tau\leq\infty$, a model with minimal assumptions which relates failure and a binary exposure is that the failure rate for exposed individuals $i\in\mathcal{R}$ is increased by an unknown factor $\phi_0\in(0,\infty)$ over the failure rate for those unexposed. Allowing additional flexibility by leaving the common baseline hazard function $\lambda_0(t)$ unspecified, and also letting the binary exposure indicator variable $Z_i(t)$ of individual $i$ depend on time, taking the value 0 or 1 when $i$ is unexposed or exposed at time $t$, respectively, results in the rate of observed failure at time $t$ for $i$ of

$$\lambda_i(t)=Y_i(t)\lambda_0(t)\phi_0^{Z_i(t)}, \tag{1}$$

where $Y_i(t)$ is the censoring indicator for individual $i$, taking the value 1 when individual $i$ is observed at time just prior to $t$. At any time $t$, we can divide the collection of individuals at risk at time $t$, $\mathcal{R}(t)=\{i: Y_i(t)=1\}$ of size $n(t)=|\mathcal{R}(t)|$, into the two groups, $\mathcal{R}_k(t)=\{i\in\mathcal{R}(t): Z_i(t)=k\}$ of sizes $n_k(t)=|\mathcal{R}_k(t)|$, $k=0,1$.

Turning for the moment to the fixed time covariate case, letting $t_{1,j}<t_{2,j}<\cdots$ be the collection of all failure times among individuals having exposure $j$, with $R_{jk}=\sum_{l\geq 1} n_k(t_{l,j})/n(t_{l,j})$, the Mantel–Haenszel estimator is given explicitly by

$$\hat{\phi}_{\mathrm{MH}} = \frac{R_{10}}{R_{01}}, \tag{2}$$

and in particular, there is no need to solve an estimating equation. In this setting, $\hat{\phi}_{\mathrm{MH}}$ is a consistent and asymptotically normal estimate of $\phi_0$ (Robins *et al.*, 1986).

In this paper, we consider Mantel–Haenszel estimators for nested case-control studies in which controls are sampled from risk sets determined by the cohort failure times (see e.g. Langholz & Goldstein, 1996). In recent work, Zhang *et al.* (2000) defined generalized Mantel–Haenszel estimators when controls are a simple random sample from the risk set and derived the properties of the estimator for right censored cohort data. Further, Zhang (2000) developed estimators for a number of methods of sampling controls including sampling with and without replacement and geometric sampling, and showed their consistency.

By placing our models in the counting process framework as detailed in sections 3 and 4.1, we expand on the work of these authors by providing Mantel–Haenszel estimators for the entire class of control sampling methods considered by Borgan *et al.* (1995). The estimators $\hat{\phi}_{\mathrm{MH}}$ proposed for sampling, generalizing (2), continue to have a 'closed form' representation which does not require solving a nonlinear estimating equation. Although our main focus is on the use of the Mantel–Haenszel estimator under sampling, our setting also allows us to extend its scope to accommodate the time-varying intensities (1). In sections 4.2 and 4.3, we show the consistency and asymptotic normality of the Mantel–Haenszel and baseline hazard estimator (14) under very general conditions, and in section 5 apply the asymptotic theory and determine the limiting distributions under random sampling, matching and counter matching, and make efficiency comparisons against $\hat{\phi}_{\mathrm{MPL}}$, the maximum partial likelihood estimator (MPLE). It is well known that in the full-cohort setting $\hat{\phi}_{\mathrm{MH}}$ performs asymptotically as well as $\hat{\phi}_{\mathrm{MPL}}$ at the null $\phi_0 = 1$. In section 5, we show that $\hat{\phi}_{\mathrm{MH}}$ continues to have this feature under sampling.

## 2. Risk set sampling

### *2.1. General framework*

We consider a cohort with failures driven by the intensity (1), where the binary exposure status histories are not available for all cohort subjects but can be ascertained for a sample in a nested case-control study. In particular, if subject $i$ fails at time $t$, we consider designs in which the *case* $i$ is always included in the sample, and *controls* are chosen from $\mathcal{R}(t) \setminus \{i\}$, those other individuals in the risk set at time $t$. Control sampling can be specified by giving for all $t$ and $(i, \mathbf{r})$ with $i \in \mathbf{r}$ and $\mathbf{r} \subset \mathcal{R}$ a collection of probabilities $\pi_t(\mathbf{r}|i)$ for choosing the individuals in the set $\mathbf{r} \subset \mathcal{R}(t)$ to serve as controls should $i$ fail at time $t$. Exposure status is then ascertained for the members of the set $\mathbf{r}$ so selected, the *sampled risk set*, made up of the controls $\mathbf{r} \setminus \{i\}$ along with the failure $i$. For convenience, we set $\pi_t(\mathbf{r}|i) = 0$ when $i \notin \mathbf{r}$ or $Y_i(t) = 0$.

The added flexibility and efficiency gains made possible by the choice of design $\pi_t(\mathbf{r}|i)$ is substantial, opening up the possibility of using sampling designs that can take advantage of additional structure which may be available in the data. For example, in designs 3 and 4, the matching and counter-matching designs described below, we assume that for each $i \in \mathcal{R}(t)$ we have available the value $C_i(t)$ giving the strata membership of $i$ among the possible values in $\mathcal{C}$, some (small) finite set. For $l \in \mathcal{C}$ we let $\mathcal{C}_l(t) = \{i : Y_i(t) = 1, C_i(t) = l\}$ and $c_l(t) = |\mathcal{C}_l(t)|$, the $l^{\mathrm{th}}$ sampling stratum, and its size, at time $t$.

Each design $\pi_t(\mathbf{r}|i)$ has an associated probability distribution on the subsets of $\mathcal{R}$ defined by

$$\pi_t(\mathbf{r}) = n(t)^{-1} \sum_{i \in \mathbf{r}} \pi_t(\mathbf{r}\,|\,i), \tag{3}$$

which sums to one by virtue of

$$\sum_{\mathbf{r} \subset \mathcal{R}} \sum_{i \in \mathbf{r}} \pi_t(\mathbf{r}\,|\,i) = \sum_{i \in \mathcal{R}} \sum_{\mathbf{r} \subset \mathcal{R},\, \mathbf{r} \ni i} \pi_t(\mathbf{r}\,|\,i) = \sum_{i \in \mathcal{R}} Y_i(t) = n(t). \tag{4}$$

In addition, we define the associated weights $w_i(t, \mathbf{r})$, set to 0 when $i$ is not at risk, by

$$w_i(t, \mathbf{r}) = \frac{\pi_t(\mathbf{r}\,|\,i)}{n(t)^{-1} \sum_{l \in \mathbf{r}} \pi_t(\mathbf{r}\,|\,l)}, \quad \text{so that} \quad \pi_t(\mathbf{r}\,|\,i) = \pi_t(\mathbf{r}) w_i(t, \mathbf{r}). \tag{5}$$

### *2.2. Some specific designs*

The sampling framework accommodates a wide range of designs (e.g. Borgan *et al.*, 1995; Borgan & Langholz, 1998; Andrieu *et al.*, 2000; Langholz & Goldstein, 2001). Here, we highlight a few:

*Design 1: the full cohort.* When information on all subjects is available, we may take $\pi_t(\mathbf{r}\,|\,i)$ to be the indicator of the set of those at risk at time $t$ $\pi_t(\mathbf{r}\,|\,i) = \mathbf{1}(\mathbf{r} = \mathcal{R}(t))$ and so $w_i(t, \mathbf{r}) = \mathbf{1}(i \in \mathcal{R}(t), \mathbf{r} = \mathcal{R}(t))$.
The framework under which the classical Mantel–Haenszel estimator can be applied is recovered under this scheme when the exposures are time fixed.

When the collection of exposure status data on the full cohort is impractical and no additional information on cohort members is available, the simple random sampling design is a natural choice:

*Design 2: simple random sampling.* At each failure time, a simple random sample of $m-1$ individuals is chosen from those at risk to serve as controls for the failure; for $i \in \mathbf{r} \subset \mathcal{R}(t)$ with $|\mathbf{r}| = m$,

$$\pi_t(\mathbf{r}\,|\,i) = \binom{n(t)-1}{m-1}^{-1}, \tag{6}$$

and the probabilities (3) and weights (5) are given by

$$\pi_t(\mathbf{r}) = \binom{n(t)}{m}^{-1} \quad \text{and} \quad w_i(t, \mathbf{r}) = \frac{n(t)}{m}.$$

The matching design could be used to control for confounding by stratifying by a potential confounder.

*Design 3: simple random sampling within matching strata, with specification* $\mathbf{m} = (m_l)_{l \in \mathcal{C}}$, $m_l \geq 1$. If subject $i$ fails at time $t$, then a simple random sample of $m_{C_i(t)} - 1$ controls are chosen from $\mathcal{C}_{C_i(t)}(t)$, the failure's stratum at time $t$, to serve as controls for the failure. Hence, the sampling probabilities of this scheme are given by

$$\pi_t(\mathbf{r}|i) = \binom{c_{C_i(t)}(t)-1}{m_{C_i(t)}-1}^{-1} \mathbf{1}(\mathbf{r} \subset \mathcal{C}_{C_i(t)}(t), \mathbf{r} \ni i, |\mathbf{r}| = m_{C_i(t)}), \tag{7}$$

and for $\mathbf{r} \subset \mathcal{C}_l(t)$, $|\mathbf{r}| = m_l$ and $i \in \mathbf{r}$, the probabilities (3) and weights (5) are given by

$$\pi_t(\mathbf{r}) = \frac{c_l(t)}{n(t)} \binom{c_l(t)}{m_l}^{-1} \quad \text{and} \quad w_i(t, \mathbf{r}) = \frac{n(t)}{m_l}.$$

In section 5, we show that significant efficiency gains over random sampling can be achieved by the counter-matching design when the strata $\mathcal{C}$ are sufficiently correlated to exposure.

*Design 4: counter matching, with specification* $\mathbf{m}=(m_l)_{l\in\mathcal{C}}$, $m_l \geq 1$. If subject $i$ fails at time $t$, then $m_l$ controls are randomly sampled without replacement from each $\mathcal{C}_l(t)$ except for the failure's stratum, from which $m_{C_i(t)}-1$ controls are sampled. Let $\mathcal{P}_{\mathcal{C}}(t)$ denote the set of all subsets of $\mathcal{R}(t)$ with $m_l$ individuals of type $l$ for all $l\in\mathcal{C}$. Then for $\mathbf{r}\in\mathcal{P}_{\mathcal{C}}(t)$ and $i\in\mathbf{r}$, the sampling probabilities of this scheme are given by

$$\pi_t(\mathbf{r}|i)=\left[\prod_{l\in\mathcal{C}}\binom{c_l(t)}{m_l}\right]^{-1}\frac{c_{C_i(t)}(t)}{m_{C_i(t)}}, \tag{8}$$

and the probabilities (3) and weights (5) are given by

$$\pi_t(\mathbf{r})=\left[\prod_{l\in\mathcal{C}}\binom{c_l(t)}{m_l}\right]^{-1} \quad \text{and} \quad w_i(t,\mathbf{r})=c_{C_i(t)}(t)/m_{C_i(t)}.$$

## 3. Mantel–Haenszel estimators for sampled risk set data

Let $N_{i,\mathbf{r}}(t)$ be the counting process that records the number of times in $(0,t]$ that $i$ fails and $\mathbf{r}$ is chosen as its sampled risk set and define

$$N^k_{\mathbf{r}}(t)=\sum_{i\in\mathcal{R}_k(t)}N_{i,\mathbf{r}}(t) \quad \text{and} \quad N_{\mathbf{r}}(t)=\sum_{i\in\mathbf{r}}N_{i,\mathbf{r}}(t), \tag{9}$$

recording, respectively, the number of times in $(0,t]$ that $\mathbf{r}$ was chosen as the sampled risk set for a failure in $\mathcal{R}_k(t)$, that is, one having exposure $k$, and the total number of times in $(0,t]$ that $\mathbf{r}$ was chosen as the sampled risk.

Now let

$$A^k_{\mathbf{r}}(t)=\sum_{i\in\mathcal{R}_k(t)}\pi_t(\mathbf{r}\,|\,i)=\pi_t(\mathbf{r})\sum_{i\in\mathcal{R}_k(t)}w_i(t,\mathbf{r}), \quad k=0,1 \tag{10}$$

and $A_{\mathbf{r}}(t)=A^0_{\mathbf{r}}(t)+A^1_{\mathbf{r}}(t)$. By (3), for sets $\mathbf{r}$ with $\pi_t(\mathbf{r})\neq 0$ we have $A_{\mathbf{r}}(t)=n(t)\pi_t(\mathbf{r})$ and

$$\frac{A^k_{\mathbf{r}}(t)}{A_{\mathbf{r}}(t)}\leq 1 \quad \text{and} \quad \sum_{\mathbf{r}\subset\mathcal{R},\,k\in\{0,1\}}A^k_{\mathbf{r}}(t)=n(t). \tag{11}$$

Now for $j,k=0,1$ set

$$R_{jk}(t)=\int_0^t\sum_{\mathbf{r}\subset\mathcal{R}}A^{-1}_{\mathbf{r}}(s)A^k_{\mathbf{r}}(s)\,\mathrm{d}N^j_{\mathbf{r}}(s) \quad \text{and} \quad R_{jk}=R_{jk}(\tau). \tag{12}$$

The Mantel–Haenszel estimator of $\phi_0$ in this more general context is then given by

$$\hat{\phi}_{\mathrm{MH}}=\frac{R_{10}}{R_{01}},$$

i.e. with the definitions of $R_{jk}$ extended as in (12), the estimator has exactly the same form (2) as before. We also consider the variance estimator

$$\hat{\sigma}^2=\hat{\phi}_{\mathrm{MH}}\frac{\int_0^\tau\sum_{\mathbf{r}\subset\mathcal{R}}A^{-2}_{\mathbf{r}}(s)A^0_{\mathbf{r}}(s)A^1_{\mathbf{r}}(s)\,\mathrm{d}N_{\mathbf{r}}(s)}{\left(\int_0^\tau\sum_{\mathbf{r}\subset\mathcal{R}}A^{-1}_{\mathbf{r}}(s)\frac{A^0_{\mathbf{r}}(s)A^1_{\mathbf{r}}(s)}{A^0_{\mathbf{r}}(s)+\hat{\phi}_{\mathrm{MH}}A^1_{\mathbf{r}}(s)}\,\mathrm{d}N_{\mathbf{r}}(s)\right)^2}. \tag{13}$$

In theorems 1 and 2 we give conditions under which $\hat{\phi}_{\mathrm{MH}}$ is consistent and asymptotically normal, and show that $n\hat{\sigma}^2$ is consistent for the variance of the asymptotic distribution. We note that for designs 1 and 2, $\hat{\phi}_{\mathrm{MH}}$ reduces to the 'classical' Mantel–Haenszel estimator and

$\hat{\sigma}^2$ is the previously described 'conditional variance' estimator (Breslow, 1981; Robins *et al.*, 1986).

Where estimates of $\phi_0$ can be used to assess the magnitude of the effect that exposure has on failure, estimates of the integrated baseline hazard

$$\Lambda_0(t) = \int_0^t \lambda_0(s)\,\mathrm{d}s$$

can in turn be used to provide estimates of absolute risk. We consider the integrated baseline hazard function estimate

$$\hat{\Lambda}_n(t, \hat{\phi}_{\mathrm{MH}}) = \int_0^t \sum_{\mathbf{r}\subset\mathcal{R}} \frac{\mathrm{d}N_{\mathbf{r}}(s)}{\sum_{i\in\mathbf{r}} \hat{\phi}_{\mathrm{MH}}^{Z_i(s)} w_i(s,\mathbf{r})}, \tag{14}$$

given in terms of the weights defined in (5), where the ratio in the integral is regarded as 0 if there is no one at risk. In theorem 3, we give conditions under which

$$\sqrt{n}\left(\hat{\Lambda}_n(\cdot, \hat{\phi}_{\mathrm{MH}}) - \Lambda_0(\cdot)\right)$$

converges weakly as $n \to \infty$ to a mean-zero Gaussian process, and provide a uniformly consistent estimator for its variance function.

## 4. Properties of the Mantel–Haenszel estimators

### *4.1. The counting process model for sampling*

Much of the analysis here follows the work of Borgan *et al.* (1995) closely, which is hereafter referred to as BGL. Assume that the censoring and failure information are defined on a probability space with a standard filtration $\mathcal{F}_t$, and that the censoring indicators $Y_i(t)$, exposures $Z_i(t)$, design $\pi_t(\mathbf{r}\,|\,i)$ and strata variables $C_i(t)$ are left continuous and adapted, and hence predictable and locally bounded. We make the assumption of independent sampling as in BGL that the intensity processes with respect to the filtration $\mathcal{F}_t$ is the same as that with respect to this filtration augmented with the sampling information, that is, that selecting an individual as a control does not influence the likelihood of failure for that individual in the future.

Combining (1) with the design probabilities, into which censoring has already been incorporated, we see $N_{i,\mathbf{r}}(t)$ has intensity of the form

$$\lambda_{i,\mathbf{r}}(t) = \phi_0^{Z_i(t)} \pi_t(\mathbf{r}\,|\,i)\lambda_0(t), \tag{15}$$

and subtracting the integrated intensity results in

$$M_{i,\mathbf{r}}(t) = N_{i,\mathbf{r}}(t) - \int_0^t \lambda_{i,\mathbf{r}}(s)\,\mathrm{d}s,$$

orthogonal local square integrable martingales with predictable quadratic variation

$$\mathrm{d}\langle M_{i,\mathbf{r}}\rangle_t = \lambda_{i,\mathbf{r}}(t)\,\mathrm{d}t.$$

Similarly, with $A_{\mathbf{r}}^k(t)$ given in (10) for $k \in \{0,1\}$, by linearity the counting processes $N_{\mathbf{r}}^k(t)$ and $N_{\mathbf{r}}(t)$ defined in (9) give rise to the orthogonal local square integrable martingales

$$M_{\mathbf{r}}^k(t) = \sum_{i\in\mathcal{R}_k(t)} M_{i,\mathbf{r}}(t) \quad \text{and} \quad M_{\mathbf{r}}(t) = \sum_{i\in\mathbf{r}} M_{i,\mathbf{r}}(t)$$

with respective intensities

$$\lambda_{\mathbf{r}}^k(t) = \phi_0^k A_{\mathbf{r}}^k(t)\lambda_0(t) \quad \text{and} \quad \lambda_{\mathbf{r}}(t) = \left(A_{\mathbf{r}}^0(t) + \phi_0 A_{\mathbf{r}}^1(t)\right)\lambda_0(t), \tag{16}$$

and predictable quadratic variations

$$\mathrm{d}\langle M^k_{\mathbf{r}}\rangle_t = \lambda^k_{\mathbf{r}}(t)\,\mathrm{d}t \quad \text{and} \quad \mathrm{d}\langle M_{\mathbf{r}}\rangle_t = \lambda_{\mathbf{r}}(t)\,\mathrm{d}t. \tag{17}$$

For $\mathbf{v}$ a multisubset of $\{0,1\}$ of size 2 or 3, e.g. $\mathbf{v}=\{0,1,1\}$, define

$$H_{\mathbf{v}}(t) = \sum_{\mathbf{r}\subset\mathcal{R}} A^{1-|\mathbf{v}|}_{\mathbf{r}}(t)\prod_{k\in\mathbf{v}} A^k_{\mathbf{r}}(t). \tag{18}$$

Then, for $j,k\in\{0,1\}$, the processes

$$W_{jk}(t) = \int_0^t \sum_{\mathbf{r}\subset\mathcal{R}} A^{-1}_{\mathbf{r}}(s)A^k_{\mathbf{r}}(s)\,\mathrm{d}M^j_{\mathbf{r}}(s) \tag{19}$$

are local square integrable martingales with predictable quadratic variation

$$\mathrm{d}\langle W_{jk}, W_{pq}\rangle_t = \mathbf{1}_{(j=p)}\phi^j_0 H_{jkq}(t)\lambda_0(t)\,\mathrm{d}t. \tag{20}$$

Using $M^j_{\mathbf{r}}(t) = N^j_{\mathbf{r}}(t) - \lambda^j_{\mathbf{r}}(t)$, (16) and (18), $R_{jk}(t)$ in (12) may be written

$$R_{jk}(t) = \phi^j_0\int_0^t H_{jk}(s)\lambda_0(s)\,\mathrm{d}s + W_{jk}(t). \tag{21}$$

As $H_{01}(t) = H_{10}(t)$,

$$G(t) = \phi_0 R_{01}(t) - R_{10}(t) = \phi_0 W_{01}(t) - W_{10}(t) \tag{22}$$

is a local square integrable martingale and, by (20), has quadratic variation

$$\mathrm{d}\langle G\rangle_t = \left(\phi^2_0 H_{011}(t) + \phi_0 H_{100}(t)\right)\lambda_0(t)\,\mathrm{d}t = \phi_0 A^{-2}_{\mathbf{r}}(t)A^0_{\mathbf{r}}(t)A^1_{\mathbf{r}}(t)\lambda_{\mathbf{r}}(t)\,\mathrm{d}t.$$

### *4.2. Asymptotics of* $\hat{\phi}_{\mathrm{MH}}$

We prove the consistency and asymptotic normality of $\hat{\phi}_{\mathrm{MH}}$ under some regularity and stability conditions.

*Condition 1.* The cumulative hazard on the interval $[0,\tau]$ is finite: $\Lambda_0(\tau) < \infty$.

*Condition 2.* For all multisubsets $\mathbf{v}$ of $\{0,1\}$ with $|\mathbf{v}|\in\{2,3\}$, there exist left continuous functions $h_{\mathbf{v}}(t)$ such that for almost all $t$ in $[0,\tau]$,

$$\frac{1}{n}H_{\mathbf{v}}(t) \to_p h_{\mathbf{v}}(t).$$

Under conditions 1 and 2 the dominated convergence theorem of Hjort & Pollard (1993) as in proposition 1 of BGL p. 1762, with dominating functions $D_n(t) = D(t) = 1$ by (11), yields

$$\int_0^t \frac{1}{n}H_{\mathbf{v}}(s)\lambda_0(s)\,\mathrm{d}s \to_p I_{\mathbf{v}}(t) \quad \text{where} \quad I_{\mathbf{v}}(t) = \int_0^t h_{\mathbf{v}}(s)\lambda_0(s)\,\mathrm{d}s. \tag{23}$$

**Proposition 1**
*Let conditions 1 and 2 hold. Then for every* $t\in[0,\tau]$,

$$n^{-1}\langle W_{jk}, W_{pq}\rangle_t \to_p \mathbf{1}_{(j=p)}\phi^j_0 I_{jkq}(t) \quad \textit{and} \quad n^{-1}R_{jk}(t) \to_p \phi^j_0 I_{jk}(t). \tag{24}$$

*Proof.* The first claim follows by (20) and (23). By (21), we have

$$\frac{1}{n}R_{jk}(t) = \phi^j_0\int_0^t \frac{1}{n}H_{jk}(s)\lambda_0(s)\,\mathrm{d}s + \frac{1}{n}W_{jk}(t).$$

By (23) the first term converges to $\phi_0^j I_{jk}(t)$. The second term converges to zero in probability by (20) and a standard argument using Lenglart's inequality (see Andersen *et al.*, 1993), and the second claim in (24) follows.

We now show the consistency of $\hat{\phi}_{\mathrm{MH}}$ upon additionally adopting

*Condition 3.* $I_{01}(\tau)$, given in (23), is strictly positive.

**Theorem 1**
*Under conditions 1–3, the estimate $\hat{\phi}_{MH}$ defined in (2) is consistent,*

$$\hat{\phi}_{\mathrm{MH}} \to_p \phi_0 \quad \text{as } n \to \infty.$$

*Proof.* Using proposition 1, and that $I_{10} = I_{01}$, we have

$$\hat{\phi}_{\mathrm{MH}} = \frac{n^{-1} R_{10}}{n^{-1} R_{01}} \to_p \frac{\phi_0 I_{10}}{I_{01}} = \phi_0.$$

**Lemma 1**
*Under conditions 1–3, the processes $\{n^{-1/2} W_{jk}(\cdot)\}$ given in (19) converge jointly in $D[0,\tau]$ to mean-zero Gaussian processes $\{w_{jk}(\cdot)\}$ with covariations*

$$\mathrm{d}\langle w_{jk}, w_{pq}\rangle_t = \mathbf{1}_{(j=p)} \phi_0^j h_{jkq}(t) \lambda_0(t)\, \mathrm{d}t,$$

*and hence $n^{-1/2} G(t) = n^{-1/2}(\phi_0 W_{01}(t) - W_{10}(t))$ in (22) converges in $D[0,\tau]$ to the mean-zero Gaussian process $g(\cdot)$ with*

$$\langle g\rangle_t = \int_0^t \left(\phi_0^2 h_{011}(s) + \phi_0 h_{100}(s)\right) \lambda_0(s)\, \mathrm{d}s.$$

*Further, for $t \in [0,\tau]$, and any consistent sequence $\hat{\phi}_n \to_p \phi_0$, as $n \to \infty$,*

$$n^{-1} \hat{\phi}_n \int_0^t \sum_{\mathbf{r} \subset \mathcal{R}} A_{\mathbf{r}}^{-2}(s)\; A_{\mathbf{r}}^0(s)\; A_{\mathbf{r}}^1(s)\, \mathrm{d}N_{\mathbf{r}}(s) \to_p \langle g\rangle_t. \tag{25}$$

*Proof.* We apply the martingale central limit theorem of Rebolledo, as presented in theorem II.5.1 of Andersen *et al.* (1993). The processes $\{n^{-1/2} W_{jk}\}$ are local square integrable martingales, whose predictable quadratic variation converges by proposition 1 to the continuous functions given in (24). Regarding the Lindeberg Condition, using (11) for the two final inequalities,

$$\begin{aligned}
&\frac{1}{n} \int_0^\tau \sum_{\mathbf{r} \subset \mathcal{R}} \left(\frac{A_{\mathbf{r}}^k(t)}{A_{\mathbf{r}}(t)}\right)^2 \mathbf{1}\left(n^{-1/2} \left|\frac{A_{\mathbf{r}}^k(t)}{A_{\mathbf{r}}(t)}\right| > \epsilon\right) \lambda_{\mathbf{r}}^j(t)\, \mathrm{d}t \\
&\quad \le \frac{\phi_0^j}{\epsilon^\delta n^{1+\delta/2}} \int_0^\tau \sum_{\mathbf{r} \subset \mathcal{R}} \left|\frac{A_{\mathbf{r}}^k(t)}{A_{\mathbf{r}}(t)}\right|^{2+\delta} A_{\mathbf{r}}^j(t) \lambda_0(t)\, \mathrm{d}t \\
&\quad \le \frac{\phi_0^j}{\epsilon^\delta n^{1+\delta/2}} \int_0^\tau \sum_{\mathbf{r} \subset \mathcal{R}} A_{\mathbf{r}}^j(t) \lambda_0(t)\, \mathrm{d}t \le \frac{\phi_0^j}{\epsilon^\delta n^{\delta/2}} \Lambda_0(\tau) \to_p 0.
\end{aligned}$$

For all $t \in [0,\tau]$ and $i,j,k \in \{0,1\}$, by Rebolledo's theorem as in theorem II.5.1 of Andersen *et al.* (1993), the scaled optional variation

$$n^{-1}[W_{ik}, W_{ij}]_t = \frac{1}{n}\int_0^t \sum_{\mathbf{r}\subset\mathcal{R}} A_{\mathbf{r}}^{-2}(s)A_{\mathbf{r}}^k(s)A_{\mathbf{r}}^j(s)\,\mathrm{d}N_{\mathbf{r}}^i(s)$$

converges to the same limit (24), as that of the scaled predictable variation. The convergence in (25) now follows.

**Theorem 2**
*Under conditions 1–3, with* $\hat{\phi}_{\mathrm{MH}}$ *given in (2),*

$$\sqrt{n}\left(\hat{\phi}_{\mathrm{MH}} - \phi_0\right) \to_d \mathcal{N}(0, \sigma^2), \tag{26}$$

*where*

$$\sigma^2 = \frac{\int_0^\tau \left(\phi_0^2 h_{011}(t) + \phi_0 h_{100}(t)\right)\lambda_0(t)\,\mathrm{d}t}{\left(\int_0^\tau h_{01}(t)\lambda_0(t)\,\mathrm{d}t\right)^2}, \tag{27}$$

*which can be consistently estimated by* $n\hat{\sigma}^2$ *given in (13).*

*Proof.* As

$$\sqrt{n}\left(\hat{\phi}_{\mathrm{MH}} - \phi_0\right) = \frac{n^{-1/2}(R_{10} - \phi_0 R_{01})}{n^{-1}R_{01}} = -\frac{n^{-1/2}G(\tau)}{n^{-1}R_{01}},$$

(26) follows using proposition 1 and lemma 1. The consistency of $n\hat{\sigma}^2$ follows from (25) for the numerator, and the consistency of $\hat{\phi}_{\mathrm{MH}}$ and (24) for the denominator.

Under the null $\phi_0 = 1$, we note that as $h_{001}(t) + h_{011}(t) = h_{01}(t)$, the asymptotic variance (27) simplifies to

$$\sigma^2 = \frac{1}{\int_0^\tau h_{01}(t)\lambda_0(t)\,\mathrm{d}t}. \tag{28}$$

### 4.3. Properties of the baseline hazard estimator

To study the estimate (14) we impose the following additional conditions.

*Condition 4.* The ratio $n(t)/n$ is uniformly bounded away from zero in probability as $n \to \infty$.

*Condition 5.* There exist functions $e$ and $\psi$ such that for all $t \in [0, \tau]$ as $n \to \infty$,

$$\sum_{\mathbf{r}\subset\mathcal{R}} \pi_t(\mathbf{r})\left\{\frac{A_{\mathbf{r}}^1(t)}{A_{\mathbf{r}}^0(t) + \phi_0 A_{\mathbf{r}}^1(t)}\right\} \to_p e(\phi_0, t), \tag{29}$$

and

$$n\sum_{\mathbf{r}\subset\mathcal{R}} \pi_t(\mathbf{r})^2\{A_{\mathbf{r}}^0(t) + \phi_0 A_{\mathbf{r}}^1(t)\}^{-1} \to_p \psi(\phi_0, t). \tag{30}$$

Letting $t_1 < t_2 < \cdots$ be the collection of all failure times, and $\tilde{\mathcal{R}}_j$ the sampled risk set at failure time $t_j$, we rewrite the cumulative baseline hazard estimate (14) as

$$\hat{\Lambda}_n(t, \hat{\phi}_{\mathrm{MH}}) = \sum_{t_j \le t} \frac{1}{\sum_{i\in\tilde{\mathcal{R}}_j} \hat{\phi}_{\mathrm{MH}}^{Z_i(t_j)} w_i(t_j, \tilde{\mathcal{R}}_j)},$$

where the weights $w_i(t, \mathbf{r})$ are given in (5).

**Theorem 3**

*Let conditions 1–5 hold, and with $e(\phi_0, u)$ as in (29) set*

$$B(t, \phi_0) = \int_0^t e(\phi_0, u)\lambda_0(u)\,\mathrm{d}u.$$

*Then $n^{1/2}(\hat{\phi}_{\mathrm{MH}} - \phi_0)$ and the process*

$$X_n(\cdot) = n^{1/2}\left(\hat{\Lambda}_n(\cdot, \hat{\phi}_{\mathrm{MH}}) - \Lambda_0(\cdot)\right) + n^{1/2}(\hat{\phi}_{\mathrm{MH}} - \phi_0)B(\cdot, \phi_0)$$

*are asymptotically independent. The limiting distribution of $X_n(\cdot)$ is, with $\psi(\phi_0, t)$ as in (30), that of a mean-zero Gaussian martingale with variance function*

$$\omega^2(t, \phi_0) = \int_0^t \psi(\phi_0, u)\lambda_0(u)\,\mathrm{d}u.$$

*In particular, the scaled difference between the estimated and true integrated baseline hazard*

$$\sqrt{n}\left(\hat{\Lambda}_n(\cdot, \hat{\phi}_{\mathrm{MH}}) - \Lambda_0(\cdot)\right)$$

*converges weakly as $n \to \infty$ to a mean-zero Gaussian process with covariance function*

$$\sigma^2_\Lambda(s, t) = \omega^2(s \wedge t) + B(s, \phi_0)\sigma^2 B(t, \phi_0).$$

*The function $\sigma^2_\Lambda(s, t)$ can be estimated uniformly consistently by $n\hat{\sigma}^2_\Lambda(s, t)$ where*

$$\hat{\sigma}^2_\Lambda(s, t) = \hat{\omega}^2(s \wedge t; \hat{\phi}_{\mathrm{MH}}) + \hat{B}_n(s; \hat{\phi}_{\mathrm{MH}})\hat{\sigma}^2\hat{B}_n(t; \hat{\phi}_{\mathrm{MH}}),$$

$$\hat{\omega}^2(t; \phi) = \sum_{t_j \le t} \frac{1}{\left\{\sum_{i \in \tilde{\mathcal{R}}_j} \phi^{Z_i(t_j)} w_i(t_j, \tilde{\mathcal{R}}_j)\right\}^2} \quad \textit{and}$$

$$\hat{B}_n(t, \phi) = \sum_{t_j \le t} \frac{\sum_{i \in \tilde{\mathcal{R}}_j} Z_i(t_j)\phi^{Z_i(t_j)-1} w_i(t_j, \tilde{\mathcal{R}}_j)}{\left\{\sum_{i \in \tilde{\mathcal{R}}_j} \phi^{Z_i(t_j)} w_i(t_j, \tilde{\mathcal{R}}_j)\right\}^2}.$$

*Proof.* The form of $\hat{\Lambda}_n$ is the same as in BGL, and noting in particular that Condition 4 in BGL can be satisfied by letting $X_{\mathbf{r}}(t) = 1$ and $D(t)$ a constant, we have that $X_n(\cdot)$ is asymptotically equivalent to the local square integrable martingale,

$$Y_n(\cdot) = n^{1/2} \int_0^{\cdot} \sum_{\mathbf{r} \subset \mathcal{R}} \frac{\mathrm{d}M_{\mathbf{r}}(u)}{\sum_{i \in \mathbf{r}} \phi_0^{Z_i(u)} w_i(u, \mathbf{r})},$$

and the proof of the claims made of the asymptotic distribution of $X_n$ now follow as there.

Regarding the asymptotic independence, for any locally bounded predictable processes $H_{\mathbf{r}}$, it is straightforward to verify

$$\langle \phi_0 W_{01} - W_{10}, \int_0^{\cdot} H_{\mathbf{r}}\,\mathrm{d}M_{\mathbf{r}} \rangle_t = 0.$$

Hence, by the asymptotic joint normality provided by Rebolledo's theorem II.5.1 in Andersen *et al.* (1993), functions of the collections $\{\int_0^{\cdot} H_{\mathbf{r}}\,\mathrm{d}M_{\mathbf{r}}\}_{\mathbf{r}}$ and $\phi_0 W_{01} - W_{10}$, in particular $\sqrt{n}(\hat{\phi}_{\mathrm{MH}} - \phi_0)$ and $X_n(\cdot)$, are asymptotically independent.

The claim that $\sigma^2_\Lambda(s, t)$ can be estimated uniformly consistent by $n\hat{\sigma}^2_\Lambda(s, t)$ follows as in BGL, based on the fact that $n\hat{\omega}^2(t, \phi_0)$ is the optional variation process of the local square integrable martingale $Y_n(\cdot)$, which by Rebolledo's theorem as cited above, converges uniformly in probability to its predictable variation $\omega^2(t, \phi_0)$; the uniform convergence of $\hat{B}_n(\cdot, \hat{\phi}_n)$ to $B(\cdot, \phi_0)$ is as in BGL, proposition 2.

## 5. Applications

In this section, we first show that for any design that meets the conditions in section 4.2, the Mantel–Haenszel estimator and the MPLE have the same asymptotic variance at the null, and away from the null that the MPL estimator is at least as efficient as the Mantel–Haenszel estimator. We then apply our results to the designs discussed in section 2.2. Although our asymptotic results hold under the weaker stability conditions of sections 4.2 and 4.3, here we assume that the censoring, covariate and strata variables are independent and identically distributed copies of $Y(t), Z(t)$, and $C(t)$, respectively, left continuous and adapted processes having right hand limits. The strata variable required for designs 3 and 4 gives the 'type' of individual among the possible values in a (small) finite set $\mathcal{C}$; the strata variable may be used to model any additional information, a surrogate of exposure in particular.

### 5.1. Relative efficiency of the Mantel–Haenszel estimator to the MPLE

We now show that at the null $\phi_0 = 1$, $\phi_{\mathrm{MH}}$ and $\phi_{\mathrm{MPL}}$ have equal efficiency. For $\mathbf{r} \subset \mathcal{R}$ let

$$S_{\mathbf{r}}^{(0)}(\phi, t) = A_{\mathbf{r}}^0(t) + \phi A_{\mathbf{r}}^1(t), \quad S_{\mathbf{r}}^{(1)}(\phi, t) = A_{\mathbf{r}}^1(t) \quad \text{and} \quad E_{\mathbf{r}}(\phi, t) = \frac{S_{\mathbf{r}}^{(1)}(\phi, t)}{S_{\mathbf{r}}^{(0)}(\phi, t)}. \tag{31}$$

Referring now to (3.4) of BGL (where $\beta = 0$ there corresponds to $\phi = 1$ here), we see $S_{\mathbf{r}}^{(2)}(1, t) = A_{\mathbf{r}}^1(t)$ as $Z^2 = Z$ when $Z \in \{0, 1\}$. In the null case, using (3.10) of BGL, the inverse variance of the MPLE is the integral of the baseline hazard against the limit of

$$\begin{aligned}
&\frac{1}{n} \sum_{r \subset \mathcal{R}} \left( \frac{S_{\mathbf{r}}^{(2)}(1, t)}{S_{\mathbf{r}}^{(0)}(1, t)} - \left( \frac{S_{\mathbf{r}}^{(1)}(1, t)}{S_{\mathbf{r}}^{(0)}(1, t)} \right)^2 \right) S_{\mathbf{r}}^{(0)}(t) \\
&= \frac{1}{n} \sum_{r \subset \mathcal{R}} \left( \frac{A_{\mathbf{r}}^1(t)}{A_{\mathbf{r}}^0(t) + A_{\mathbf{r}}^1(t)} - \left( \frac{A_{\mathbf{r}}^1(t)}{A_{\mathbf{r}}^0(t) + A_{\mathbf{r}}^1(t)} \right)^2 \right) [A_{\mathbf{r}}^0(t) + A_{\mathbf{r}}^1(t)] \\
&= \frac{1}{n} \sum_{r \subset \mathcal{R}} \left( \frac{A_{\mathbf{r}}^1(t)[A_{\mathbf{r}}^0(t) + A_{\mathbf{r}}^1(t)]}{A_{\mathbf{r}}^0(t) + A_{\mathbf{r}}^1(t)} - \frac{A_{\mathbf{r}}^1(t)^2}{A_{\mathbf{r}}^0(t) + A_{\mathbf{r}}^1(t)} \right) \\
&= \frac{1}{n} \sum_{r \subset \mathcal{R}} \frac{A_{\mathbf{r}}^1(t) A_{\mathbf{r}}^0(t)}{A_{\mathbf{r}}^0(t) + A_{\mathbf{r}}^1(t)} = h_{n,\,01}(t) \to_p h_{01}(t),
\end{aligned}$$

yielding agreement with (28). Hence, the asymptotic variances of the MPLE and of the Mantel–Haenszel estimator, at the null, are equal for sampling in general.

To characterize the relative efficiency away from the null ($\phi \neq 1$), Anderson & Bernstein (1985) showed that in the full-cohort situation $\hat{\phi}_{\mathrm{MH}}$ is one Newton step away from $\phi = 1$. It is easily shown that this result holds for $\hat{\phi}_{\mathrm{MH}}$ under risk set sampling. Although estimators which are one Newton step away from an $\sqrt{n}$-consistent estimator can be asymptotically efficient (see e.g. theorem 4.3 of Lehmann & Casella, 1998), we expect away from the null that $\hat{\phi}_{\mathrm{MH}}$, one step away from the inconsistent estimator $\phi = 1$, will generally be less efficient than $\hat{\phi}_{\mathrm{MPL}}$.

### 5.2. Relative efficiency for specific designs

For each of the designs 1–4, we verify that conditions 1–5 are satisfied and determine the standardized asymptotic distributions of $\hat{\phi}_{\mathrm{MH}}$ and $\hat{\Lambda}_n$. We assume that $\tau < \infty$; as $\lambda_0$ is already assumed bounded away from infinity, the finite interval Condition 1 holds. For designs 1 and 2, to satisfy condition 3, we assume that

$$f_k(t) = P(Z(t) = k \,|\, Y(t) = 1) \quad \text{for } k = 0, 1,$$

are bounded away from 0 over some non-trivial interval of time $[a, b] \subset [0, \tau]$.

For design 3, to satisfy condition 3, letting for $k = 0, 1$ and $l \in \mathcal{C}$,

$$q_l(t) = P(C(t) = l \,|\, Y(t) = 1) \quad \text{and} \quad f_{k,l}(t) = P(Z(t) = k \,|\, C(t) = l, Y(t) = 1),$$

we assume there exists $l \in \mathcal{C}$ with $m_l \geq 2$ such that over some non-trivial interval $[a, b] \subset [0, \tau]$ the functions $q_l(t)$ and $f_{k,l}(t)$ are bounded away from zero. That is, there is some strata in which a comparison of individuals can be made, and in that strata, the covariate value is not a constant.

For design 4, to satisfy condition 3 we assume that either: (i) the assumption for design 3 holds, or (ii) there exists an unequal pair $l_1, l_2 \in \mathcal{C}$ with $q_{l_1}(t), q_{l_2}(t), f_{j,l_1}(t), f_{k,l_2}(t)$ bounded away from zero. That is, we need to assume either that a meaningful comparison can be drawn: (i) within a strata or (ii) between two different strata.

Condition 4 is satisfied when $\tau < \infty$ assuming that

$$\inf_{t \in [0,\, \tau]} p(t) > 0, \quad \text{where} \quad p(t) = P(Y(t) = 1);$$

one needs only to invoke the strong law of large numbers in $D[0, 1]$ of Rao (1963) (after reversing the time axis), similar to BGL. In summary, in each of the examples which follow, we need to verify only conditions 2, 3 and 5. Throughout we continue to let $n(t) = |\mathcal{R}(t)|$, and define $\rho_n(t) = n(t)/n$.

*Design 1: full cohort.* Sampling all individuals who are at risk at the time of failure gives $\pi_t(\mathbf{r} \,|\, i) = \mathbf{1}(\mathbf{r} = \mathcal{R}(t))$, and with $n_k(t) = |\mathcal{R}_k(t)|$,

$$A^k_{\mathbf{r}}(t) = \sum_{i \in \mathcal{R}_k(t)} \pi_t(\mathbf{r} \,|\, i) = n_k(t) \mathbf{1}(\mathbf{r} = \mathcal{R}(t)),$$

and $A_{\mathcal{R}(t)}(t) = n(t)$. Using (18),

$$\frac{1}{n} H_{\mathbf{v}}(t) = \frac{1}{n} \sum_{\mathbf{r} \subset \mathcal{R}} A^{|\mathbf{v}|-1}_{\mathbf{r}}(t) \prod_{k \in \mathbf{v}} A^k_{\mathbf{r}}(t) = \rho_n(t) \prod_{k \in \mathbf{v}} \frac{n_k(t)}{n(t)} \to_p p(t) \prod_{k \in \mathbf{v}} f_k(t) = h_{\mathbf{v}}(t);$$

hence condition 2 is satisfied. Using that $\lambda_0$ is bounded away from zero, Condition 3 is satisfied as $f_0(t)$ and $f_1(t)$ are assumed bounded away from zero over some interval. By (27) the variance of the limiting normal is

$$\sigma^2 = \frac{\int_0^\tau \left( \phi_0^2 f_1(t) + \phi_0 f_0(t) \right) f_0(t) f_1(t) p(t) \lambda_0(t) \, \mathrm{d}t}{\left( \int_0^\tau f_0(t) f_1(t) p(t) \lambda_0(t) \, \mathrm{d}t \right)^2}.$$

Condition 5 can be seen to be satisfied by

$$e(\phi_0, t) = \frac{f_1(t)}{f_0(t) + \phi_0 f_1(t)} \quad \text{and} \quad \psi(\phi_0, t) = \frac{1}{f_0(t) + \phi_0 f_1(t)}.$$

Specializing further to the null case $\phi_0 = 1$,

$$\sigma^2 = \frac{1}{\int_0^\tau p(t) f_0(t) f_1(t) \lambda_0(t) \, \mathrm{d}t}, \quad e(\phi_0, t) = f_1(t) \quad \text{and} \quad \psi(\phi_0, t) = 1. \tag{32}$$

*Design 2: simple random sampling.* Letting $\mathbf{r}_k(t) = \{i \in \mathbf{r}, Z_i(t) = k\}$ and $r_k(t) = |\mathbf{r}_k(t)|$, the sampling probabilities (6) yield that for $\mathbf{r} \subset \mathcal{R}(t)$ with $|\mathbf{r}| = m$,

$$\pi_t(\mathbf{r}) = \binom{n(t)}{m}^{-1}, \quad A^k_{\mathbf{r}}(t) = r_k(t) \binom{n(t) - 1}{m - 1}^{-1}, \quad A_{\mathbf{r}}(t) = \frac{1}{m} \binom{n(t) - 1}{m - 1}$$

and so

$$h_{n,\mathbf{v}}(t)=\frac{1}{n}\sum_{\mathbf{r}\subset\mathcal{R}} a_{\mathbf{v}}^{|\mathbf{v}|-1}(t)\prod_{k\in\mathbf{v}} A_{\mathbf{r}}^{k}(t)=\frac{1}{nm^{|\mathbf{v}|-1}}\binom{n(t)-1}{m-1}^{-1}\sum_{|\mathbf{r}|=m,\,\mathbf{r}\subset\mathcal{R}(t)}\prod_{k\in\mathbf{v}} r_k(t)$$
$$=\frac{n(t)}{nm^{|\mathbf{v}|}}\binom{n(t)}{m}^{-1}\sum_{|\mathbf{r}|=m,\,\mathbf{r}\subset\mathcal{R}(t)}\prod_{k\in\mathbf{v}} r_k(t)=\frac{\rho_n(t)}{m^{|\mathbf{v}|}}E\prod_{k\in\mathbf{v}} X_k(t),$$

where

$$(X_0(t),X_1(t))\sim\mathcal{H}((n_0(t),n_1(t)),m), \tag{33}$$

the (hypergeometric) number of type 0 and 1 items in a simple random sample of $m$ items from a population with $n_0(t)$ and $n_1(t)$ items of type 0 and 1 respectively. Taking limits for $j,k$ distinct for $|\mathbf{v}|=2$

$$h_{jk}(t)=\binom{m-1}{m}p(t)f_j(t)f_k(t),$$

while for $|\mathbf{v}|=3$,

$$h_{jjk}(t)=\frac{p(t)}{m^3}\left((m)_2 f_j(t)f_k(t)+(m)_3 f_j^2(t)f_k(t)\right),$$

satisfying Condition 2. Condition 3 is verified here as it was for design 1.

By (27), the variance of the asymptotic distribution is

$$\sigma^2=\frac{\phi_0\int_0^\tau p(t)f_0(t)f_1(t)\left[(1+\phi_0)+(f_0(t)+\phi_0 f_1(t))(m-2)\right]\lambda_0(t)\,\mathrm{d}t}{(m-1)\left(\int_0^\tau p(t)f_0(t)f_1(t)\lambda_0(t)\,\mathrm{d}t\right)^2}, \tag{34}$$

and condition 5 is satisfied with

$$e(\phi_0,t)=\sum_{x_0+x_1=m}\frac{x_1}{x_0+\phi_0 x_1}\binom{m}{x_0,x_1}f_0^{x_0}(t)f_1^{x_1}(t)\quad\text{and}$$
$$\psi(\phi_0,t)=\frac{m}{p(t)}\sum_{x_0+x_1=m}\frac{1}{x_0+\phi_0 x_1}\binom{m}{x_0,x_1}f_0^{x_0}(t)f_1^{x_1}(t).$$

Under the null $\phi_0=1$, expression (34) simplifies to

$$\sigma^2=\binom{m}{m-1}\frac{1}{\int_0^\tau p(t)f_0(t)f_1(t)\lambda_0(t)\,\mathrm{d}t},$$

giving an asymptotic relative efficiency of $(m-1)/m$ with respect to the full-cohort variance (32), the same relative efficiency as $\hat{\phi}_{\mathrm{MPL}}$, as expected by the computation at the end of section 4.2. Lastly, in the null case $e(\phi_0,t)=f_1(t)$ and $\psi(\phi_0,t)=p(t)^{-1}$.

Previous efficiency work used a recursive representation of the factorial moments of the extended hypergeometric distribution (Harkness, 1965) to derive an asymptotic variance expression for 'small strata' case-control data (Breslow, 1981; Hauck & Donner, 1988). The expressions derived in these papers when there is a single case per set correspond to (34), a simplification that has not been previously described.

Figure 1 shows efficiency curves relative to $\hat{\phi}_{\mathrm{MPL}}$ as a function of $\log\phi$ by $m$ when $f_1(t)\equiv 0.2$. As noted previously in Breslow (1981), the Mantel–Haenszel estimator has high efficiency relative to $\hat{\phi}_{\mathrm{MPL}}$ over a fairly large region around the null.

For the next two designs, define for $\mathbf{r}\subset\mathcal{R}(t)$

$$\mathbf{r}_{k,l}(t)=\mathbf{r}\cap\mathcal{R}_k(t)\cap\mathcal{C}_l(t),\quad r_{k,l}(t)=|\mathbf{r}_{k,l}(t)|,\quad n_{k,l}(t)=|\mathcal{R}_k(t)\cap\mathcal{C}_l(t)|,$$

and let

$$\mathbf{X}_l(t) \sim H((n_{0,l}(t), n_{1,l}(t)), m_l), \quad \text{for } l \in \mathcal{C} \tag{35}$$

be independent multivariate hypergeometric vectors, as in (33).

*Design 3: matching.* Recalling the sampling probabilities (7) for the matching design, for $\mathbf{r} \subset \mathcal{C}_l(t)$ with $|\mathbf{r}| = m_l$, we have

$$A^k_{\mathbf{r}}(t) = \binom{c_l(t)}{m_l}^{-1} \frac{c_l(t)}{m_l} r_{k,l}(t) \quad \text{and} \quad A_{\mathbf{r}}(t) = \frac{1}{c_l(t)} \binom{c_l(t)}{m_l}.$$

Hence,

$$\begin{aligned} h_{n,\mathbf{v}}(t) &= \frac{1}{n} \sum_{\mathbf{r} \subset \mathcal{R}} a_{\mathbf{r}}^{|\mathbf{v}|-1}(t) \prod_{k \in \mathbf{v}} A^k_{\mathbf{r}}(t) = \frac{1}{n} \sum_{l \in \mathcal{C}} \sum_{\mathbf{r} \subset \mathcal{C}_l(t),\, |\mathbf{r}| = m_l} a_{\mathbf{r}}^{|\mathbf{v}|-1}(t) \prod_{k \in \mathbf{v}} A^k_{\mathbf{r}}(t) \\ &= \frac{n(t)}{n} \sum_{l \in \mathcal{C}} \frac{c_l(t)}{n(t)} \binom{c_l(t)}{m_l}^{-1} \sum_{\mathbf{r} \subset \mathcal{C}_l(t),\, |\mathbf{r}| = m_l} \prod_{k \in \mathbf{v}} m_l^{-1} r_{k,l}(t) \\ &= \rho_n(t) \sum_{l \in \mathcal{C}} \frac{c_l(t)}{n(t)} E \prod_{k \in \mathbf{v}} m_l^{-1} X_{k,l}(t) \end{aligned}$$

with $\mathbf{X}_l(t)$ as in (35). For $j \neq k$ distinct, taking limits we find

$$h_{jk}(t) = p(t) \sum_{l \in \mathcal{C}} \binom{m_l - 1}{m_l} q_l(t) f_{j,l}(t) f_{k,l}(t), \tag{36}$$

while for $|\mathbf{v}| = 3$,

$$h_{jjk}(t) = p(t) \sum_{l \in \mathcal{C}} q_l(t) \left( \frac{m_l - 1}{m_l^2} f_{j,l}(t) f_{k,l}(t) + \frac{(m_l - 1)_2}{m_l^2} f^2_{j,l}(t) f_{k,l}(t) \right)$$

and condition 2 is satisfied. Condition 3 is satisfied in a manner similarly as for design 2, with the additional assumption that $m_l \geq 2$, ensuring that $(m_l - 1)/m_l$ in (36) is positive.

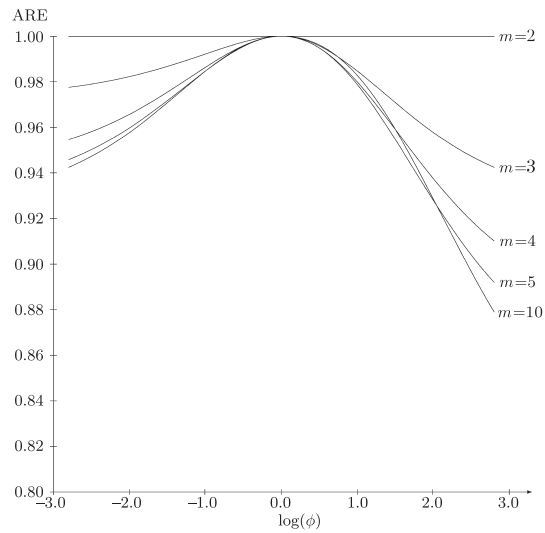


*Fig. 1.* Asymptotic efficiency by exposure rate ratio $\phi$ of Mantel–Haenszel relative to the partial likelihood estimator for simple random sampling of $m - 1$ controls. Probability of exposure $P(Z(t) = 1 \,|\, Y(t) = 1) = 0.2$.

In particular, from (27) the variance of the limiting normal is

$$\sigma^2 = \frac{\phi_0 \int_0^\tau p(t) \sum_{l\in\mathcal{C}} \left(\frac{m_l-1}{m_l^2}\right) q_l(t) f_{0,l}(t) f_{1,l}(t)[(1+\phi_0)+(f_{0,l}(t)+\phi_0 f_{1,l}(t))(m_l-2)]\lambda_0(t)\,\mathrm{d}t}{\left(\int_0^\tau p(t) \sum_{l\in\mathcal{C}} \left(\frac{m_l-1}{m_l}\right) q_l(t) f_{0,l}(t) f_{1,l}(t)\lambda_0(t)\,\mathrm{d}t\right)^2},$$

and condition 5 is satisfied with

$$e(\phi_0, t) = \sum_{l\in\mathcal{C}} q_l(t) \sum_{x_{0,l}+x_{1,l}=m_l} \left(\frac{x_{1,l}}{x_{0,l}+\phi_0 x_{1,l}}\right) \binom{m_l}{x_{0,l}, x_{1,l}} f_{0,l}^{x_{0,l}}(t) f_{1,l}^{x_{1,l}}(t)$$

and

$$\psi(\phi_0, t) = p(t)^{-1} \sum_{l\in\mathcal{C}} q_l(t) m_l \sum_{x_{0,l}+x_{1,l}=m_l} \left(\frac{1}{x_{0,l}+\phi_0 x_{1,l}}\right) \binom{m_l}{x_{0,l}, x_{1,l}} f_{0,l}^{x_{0,l}}(t) f_{1,l}^{x_{1,l}}(t).$$

Specializing further, under the null $\phi_0 = 1$,

$$\sigma^2 = \frac{1}{\left(\int_0^\tau p(t) \sum_{l\in\mathcal{C}} \left(\frac{m_l-1}{m_l}\right) q_l(t) f_{0,l}(t) f_{1,l}(t)\lambda_0(t)\,\mathrm{d}t\right)},$$

$$e(\phi_0, t) = \sum_{l\in\mathcal{C}} q_l(t) f_{1,l}(t) \quad \text{and} \quad \psi(\phi_0, t) = p(t)^{-1}.$$

*Design 4: counter matching.* Recalling the sampling probabilities (8) for the counter-matching design, where $\mathcal{C}$ is a set of types, $\mathcal{P}_\mathcal{C}(t) \subset \mathcal{R}(t)$ the collection of sets $\mathbf{r}$ with $m_l$ subjects of type $l$ at time $t$, $c_l(t)$ is the number of type $l$ subjects in $\mathcal{R}(t)$, and $C_i(t)$ the type of subject $i$ at time $t$. By (3) for $\mathbf{r} \in \mathcal{P}_\mathcal{C}(t)$,

$$\pi_t(\mathbf{r}) = \left[\prod_{l\in\mathcal{C}} \binom{c_l(t)}{m_l}\right]^{-1},$$

and letting $\mathbf{r}_{k,l}(t) = \{i \in \mathbf{r} : Z_i(t) = k, C_i(t) = l\}$, and $r_{k,l}(t) = |\mathbf{r}_{k,l}(t)|$,

$$A_\mathbf{r}^k(t) = \sum_{i\in\mathcal{R}_k(t)} \pi_t(\mathbf{r}\,|\,i) = \left[\prod_{l\in\mathcal{C}} \binom{c_l(t)}{m_l}\right]^{-1} \left(\sum_{l\in\mathcal{C}} r_{k,l}(t) \frac{c_l(t)}{m_l}\right),$$

and $A_\mathbf{r}(t) = \frac{1}{n(t)} \left[\prod_{l\in\mathcal{C}} \binom{c_l(t)}{m_l}\right]$.

As for design 3, we can write $h_{n,\mathbf{v}}(t)$ as an expectation

$$h_{n,\mathbf{v}}(t) = \rho_n(t) E\left(\prod_{k\in\mathbf{v}} \sum_{l\in\mathcal{C}} \frac{X_{k,l}(t) c_l(t)}{m_l n(t)}\right) = \rho_n(t) E\left(\sum_{l_p\in\mathcal{C},\, p=1,\ldots,|\mathbf{v}|} \prod_{k\in\mathbf{v}} \frac{X_{k,l_p}(t) c_{l_p}(t)}{m_{l_p} n(t)}\right),$$

and for $|\mathbf{v}| = 2$ with $j \neq k$ distinct, taking limits we find $h_{jk}(t)$ is $p(t)$ times

$$\sum_{l\in\mathcal{C}} \left(\frac{m_l-1}{m_l}\right) f_{j,l}(t) f_{k,l}(t) q_l^2(t) + \sum_{l_1\neq l_2} f_{j,l_1}(t) f_{k,l_2}(t) q_{l_1}(t) q_{l_2}(t), \tag{37}$$

which can be further simplified to yield

$$h_{jk}(t)=p(t)\left(f_j(t)f_k(t)-\sum_{l\in\mathcal{C}}\left(\frac{1}{m_l}\right)f_{j,l}(t)f_{k,l}(t)q_l^2(t)\right). \tag{38}$$

Applying the assumptions made at the beginning of this section in version (i) on the first sum in (37) or in version (ii) on the second sum in (37), condition 3 is satisfied. Similar calculations yield

$$h_{jkk}(t)=p(t)\left\{f_j(t)f_k^2(t)+\left(\sum_{l\in\mathcal{C}}\left(\frac{1}{m_l}\right)f_{j,l}(t)f_{1,k}(t)q_l^2(t)\right)\left(\sum_{l\in\mathcal{C}}(1-3f_{k,l}(t))q_l(t)\right)\right.$$
$$\left.-\sum_{l\in\mathcal{C}}\left(\frac{1}{m_l^2}\right)f_{j,l}(t)f_{k,l}(t)(1-2f_{k,l}(t))q_l^3(t)\right\}.$$

Hence, condition 2 is satisfied, and $\sigma^2$ can now be calculated by (27).

For the parameters in the limiting distribution for the baseline hazard estimator, we have

$$e(\phi_0,t)=\sum_{x_{0,\alpha}+x_{1,\alpha}=m_\alpha,\,\alpha\in\mathcal{C}}\left(\frac{\sum_{l\in\mathcal{C}}x_{1,l}\frac{q_l(t)}{m_l}}{\sum_{l\in\mathcal{C}}(x_{0,l}+\phi_0 x_{1,l})\frac{q_l(t)}{m_l}}\right)\prod_{\alpha\in\mathcal{C}}\binom{m_\alpha}{x_{0,\alpha},x_{1,\alpha}}f_\alpha^{x_0}(t)f_\alpha^{x_1}(t),$$

and

$$\psi(\phi_0,t)=p(t)^{-1}\sum_{x_{0,\alpha}+x_{1,\alpha}=m_\alpha,\,\alpha\in\mathcal{C}}\left(\frac{1}{\sum_{l\in\mathcal{C}}(x_{0,l}+\phi_0 x_{1,l})\frac{q_l(t)}{m_l}}\right)\prod_{\alpha\in\mathcal{C}}\binom{m_\alpha}{x_{0,\alpha},x_{1,\alpha}}f_\alpha^{x_0}(t)f_\alpha^{x_1}(t).$$

We specialize further to the case where there are two strata, $|\mathcal{C}|=2$, and the binary strata variable $C(t)\in\{0,1\}$ is a (perhaps easily available) surrogate for the true binary exposure $Z(t)\in\{0,1\}$. Recalling

$$f_{k,l}(t)=P(Z(t)=k\,|\,C(t)=l,\,Y(t)=1)\quad k,l\in\{0,1\},$$

we have

$$f_{k,l}(t)q_l(t)=P(Z(t)=k\,|\,C(t)=l,\,Y(t)=1)P(C(t)=l\,|\,Y(t)=1)$$
$$=P(Z(t)=k,\,C(t)=l\,|\,Y(t)=1)=\pi_{k,l}(t)$$

say, and

$$\delta(t)=P(C(t)=1\,|\,Z(t)=1,\,Y(t)=1)\quad\text{and}\quad\gamma(t)=P(C(t)=0\,|\,Z(t)=0,\,Y(t)=1),$$

the sensitivity and specificity of $Z(t)$ for $C(t)$. As

$$\pi_{11}(t)=\delta(t)f_1(t),\qquad\pi_{10}(t)=(1-\delta(t))f_1(t)$$
$$\pi_{01}(t)=(1-\gamma(t))f_0(t),\qquad\pi_{00}(t)=\gamma(t)f_0(t),$$

and (38) gives $h_{01}(t)$ for, say $m_0=m_1=1$, as $p(t)$ times

$$f_0(t)f_1(t)-(f_{0,1}(t)f_{1,1}(t)q_1^2(t)+f_{0,0}(t)f_{1,0}(t)q_0^2(t))$$
$$=f_0(t)f_1(t)-(\pi_{0,1}(t)\pi_{1,1}(t)+\pi_{0,0}(t)\pi_{1,0}(t))$$
$$=f_0(t)f_1(t)-((1-\gamma(t))f_0(t)\delta(t)f_1(t)+\gamma(t)f_0(t)(1-\delta(t))f_1(t))$$
$$=f_0(t)f_1(t)((1-\delta(t))(1-\gamma(t))+\gamma(t)\delta(t)). \tag{39}$$

In a similar way, $h_{011}(t)$ and $h_{001}(t)$ can be expressed in terms of the sensitivity, specificity and probability of exposure integrated against the baseline hazard. Using (26) and the partial likelihood variance given in (A3) from Langholz & Borgan (1995), asymptotic efficiencies for $\hat{\phi}_{\mathrm{MH}}$ relative to $\hat{\phi}_{\mathrm{MPL}}$ can be computed.

Figure 2 shows the asymptotic relative efficiencies by $\log(\phi)$ with $P(Z(t)=1\,|\,Y(t)=1)=0.2$ for $m_0,m_1\in\{1,2\}$ when the conditional distribution of $(Z(t),C(t))$ given $Y(t)=1$ does not

depend on $t$, which holds, approximately, for rare outcomes when censoring does not depend on $(Z(t), C(t))$. Although there is some difference in the relative efficiencies by choice of $m_0$ and $m_1$ and the sensitivity and specificity of $C$ for $Z$, $\hat{\phi}_{\mathrm{MH}}$ has fairly high efficiency in a wide range of situations.

Under the null $\phi_0 = 1$ (27) simplifies to yield

$$\sigma^2 = \frac{1}{\int_0^\tau p(t)\left(f_0(t)f_1(t) - \sum_l (\frac{1}{m_l}) f_{0,l}(t) f_{1,l}(t) q_l^2(t)\right)\lambda_0(t)}, \tag{40}$$

and using $x_{0,l} + x_{1,l} = m_l$ and $EX_{k,l} = m_l f_{k,l}(t)$, we have

$$e(\phi_0, t) = \sum_{l \in \mathcal{C}} f_{1,l}(t) q_l(t) \quad \text{and} \quad \psi(\phi_0, t) = p(t)^{-1}.$$

When $(m_0, m_1) = (1, 1)$, so that the design matches one control with 'surrogate exposure' $C(t)$ value opposite to the exposure $Z(t)$ of the case, substituting (39) into (40) yields

$$\sigma^2 = \left(\int_0^\tau p(t) f_0(t) f_1(t)((1-\delta(t))(1-\gamma(t)) + \gamma(t)\delta(t))\lambda_0(t)\,\mathrm{d}t\right)^{-1},$$

which is the equal to the asymptotic variance for the 1:1 counter-matching design when using the $\hat{\phi}_{\mathrm{MPL}}$ (Langholz & Clayton, 1994), as anticipated by the argument at the end of section 4.2. We note that, as in Langholz & Clayton (1994), when the sensitivity and specificity are close to 1 (or 0), the counter-matching design has efficiency close to that of the full cohort.

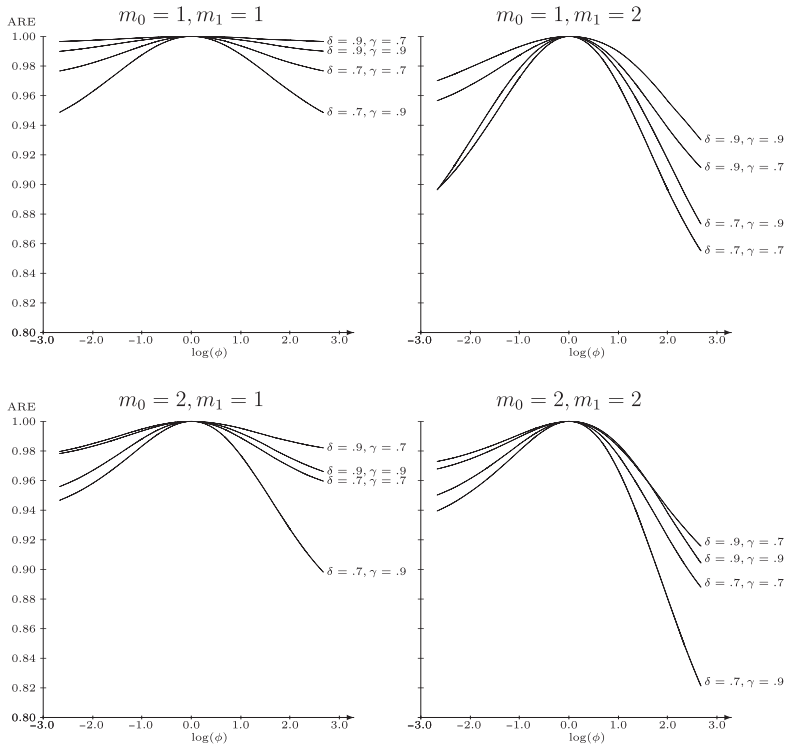


*Fig. 2.* Asymptotic efficiency by exposure rate ratio $\phi$ of Mantel–Haenszel relative to the partial likelihood estimator for counter matching by sensitivity ($\delta = P(Z(t) = 1 | C(t) = 1, Y(t) = 1)$) and specificity ($\gamma = P(Z(t) = 0 | C(t) = 0, Y(t) = 1)$). Probability of exposure $P(Z(t) = 1 | Y(t) = 1) = 0.2$.

## 6. Discussion

We have described the Mantel–Haenszel estimator for the rate ratio in a proportional hazards model, and associated baseline hazard estimator, for a large class of case-control sampling designs within the context of risk set sampling from cohort data. We have further demonstrated necessary conditions for consistency and asymptotic normality as well as provided efficiency calculations for a number of designs. Our results generalize the work of Zhang *et al.* (2000) and Zhang (2000) by allowing for quite general censoring and sampling, recurrent events, and time-dependent exposure variable. Although, in general, $\hat{\phi}_{\mathrm{MH}}$ is less efficient than $\hat{\phi}_{\mathrm{MPL}}$, we showed that for general sampling, when $\phi_0 = 1$, $\hat{\phi}_{\mathrm{MH}}$ has efficiency equal to $\hat{\phi}_{\mathrm{MPL}}$. Further, for the specific sampling designs we studied, we found that the efficiency loss for $\hat{\phi}_{\mathrm{MH}}$ was not large within a range of $\phi_0$ of practical interest. Like other Mantel–Haenszel estimators and unlike $\hat{\phi}_{\mathrm{MPL}}$, $\hat{\phi}_{\mathrm{MH}}$ under sampling has a simple closed form.

We have described and analysed a number of extensions to the Mantel–Haenszel estimator in Goldstein and Langholz (2006) including handling multilevel exposure, estimators based on

$$R_{jk}(t) = \int_0^t \sum_{\mathbf{r} \subset \mathcal{R}} a(A^0_{\mathbf{r}}(s), A^1_{\mathbf{r}}(s))\ A^k_{\mathbf{r}}(s)\, \mathrm{d}N^j_{\mathbf{r}}(s),$$

where $a(u, v)$ is a positive symmetric function along the lines of Zhang *et al.* (2000) and Zhang (2000), and robust variance estimators based on the optional variation parallel to the estimator described by Liang (1985).

An important generalization relevant to the matching design is to the stratified proportional hazards model with $\lambda_i(t) = Y_i(t)\lambda_{C_i(t)}(t)\phi_0^{Z_i(t)}$, where $\lambda_c(t)$ is the baseline hazard function for matching stratum $c$ (e.g. Andersen *et al.*, 1993). Even in this extended model, it remains true that $R_{10}(t) - \phi_0 R_{01}(t)$ is a local square integrable martingale. We can guarantee the consistency of $\hat{\phi}_{\mathrm{MH}}$ in this situation by letting condition 1 hold with $\lambda_l(t)$ replacing $\lambda_0(t)$, and condition 2 hold with

$$H_{\mathbf{v},l}(t) = \sum_{r \subset \mathcal{C}_l(t)} a^{|v|-1}_{\mathbf{r}}(t) \prod_{k \in \mathbf{v}} A^k_{\mathbf{r}}(t)$$

and its scaled limit $h_{\mathbf{v},l}(t)$, replacing $H_{\mathbf{v}}(t)$ and $h_{\mathbf{v}}(t)$, respectively, for each $l \in \mathcal{C}$. Now lemma 1 holds for each $l \in \mathcal{C}$, and summing over all $l$ in (the finite set) $\mathcal{C}$ gives the asymptotic normality of $\hat{\phi}_{\mathrm{MH}}$ for the matching design with strata-specific baseline hazard $\lambda_l(t)$, with variance

$$\sigma^2 = \frac{\int_0^\tau \sum_{l \in \mathcal{C}} (\phi_0^2 h_{011,l}(t) + \phi_0 h_{100,l}(t))\lambda_l(t)\, \mathrm{d}t}{(\int_0^\tau \sum_{l \in \mathcal{C}} h_{01,l}(t)\lambda_l(t)\, \mathrm{d}t)^2}.$$

### Acknowledgement

This work was supported by United States National Cancer Institute grant CA14089.

### References

Anderson, J. R. & Bernstein, L. (1985). Asymptotically efficient two-step estimators of the hazards ratio for follow-up studies and survival data. *Biometrics* **41**, 733–739.

Andersen, P. K., Borgan, Ø., Gill, R. D. & Keiding, N. (1993). *Statistical models based on counting processes*. Springer Verlag, New York.

Andrieu, N., Goldstein, A. M., Thomas, D. C. & Langholz, B. (2000). Counter-matching in gene-environment interaction studies: Efficiency and feasibility. *Am. J. Epidemiol.* **153**, 265–274.

Borgan, Ø. & Langholz, B. (1998). Risk set sampling designs for proportional hazards models. In *statistical analysis of medical data: new developments* (eds B. S. Everitt & G. Dunn), 75–100. Arnold, London.

Borgan, Ø., Goldstein, L. & Langholz, B. (1995). Methods for the analysis of sampled cohort data in the Cox proportional hazards model. *Ann. Statist.* **23**, 1749–1778.

Breslow, N. E. (1981). Odds ratio estimators when the data are sparse. *Biometrika* **68**, 73–84.

Breslow, N. E. (1996). Statistics in epidemiology: The case-control study. *Journal of the American Statistical Association* **91**, 14–28.

Breslow, N. E. & Day, N. E. (1980). *The design and analysis of case-control studies*. IARC Scientific Publications, Lyon.

Goldstein, L. & Langholz, B. (2006). *Cohort sampling schemes for the Mantel–Haenszel estimator: extensions to multilevel covariates, stratified models, and robust variance estimators*. http://arxiv.org/abs/math.ST/0609333 (accessed September 12, 2006)..

Harkness, W. (1965). Properties of the extended hypergeometric distribution. *Ann. Math. Statist.* **36**, 938–945.

Hauck, W. W. & Donner, A. (1988). The asymptotic relative efficiency of the Mantel–Haenszel estimator in the increasing-number-of-strata case. *Biometrics* **44**, 379–384.

Hjort, N. L. & Pollard, D. (1993). *Asymptotics of minimizers of convex processes*. Statistical Research Report 5/93, Institute of Mathematics, University of Oslo, Oslo.

Langholz, B. & Borgan, Ø. (1995). Counter-matching: a stratified nested case-control-sampling method. *Biometrika* **82**, 69–79.

Langholz, B. & Clayton, D. (1994). Sampling strategies in nested case-control studies. *Environ. Health Perspect.* **102**(Suppl. 8), 47–51.

Langholz, B. & Goldstein, L. (1996). Risk set sampling in epidemiologic cohort studies. *Statistical Science* **11**, 35–53.

Langholz, B. & Goldstein, L. (2001). Conditional logistic analysis of case-control studies with complex sampling. *Biostatistics* **2**, 63–84.

Lehmann, E. & Casella, G. (1998). *Theory of point estimation*, 2nd edn. Springer, New York.

Liang, K.-Y. (1985). Odds ratio inference with dependent data. *Biometrika* **72**, 678–682.

Mantel, N. & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* **22**, 719–748.

Rao, R. R. (1963). The law of large numbers for $D[0, 1]$-valued random variables. *Theory Probab. Appl.* **8**, 70–74.

Robins, J. M., Gail, M. H. & Lubin, J. H. (1986). More on 'Biased selection of controls for case-control analysis of cohort studies'. *Biometrics* **42**, 293–299.

Zhang, Z.-Z. (2000). On consistency of Mantel–Haenszel type estimators in nested case-control studies. *J. Jpn Statist. Soc.* **30**, 205–211.

Zhang, Z.-Z., Fujii, Y. & Yanagawa, T. (2000). On Mantel–Haenszel type estimators in simple nested case-control studies. *Commun. Statist. A – Theory Methods* **29**, 2507–2521.

*Received October 2005, in final form September 2006*

Larry Goldstein, USC Department of Mathematics, KAP-108, Los Angeles, CA 90089-2532, USA.
E-mail: larry@math.usc.edu