

Adjustment for Selection Bias in Observational Studies with Application to the Analysis of Autopsy Data

S. Haneuse^{a, b} J. Schildcrout^g P. Crane^{b, c} J. Sonnen^d J. Breitner^{e, f} E. Larson^a

^aGroup Health Center for Health Studies, ^bDepartment of Biostatistics, ^cDivision of General Internal Medicine, ^dDepartment of Pathology, ^eVA Puget Sound Health Care System and ^fDepartment of Psychiatry and Behavioral Sciences, University of Washington, Seattle, Wash., and ^gDepartment of Biostatistics, Vanderbilt University, Nashville, Tenn., USA

Key Words

Alzheimer's disease · Autopsy · Bootstrap · Dementia · Regression analysis · Selection bias · Weighted estimating equations

Abstract

Background: The interpretation of neuropathological studies of dementia and Alzheimer's disease is complicated by potential selection mechanisms that can drive whether or not a study participant is observed to undergo autopsy. Notwithstanding this, there appears to have been little emphasis placed on potential selection bias in published reports from population-based neuropathological studies of dementia. **Methods:** We provide an overview of methodological issues relating to the identification of and adjustment for selection bias. When information is available on factors that govern selection, inverse-probability weighting provides an analytic approach to adjust for selection bias. The weights help alleviate bias by serving to bridge differences between the population from which the observed data may be viewed as a representative sample and the target population, identified as being of scientific interest. **Results:** We illustrate the methods with data obtained from the Adult Changes in Thought study. Adjustment for potential selection bias yields substantially strengthened association between neuro-

pathological measurements and risk of dementia. **Conclusions:** Armed with analytic techniques to adjust for selection bias and to ensure generalizability of results from population-based neuropathological studies, researchers should consider incorporating information related to selection into their data collection schemes. Copyright © 2009 S. Karger AG, Basel

Introduction

Population-based neuropathological (NP) studies of dementia and Alzheimer disease (AD) rely on measurements obtained from study participants who undergo autopsy. While scientific interest often lies in understanding associations in the general population, mechanisms at play during the study may lead to the observed autopsy sample no longer being representative. As such establishing generalizability, or external validity, of results beyond the study sample requires careful consideration of potential selection bias.

Recently, Zaccai et al. [1] conducted a comprehensive review of population-based NP studies of dementia and identified six such studies: the Hisayama study, the Cambridge City over 75 Cohort Study, the Vantaa 85+ Study, the Medical Research Council Cognitive Function and

Ageing Study, the Honolulu-Asia Aging Study and the Cache County Study of Aging and Memory's. While the review identified three major mechanisms related to selection (nonresponse, attrition/death and willingness for brain donation), it also noted that none of the reviewed studies had reported attempts to generalize beyond the study sample. In the absence of adjusting for selection, therefore, one can only interpret the results from these studies as pertaining to a population of autopsied individuals; to date, we are aware of only one population-based study of NP risk factor associations for dementia that attempted to adjust for potential selection [2]. Beyond the context of population-based studies, previous attempts at generalizing results based on autopsy data have been limited [3–5].

We therefore see an opportunity to review methodological issues related to bias in autopsy-based NP studies of dementia and AD. Specifically, we consider two key aspects of selection bias: identification and adjustment. Identification is facilitated via a structural framework for characterizing selection bias, developed by Hernán et al. [6], while adjustment is performed using inverse-probability weighting. We illustrate the methods with a study of NP risk factors for dementia among participants in the Adult Changes in Thought (ACT) study.

Methods and Materials

The problem of adjusting for potential selection bias can usefully be cast as a missing data problem; some individuals are selected into the sample to have complete data, while other individuals are not selected and consequently have missing information. As such, methods developed for missing data problems provide convenient tools for handling selection bias. In the following we provide a general description of the techniques, applicable in any setting where selection bias may be an issue. Prior to doing so, and to provide a concrete setting for motivation, we provide a brief description of the ACT study.

Adult Changes in Thought

The ACT study is an ongoing population-based longitudinal study of incident AD and dementia, among individuals aged 65 years and older, from a population base of 23,000 members of Group Health Cooperative (GHC), a large health care provider in King County, Wash. [7]. Between 1992 and 2004, enrollment of participants consisted of two phases: an initial cohort enrolled from 1992 to 1994 and an expansion cohort enrolled from 2002 to 2004. For all enrollees, demographic, medical history and functional status information was collected at baseline and at subsequent biennial follow-up visits. A blood sample was obtained from consenting enrollees at baseline to permit apolipoprotein E (APOE) genotyping. At each visit study participants were evaluated with a protocol-based examination using the Cognitive Abil-

ities Screening Instrument (CASI) [8]. At each follow-up visit, a CASI score of 85 or less triggered a comprehensive dementia workup, with clinical dementia diagnosis following Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) criteria [9]. Based on these criteria, enrollees were required to be dementia free at baseline; subsequent diagnoses of dementia were therefore taken to be incident cases. Follow-up of participants continued until the first of a diagnosis of dementia, withdrawal or death. For participants that withdraw from ACT but remain members of Group Health Cooperative, a separate protocol has been established for postwithdrawal verification of dementia and vital status. As ACT is ongoing, analyses presented here consider data collected up until December 31, 2006.

Participants were also asked for consent for brain autopsy at the first follow-up examination for the original cohort and at baseline for the expansion cohorts. For participants who had not decided whether or not to provide consent, additional requests were made at subsequent biennial visits. In accordance with state law, next-of-kin were also required to file informed consent for autopsy after death. For autopsied individuals, evaluation of NP measures followed established methods and was performed blinded to the dementia diagnosis [2].

Notation

Suppose interest lies in determining the association between an outcome Y (e.g., clinical dementia diagnosis) and a risk factor X (e.g., some NP measure), adjusting for a set of potential confounders, C . A common approach used to characterize the association is to construct a regression model; let $\mu = E[Y | X, C]$ denote the mean of the outcome, given the risk factor of interest and potential confounders and assume

$$g(\mu) = \beta_0 + \beta_x X + \beta_c C, \quad (1)$$

where $g(\cdot)$ is the link function [11]. Common link functions include the identity for linear regression, the log for Poisson regression and the logit for logistic regression. Having identified a model that characterizes the question of interest, the primary (statistical) goal is to estimate, and perform inference on, the unknown parameter $\beta = (\beta_0, \beta_x, \beta_c)$.

Estimation

Towards estimation of the β regression parameters, researchers have a variety of statistical tools at their disposal. Here we consider semiparametric estimating equations, with an alternative being fully parametric maximum likelihood. Given a sample of size n , an estimate of β can be obtained by solving the equation

$$U(\beta) = \sum_{i=1}^n D_i^T V_i^{-1} (Y_i - \mu_i) = 0, \quad (2)$$

where D_i is the derivative of μ_i with respect to β and V_i is the assumed variance of Y_i . Intuitively, the solution to equation 2 is the value of β such that the observed and expected total outcomes (Y_i and μ_i , respectively) are, on average, equal.

Provided model 1 is correctly specified, equation 2 is unbiased (i.e. has zero expectation) regardless of the choice of V , an important robustness property not enjoyed by maximum likelihood-based approaches. Hence, the corresponding estimate of β can be shown to be consistent (asymptotically unbiased), again regardless of the choice of V_i [12]. Estimation of standard errors for the

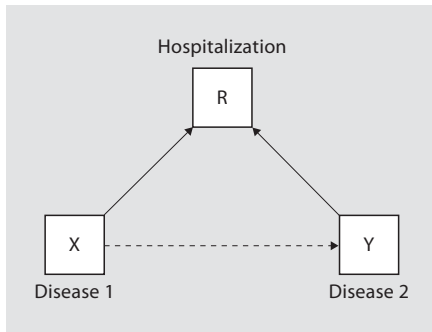


Fig. 1. Directed acyclic graph illustrating potential selection bias arising from Berkson's paradox.

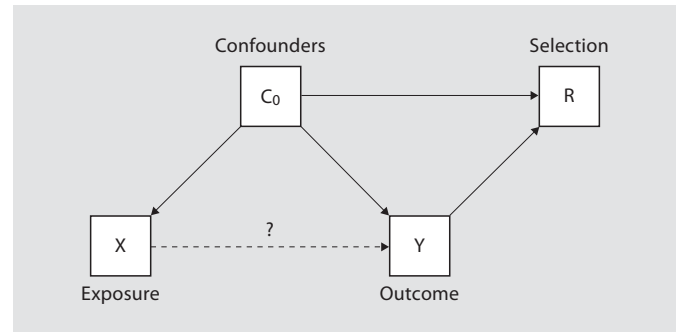


Fig. 2. Directed acyclic graph illustrating potential selection bias in the context of a case-control study.

β estimates is straightforward, with the 'sandwich estimator' well-known to be robust in the sense that valid inference is obtained, even if the variance V_i is misspecified [12].

Selection Bias

Having solved equation 2 to obtain an estimate of β , the precise interpretation of the estimate depends on several factors including the link function, the form of X and choice of C . The interpretation also requires consideration of the sampling frame and, in particular, an understanding of factors that govern whether or not an individual is observed in the study sample. Towards this it is useful to define two populations: the target population and the sampling population. The target population refers to the population of scientific interest, typically identified a priori and characterized by the hypothesis under investigation, together with inclusion/exclusion criteria. The sampling population refers to a (potentially hypothetical) population from which the observed sample may be viewed as being representative. For example, in the ACT study one could take the target population to be the population of individuals who are dementia-free at age 65. For the analysis of the autopsy data, the sampling population would be a population of individuals who are dementia-free at age 65, subsequently died and had an autopsy.

In settings where the observed (sample) data are obtained via random sampling, the target and sampling populations can easily be seen to be equivalent. Beyond random sampling, however, the two populations are often not equivalent so that conclusions drawn on the basis of the observed data, while clearly pertaining to the sampling population, may not generalize to the target population. The mechanism driving differences between the two populations can be generically referred to as the *selection mechanism*, although it is often the case that it involves several separate underlying mechanisms (such as refusal, dropout and death). Bias introduced by interpreting results based on the observed data, a sample from the sampling population, as pertaining to the target population is referred to as *selection bias* and, as we expand upon below, the extent to which such bias exists in any given application will depend on the nature of the selection mechanism [6, 10].

Towards characterizing differences between the target and sampling populations, and hence the selection mechanism, let R be a binary indicator of selection; $R = 1$ denotes selection into the sample and $R = 0$ nonselection. Using this notation, the subset of

data for which $R = 1$ is taken to be a random sample from the sampling population. Returning to the study of autopsy data from ACT, individuals that underwent autopsy would have $R = 1$, while those who did not would have $R = 0$.

Parallel to viewing selection bias in terms of differences between the sampling and target populations, one can also consider characterizing selection bias in terms of the associations corresponding to each population. To see this, note that in any given data situation one does not actually estimate the components of model 1, but rather the components of

$$g(\mu^*) = \beta_0^* + \beta_x^* X + \beta_c^* C, \quad (3)$$

where $\mu^* = E[Y | X, C, R = 1]$ denotes the mean outcome, *conditional on having been selected into the observed data*. The parameters in model 3 can be distinguished from those of model 1 in that they are stratum-specific with respect to R and, hence, pertain to the sampling population. Model 1 is marginal with respect to R (i.e., averages across those selected and those not) and therefore pertains to the target population. Selection bias occurs when one is interested in estimating β , but $\beta^* \neq \beta$.

Recently, Hernán et al. [6] developed a framework within which the causal structure of selection bias may be characterized. Specifically, suppose the selection mechanism depends on a set of covariates L , selection bias (i.e. $\beta^* \neq \beta$) results when L contains *both* (1) X or a cause of X *and* (2) Y or a cause of Y . That is, selection into the observed sample is *jointly* driven by the exposure, or a cause of exposure, *and* the outcome, or a cause of the outcome. If selection into the observed sample is independent of either of these, no adjustment is necessary. Figures 1 and 2 illustrate the framework in the context of two simple examples. Specifically figure 1 illustrates Berkson's paradox, often attributed to studies of hospitalized populations, where conditioning on $R = 1$ (i.e. hospitalization) can create an association between two diseases, even if they are independent in the general population. Figure 2 illustrates the potential for selection bias in a case-control study. Since the case-control design is typically employed in the context of a rare outcome, selection is naturally driven by case status. In settings where actual participation of invited enrollees is driven by a confounder of the exposure/outcome relationship, selection bias can be induced. Hernán et al. [6] provide an excellent and detailed exposition, with further illustrations and discussion.

Weighted Analyses

In settings where adjustment is necessary, inverse probability weighting provides a means to obtain consistent estimates of β (i.e. the parameters in model 1) [13]. Suppose the underlying selection mechanism depends on L via the model $\pi(\gamma) = P(R = 1 | L)$, where γ is an unknown parameter vector. Practically, it is common to specify $\pi(\gamma)$ as a logistic regression model. Key to adjusting for differences between the target and sampling populations, therefore, is that all components of L must be observable on all individuals, including those not selected. In the missing data literature, this corresponds to the 'missing-at-random' (MAR) assumption [14]. In observational studies, the latter is not verifiable from the data alone and will typically rely on scientific input and the published literature.

Given an estimate, an estimate of β (from model 1) is obtained by solving the weighted estimating equation

$$U(\beta, \hat{\gamma}) = \sum_{i=1}^n w_i(\hat{\gamma}) D_i^T V_i^{-1} (Y_i - \mu_i) = 0, \quad (4)$$

where the weight $w_i(\hat{\gamma}) = \pi_i(\hat{\gamma})^{-1}$ is the inverse estimated probability of selection for the i -th individual in the sample. The introduction of the weights (essentially) generates $w_i(\hat{\gamma})$ duplicates of the i -th individual, the effect of which is to create a pseudosample that, heuristically, we would have observed had 'selection' not occurred. As such, the pseudosample may be viewed as a random sample from the target population so that the estimate obtained by solving equation 4 is interpretable in terms of the scientific population of interest. Practically, incorporation of weights into the estimating equation is straightforward in most statistical packages.

For inference, standard errors and 95% confidence intervals (CIs) could be constructed using a sandwich-based estimate. Robins et al. [13, section 6.4] provide analytic expressions for the standard error that accounts for the estimation of $\hat{\gamma}$. The calculation is complex, however, and not readily implemented. The bootstrap provides a practical alternative that is easily implemented in most statistical packages. In the following, we present bias-corrected and accelerated (BCa) CIs, shown to be more accurate than standard percentile-based bootstrap interval estimates [15]. Briefly, the calculation is a relatively straightforward extension of the usual BCa calculation with the key difference being that the bootstrap samples are taken from the entire dataset (i.e. those with $R = 0$ and $R = 1$). Consequently, the selection model parameters γ are reestimated for each bootstrap replicate, as are the weights. Although not presented here, a detailed description of the algorithm is available from the corresponding author upon request.

Results

To illustrate the methods, we consider their application to the autopsy data from ACT. While ACT is a rich research resource and provides the opportunity to investigate a range of questions relating to neuropathology of dementia for simplicity we examine the cross-sectional association between NP measures and dementia. In particular, we consider the association among individuals

currently alive. Within this context, the sampling population is taken to be a hypothetical population from which autopsied individuals are a random sample, while the target population is taken to be those currently alive and enrolled in ACT. In an evaluation of selection bias in an autopsy case series from the Alzheimer's Disease Patient Registry, Tsuang et al. [5] considered a similar target population. A key difference between their choice and the one presented here is that they combined individuals that died without autopsy together with those alive and enrolled. As such, while similar, their choice of target population corresponds to addressing a different question, and would consequently require a separate assessment of the MAR assumption. Beyond the specific choices of this paper and of Tsuang et al. [5], there are a multitude of other potential choices; care is therefore required in the assessment of the implications of a specific choice in terms of the scientific question being addressed.

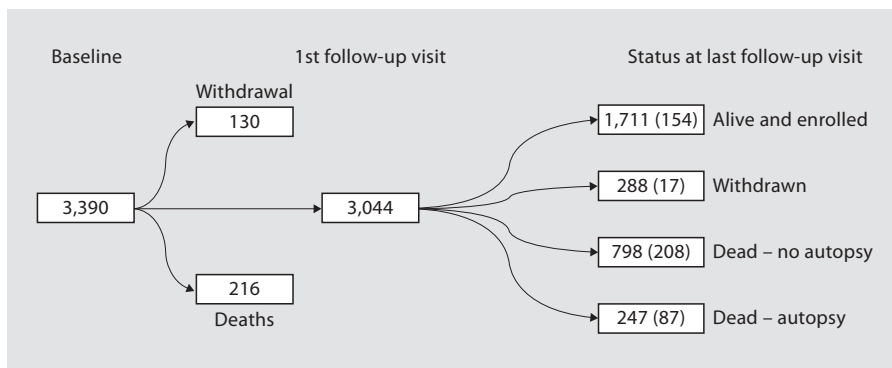
Cohort Characteristics

Table 1 provides baseline characteristics of ACT participants. Of the 3,390 enrollees, 3,044 (90%) have at least one follow-up visit; 216 (6%) died prior to a subsequent follow-up while 130 (4%) withdrew from the study prior to their first follow-up visit. Of the 3,044 enrollees with at least one follow-up visit 466 (15%) had a clinical diagnosis of dementia during follow-up (also see fig. 3). At baseline the data are complete with the exception of APOE genotyping, missing for approximately 11% of participants. Generally, older participants (at baseline) are more likely to have a dementia diagnosis during the course of follow-up, as are whites, those less educated and those with at least one APOE $\epsilon 4$ allele.

Of the 3,044 participants with at least one follow-up visit, 979 (32.2%) had consented to autopsy. Consent rates did not differ significantly between the two cohorts (31.8% for the original cohort and 33.4% for the expansion cohort; $p = 0.429$); further, rates did not differ according to gender, and baseline marital status. Consent rates did differ significantly between whites and non-whites (33.8 vs. 19.0%; $p < 0.001$), across baseline age strata (increasing linearly from 29.8% among those ≤ 70 years to 42.1% among those > 90 years; $p = 0.002$) and by education (increasing from 22.8% among those with less than a high school education to 39.3% among those with a college degree).

Table 2 provides demographic information, obtained at last known follow-up, according to the vital status and whether or not an autopsy was performed. Also shown are characteristics reported as being predictive of selec-

Fig. 3. Flowchart indicating the progression of ACT enrollees from baseline to status at last known follow-up. Among those with at least 1 follow-up visit, individuals are grouped according to vital status, whether or not they had withdrawn and, among those that died, whether or not an autopsy was performed. Number of patients with a clinical diagnosis of dementia during follow-up is given in parentheses.



tion in autopsy studies of dementia [1]. We find that of the variables we considered, each of cohort membership, dementia status, age, gender, race, education, marital status and depression are associated (at the 0.05 level) with vital/autopsy status, at the last known visit. We also find that participants who die are generally more likely to be demented, older, male, nonwhite, and not married. Further, ‘alive-withdrawn’ and ‘dead-no autopsy’ individuals tend to be less well educated; 18.1% have less than a high school education compared to 9.2% among the ‘alive-enrolled’ and ‘dead-autopsied’ individuals.

We focus on four key NP measures: (1) Braak stage, a staging criterion for AD based on neurofibrillary tangles [16], (2) cerebral microvascular infarcts, focal lesions attributed to ischemia and found only on microscopic examination [17], (3) neocortical Lewy bodies, abnormal aggregates of protein that develop inside nerve cells, and (4) cystic infarcts, which derive from artery and arteriole obstruction. The left-hand side of table 3 provides frequencies for the four NP measurements among the autopsied individuals, by dementia status at last known visit. Note that 33 individuals were excluded from the nondemented group, since their deaths occurred more than 2 years after their last visit date and they may have developed dementia between their last study visit and death. Of the remaining 214 autopsied individuals, 87 (40.7%) were clinically diagnosed with dementia.

Unweighted Analyses

Unweighted analyses of the autopsy data, using estimating equation 2, are reported in table 3. Shown are relative risk (RR) estimates based on a log-linear model for the binary dementia outcome; again we note that these estimates correspond to the β^* parameters of model 3 which, as we explore below, may or may not equal β . All four NP measurements were included in the model si-

Table 1. Baseline characteristics for ACT participants with at least 1 follow-up visit

	Nondemented		Demented	
	n	% ^a	n	% ^a
Total	2,578		466	
Cohort				
Original	1,899	73.7	418	89.7
Expansion	679	26.3	48	10.3
Age, years				
≤70	655	25.4	37	7.9
71–75	814	31.6	101	21.7
76–80	576	22.3	139	29.8
81–85	356	13.8	118	25.3
85–90	141	5.5	50	10.7
>90	36	1.4	21	4.5
Gender				
Male	1,045	40.6	187	40.1
Female	1,532	59.4	279	59.9
Missing	1		0	
Race				
White	2,316	89.9	432	92.7
Nonwhite	261	10.1	34	7.3
Missing	1		0	
Education				
<High school	298	11.6	79	17.0
High school	627	24.3	142	30.5
Some college	1,134	44.0	178	38.2
College graduate	516	20.0	67	14.4
Missing	3		0	
Marital status				
Not married	1,169	45.4	228	48.9
Married	1,408	54.6	238	51.1
Missing	1		0	
APOE ε4 allele				
None	1,767	76.7	260	63.6
At least one	538	23.3	149	36.4
Missing	273		57	

^a Percentages based on nonmissing values.

Table 2. Characteristics of ACT participants at last known follow-up, according to vital status, whether or not they had withdrawn and, among those that died, whether or not an autopsy was performed

	Alive				Dead				p value ^b
	enrolled		withdrawn		no autopsy		autopsy		
	n	% ^a	n	% ^a	n	% ^a	n	% ^a	
Total	1,711		288		798		247		
Cohort									<0.001
Original	1,106	64.6	271	93.8	590	73.9	160	64.8	
Dementia status									<0.001
Demented	154	9.0	18	6.2	208	26.1	87	35.2	
Age, years									<0.001
≤75	241	14.1	61	21.2	96	12.0	24	9.7	
76–80	535	31.3	86	29.9	180	22.6	40	16.2	
81–85	534	31.2	76	26.4	221	27.7	55	22.3	
85–90	292	17.1	50	17.4	183	22.9	69	27.9	
>90	109	6.4	15	5.2	118	14.8	59	23.9	
Gender									0.001
Female	1,049	61.3	189	65.6	430	53.9	143	57.9	
Missing	1		0		0		0		
Race									0.001
Nonwhite	181	10.6	35	12.2	72	9.0	7	2.8	
Missing	1		0		0		0		
Education									<0.001
<High school	151	8.8	54	18.8	143	17.9	29	11.7	
High school	403	23.6	81	28.1	224	28.1	61	24.7	
Some college	752	44.0	128	44.4	326	40.9	106	42.9	
College graduate	402	23.5	25	8.7	105	13.2	51	20.6	
Missing	3		0		0		0		
Marital status									0.037
Married	773	45.2	139	48.3	321	40.2	102	41.3	
Missing	1		0		0		0		
Depression									0.028
Depressed	211	12.3	49	17.0	128	16.0	37	15.0	
APOE ε4 allele									0.604
None	1,135	74.3	192	77.4	535	75.2	165	72.4	
At least one	392	25.7	56	22.6	176	24.8	63	27.6	
Missing	184		40		87		19		

^a Percentages based on nonmissing values. ^b χ^2 test for association between covariate and four-group classification.

multaneously, with adjustment for cohort membership, age (via a natural smoothing spline 4 degrees of freedom), gender, race, education and presence/absence of any APOE ε4 alleles.

The results indicate statistically significant associations between dementia and Braak stage (V/VI vs. 0–IV; RR 3.06, 95% CI 2.18, 4.29), the number of cerebral microvascular infarcts (>2 vs. ≤2; RR 2.21, 95% CI 1.59, 3.06), and the number of cystic infarcts (≥1 vs. 0; RR 1.41, 95% CI 1.01, 1.97). Although the point estimate for the presence of neocortical Lewy bodies is suggestive of an

effect, the unweighted association was not statistically significant.

Selection Models

To broaden the generalizability of analyses based on the ACT data, one could consider three mechanisms: withdrawal, death and autopsy after death. As such, one way forward would be to develop separate models for the three mechanisms. Each model would provide weights specific to the mechanism, with their combination bridging the gap between the sampling and target

Table 3. NP risk factors according to dementia status at death, together with results from an unweighted log-linear model

	Dementia status				Unweighted analysis ^b		
	nondemented		demented		RR	95% CI	p value
	n	% ^a	n	% ^a			
Total	127		87				
Braak stage							
0–IV	121	95.3	42	48.8	REF		
V or VI	6	4.7	44	51.2	3.06	2.18, 4.29	<0.001
Missing	0		1				
Cerebral microinfarcts							
≤2	118	92.9	62	72.1	REF		
>2	9	7.1	24	27.9	2.21	1.59, 3.06	<0.001
Missing	0		1				
Neocortical Lewy bodies							
0	112	95.7	73	90.1	REF		
≥1	5	4.3	8	9.9	1.84	0.90, 3.79	0.096
Missing	10		6				
Cystic infarcts							
0	93	78.2	56	65.1	REF		
≥1	26	21.8	30	34.9	1.41	1.01, 1.97	0.043
Missing	8		1				

REF = Reference group.

^a Percentages based on nonmissing values. ^b Adjusted for cohort, age, gender, race, education and APOE.

populations. An alternative approach is to directly model the difference between the sampling and target populations, averaging over the three mechanisms. The first approach is appealing in that it adopts a natural perspective towards the underlying selection mechanisms. The second approach is appealing in that it is simpler analytically, requiring fewer modeling assumptions and only a single set of weights. In both cases, however, the aim is to develop a set of weights that permit the interpretation of results from the autopsy sample to the target population (rather than the sample population). As part of future work, we plan on exploring the potential bias-variance trade-off associated with separating out the three models, as well as implications for interpretation of the corresponding target populations. Here, for simplicity, we take the sampling population (i.e. those for whom $R = 1$) to be the ‘dead-autopsy’ participants, while the target population is the ‘alive-enrolled’ population. Consequently, the selection models presented here are for $\pi(\hat{y}) = P(R = 1 | L)$ where $R = 0/1$ denotes membership to the ‘alive-enrolled’/‘dead-autopsy’ subsamples, respectively.

Following the framework of Hernán et al. [6], figure 4 provides a directed acyclic graph indicating various relationships between variables associated with dementia and selection. For clarity, confounders of the NP risk factor-dementia association have been separated into two groups; those that further act as predictors of the selection mechanism denoted C_0 , and those that do not, denoted C_1 . Finally, Z denotes covariates that are risk factors for selection but not for dementia. From figure 4 we see that $L = (Y, C_0, Z)$; it is clear that both of the Hernán et al. [6] criteria apply, indicating potential selection bias.

Table 4 considers two models for selection: a ‘saturated’ model based on the published literature and a ‘sparse’ model, which only includes those covariates in table 4 found to be statistically significant (at the $\alpha = 0.05$ level). The results indicate cohort membership; age (see fig. 5), race and gender are statistically significantly associated with selection (i.e., $R = 1$). Further, dementia status (the outcome for the main analyses) was found to be highly associated with selection, with demented individuals more likely to be in the ‘dead-autopsy’ group with an estimated odds ratio of 5.84 (95% CI 4.16, 8.19).

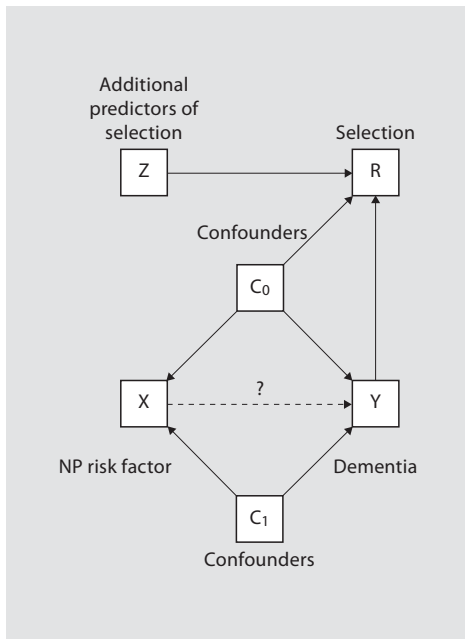


Fig. 4. Directed acyclic graph showing the conditional independencies for the dementia and selection models. Note, NP risk factors (X) are only observed on the autopsy sample; all other covariates are observed on all individuals in ACT. Confounders are split into two groups: those that influence selection (C_0) and those that do not (C_1).

Weighted Analyses

Based on the selection models reported in table 4, we evaluated estimated weights for each of the 214 individuals in the autopsy sample. As outlined above, the introduction of $w_i(\hat{\gamma})$ into the estimating equation serves to up-weight each individual's contribution to create a pseudosample. Note, each of the $w_i(\hat{\gamma})$ weights are greater than 1.0, since the estimated probabilities from the selection model will be between 0 and 1. In this setting the pseudosample may be viewed as a random sample from a hypothetical population of 'Alive-Enrolled' and 'Dead-Autopsy' individuals (i.e. a population for which $R = 0$ or $R = 1$). This is the usual formulation of the pseudosample for missing data problems, where one is interested in recovering the analysis one would have performed had all members of the sample been fully observed. For the analyses presented here, we identified the target population as the 'alive-enrolled' population. To recover an analysis based solely on these individuals, one must adjust the pseudosample to focus on this subgroup (i.e. one for which $R = 1$). This is achieved via an analogous down-weighting using the same fitted probabilities. The com-

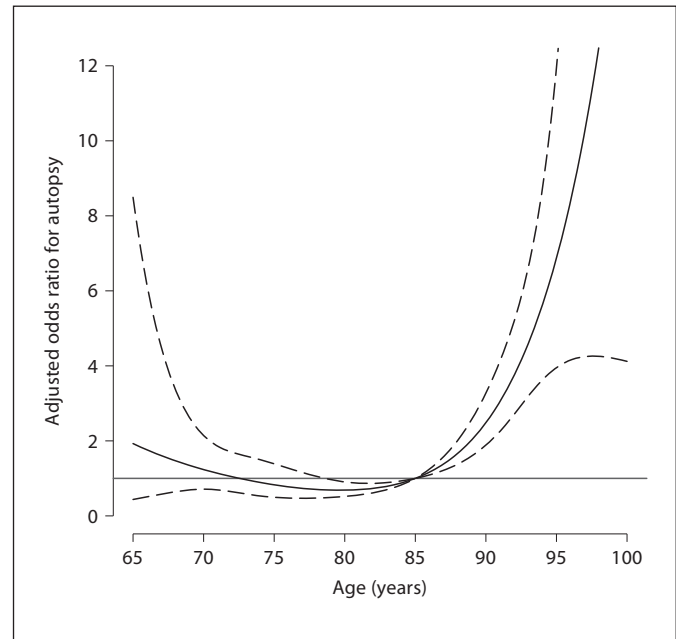


Fig. 5. Estimated odds ratio association (with pointwise 95% CI) between age and risk of 'selection', based on the saturated selection model. The functional form of the association is modeled via a natural smoothing spline (4 degrees of freedom), with age 85 years taken as the referent age group.

bination of the two weights serves to bridge the gap between the sampling and target populations; the corresponding estimating equation is equivalent to equation 5, but with an alternative weight of $\tilde{w}_i(\hat{\gamma}) = (\pi_i(\hat{\gamma})^{-1} - 1)$, rather than $w_i(\hat{\gamma})$.

Table 5 reports the results of the weighted analyses; a comparison with table 3 indicates substantial changes in RR estimates when one adjusts for selection bias using weighted estimating equations. For example, the weighted analysis using the saturated selection model yields a point estimate of 7.81 for Braak stage compared to 3.06 for the unweighted analysis. Similarly, the RR estimate for cerebral microvascular infarcts increased from 2.21 to 4.78. Comparison between the results based on the saturated and sparse selection models indicates robustness of the RR estimates to the specification of the selection model.

Also shown in table 5 are naïve 95% CIs and p values, which ignore estimation of the weights, as well as the BCA bootstrap intervals and p values. The naïve 95% CIs based on the saturated model suggest a statistically significant association for neocortical Lewy bodies and cystic in-

Table 4. Logistic regression selection models, each modeling the probability of being in either the ‘alive-enrolled’ or ‘dead-autopsy’ groups

	Saturated model			Sparse model		
	OR	95% CI	p value	OR	95% CI	p value
Cohort						
Original	REF					
Expansion	0.29	0.18, 0.47	<0.001	0.29	0.18, 0.48	<0.001
Dementia status						
Nondemented	REF			REF		
Demented	5.84	4.16, 8.19	<0.001	5.76	4.12, 8.07	<0.001
Age ^a	see figure 5		<0.001			<0.001
Race						
White	REF			REF		
Nonwhite	0.32	0.14, 0.73	0.006	0.32	0.14, 0.72	0.006
Gender						
Male	REF					
Female	0.68	0.48, 0.95	0.025	0.67	0.49, 0.91	0.011
Education						
High school or less	REF					
Some college	1.22	0.88, 1.69	0.227			
Marital status						
Not married	REF					
Married	1.06	0.75, 1.51	0.749			
Depression						
Not depressed	REF					
Depressed	1.26	0.83, 1.91	0.277			

REF = Reference group.

^a Age modeled via a natural smoothing spline with 4 degrees of freedom.

Table 5. RR estimates based on weighted estimation equations, together with inference using naïve methods (i.e. standard sandwich-based standard errors) and inference using the BCa bootstrap

	RR ^a	Naïve analysis		BCa bootstrap	
		95% CI	p value	95% CI	p value
Saturated selection model					
Braak stage	7.81	4.05, 15.04	<0.001	3.10, 14.75	<0.001
Cerebral microinfarcts	4.78	2.41, 9.51	<0.001	2.07, 10.13	<0.001
Neocortical Lewy bodies	3.63	1.15, 11.43	0.028	0.45, 19.45	0.170
Cystic infarcts	2.12	1.09, 4.12	0.027	0.94, 4.39	0.068
Sparse selection model					
Braak stage	7.87	4.09, 15.15	<0.001	3.24, 14.81	<0.001
Cerebral microinfarcts	4.54	2.34, 8.81	<0.001	2.01, 9.29	<0.001
Neocortical Lewy bodies	3.21	0.86, 11.94	0.082	0.29, 18.00	0.266
Cystic infarcts	2.10	1.09, 4.05	0.026	0.95, 4.26	0.065

^a Adjusted for cohort, age, gender, race, education and APOE.

farcts, in contrast to the unweighted analyses. There is some sensitivity for neocortical Lewy bodies to the selection model, although given the infrequency of the exposure this could reasonably be attributed to a lack of precision. Across all analyses, the bootstrap estimates of uncertainty are greater than those provided by naïve analyses; the increase in width of the bootstrap 95% CIs ranges from approximately 20 to 100%. For both neocortical Lewy bodies and cystic infarcts, the increase in uncertainty for analyses based on the combined original and expansion cohorts leads to the estimated associations no longer being statistically significant. Both Braak stage and cerebral microvascular infarcts remain highly statistically significant.

Discussion

In settings where observed data may not be representative of the target population, Hernán et al. [6] provide a useful framework with which one can establish if selection bias is present and, in particular, if $\beta^* \neq \beta$. Essentially, the characterization reduces to whether or not entry into the study (or equivalently the value of R) is jointly governed by the outcome and exposure of interest (or causes of each). We find distinguishing the (potentially hypothetical) sampling population from the target populations to be a useful, parallel characterization of potential selection bias. A third characterization is to consider the hypothetical study one would have conducted, had there not been any selection. The analyses presented here, for example, could be interpreted as mimicking a hypothetical cross-sectional analysis of individuals currently alive. An alternative would be to view the entire initial ACT cohort as the target population; the corresponding hypothetical study would be a longitudinal study where we observe all the NP measurements at baseline and follow individuals in time, recording incident dementia/AD. We are currently conducting a comprehensive analysis that considers such a hypothetical longitudinal study and moves beyond the cross-sectional associations reported here. This analysis will include separate assessments of the various underlying selection mechanisms (withdrawal, death and consent to autopsy), permitting the evaluation of associations corresponding to the full ACT cohort. Separate models for each of the mechanisms, however, require careful thought and development.

Substantively, we found that adjustment for selection bias strengthened the magnitude of associations observed between NP measurements and risk of dementia. As with

all analyses of autopsy data, however, the interpretation of results is limited by only knowing exposure status after death, and for cases after the outcome has been observed. However, for both AD and dementia there is evidence that the pathogenesis includes preclinical stages suggesting a 'chronic disease' model. For example, studies have shown the presence of abundant Alzheimer neuropathology in asymptomatic elderly individuals at least a decade before AD dementia is common [18, 19], while positron emission tomography studies have shown regional brain hypometabolism in young individuals at risk with a pattern similar to that observed in AD/dementia [20]. Further, recent studies with Pittsburgh Compound-B [(11C)PIB], a positron emission tomography imaging tracer that binds to amyloid plaques in vivo, suggest Alzheimer neuropathology occurs in both the preclinical state and in mild cognitive impairment [21, 22].

The framework and ideas presented here are broadly applicable, and likely useful, in more general settings of studies of neurological disorders [23, 24]. We note, in particular, that the choice and interpretation of R (and its substrata, 0 vs. 1) will depend heavily on the study design, as well as the specific question under investigation. Reidel-Heller et al. [25], for example, examine selection bias associated with recruitment procedures in a study of the prevalence of dementia. Information based solely on face-to-face interviews among community-dwelling individuals yielded a prevalence of 5.3%, whereas incorporating additional information from proxy interviews and institutionalized individuals increased the prevalence to 10.5%. This raises the important point that selection bias must be with respect to some underlying population of interest and that a careful interpretation of estimates may, in fact, indicate no bias. The estimated prevalence of 5.3%, for example, can reasonably be interpreted as pertaining to the rate among community-dwelling individuals; with respect to that population, differences in reported estimates may not be relevant.

Finally, we emphasize that when attempting to adjust for potential selection bias, the MAR assumption is crucial. Indeed there is an important interplay between the question under investigation, the definition of the target population, the MAR assumption and the ability to adequately estimate appropriate weights. Unfortunately, the MAR assumption cannot be verified empirically and its validity relies exclusively on scientific input and knowledge of the target population. For the ACT analyses presented here, we were able to take advantage of previously published work on factors related to selection in autopsy-based studies of dementia, together with the comprehen-

sive data collection of the parent cohort study. More generally, we note that selection bias and traditional confounding present similar challenges in observational studies, both requiring careful thought at the planning stage. We therefore encourage researchers to incorporate, as part of their sampling plans, data collection schemes aimed directly at understanding and characterizing potential sources of selection bias. Practically, this involves a priori identifying the components of L, and ensuring they are observed on all individuals.

Conclusions

Although population-based NP studies of dementia have a strong potential to suffer from selection bias, published reports appear not to have attempted to generalize

their results beyond their respective study samples. This paper reviews and demonstrates methods for identifying potential selection bias, as well as the use of weighted estimating equations to generalize results to a target population of scientific interest. The ACT example presented serves to illustrate some of the key concepts and, given the current state of the literature on autopsy studies of dementia, provides an important step forward. The next few years should see additional developments in this overlooked but important area.

Acknowledgements

Data collection and analyses were supported by NIH grant U01 AG 06781 (E. Larson). The funding agency had no impact on the decision to perform these analyses or to publish the results.

References

- Zaccai J, Ince P, Brayne C: Population-based neuropathological studies of dementia: design, methods and areas of investigation – a systematic review. *BMC Neurol* 2006;6:2.
- Sonnen JA, Larson EB, Crane PK, Haneuse S, Li G, Schellenberg GD, Craft S, Leverenz JB, Montine TJ: Pathological correlates of dementia in a longitudinal, population-based sample of aging. *Ann Neurol* 2007;62:406–413.
- MacClean C, Reed D: Predictors of atherosclerosis in the Honolulu Heart Program. *Am J Epidemiol* 1987;126:226–236.
- Gao S, Hui SL, Hall KS, Hendrie HC: Estimating disease prevalence from two-phase surveys with non-response at the second phase. *Stat Med* 2000;19:2101–2114.
- Tsuang D, Simpson KL, Li G, Barnhart RL, Edland SD, Bowen J, McCormick W, Teri L, Nochlin D, Larson EB, Thompson ML, Leverenz JB: Evaluation of selection bias in an incident-based dementia autopsy case series. *Alzheimer Dis Assoc Disord* 2005;19:67–73.
- Hernán MA, Hernandez-Diaz S, Robins JM: A structural approach to selection bias. *Epidemiology* 2004;15:615–625.
- Kukull WA, Higdon R, Bowen JD, McCormick WC, Teri L, Schellenberg GD, van Belle G, Jolley L, Larson EB: Dementia and Alzheimer disease incidence: a prospective cohort study. *Arch Neurol* 2002;59:1737–1746.
- Teng EL, Hasegawa K, Homma A, Imai Y, Larson E, Graves A, Sugimoto K, Yamaguchi T, Sasaki H, Chiu D, et al: The Cognitive Abilities Screening Instrument (CASI): a practical test for cross-cultural epidemiological studies of dementia. *Int Psychogeriatr* 1994;6:45–58, 62.
- Rothman K, Greenland S: *Modern Epidemiology*, ed 2. Philadelphia, Lippincott Williams & Wilkins, 1998.
- Szklo M: Population-based cohort studies. *Epidemiol Rev* 1998;20:81–90.
- McCullagh P, Nelder J: *Generalized Linear Models*, ed 2. Boca Raton, Chapman & Hall/CRC, 1989.
- Liang K-Y, Zeger S: Longitudinal data analysis using generalized linear models. *Biometrika* 1986;73:13–22.
- Robins JM, Rotnitzky A, Zhao LP: Estimation of regression coefficients when some regressors are not always observed. *J Am Stat Assoc* 1994;89:846–866.
- Little R, Rubin D: *Statistical Analysis with Missing Data*, ed 2. Hoboken, Wiley, 2002.
- Efron B, Tibshirani R: *An Introduction to Bootstrap*. London, Chapman & Hall, 1993.
- Braak H, Braak E: Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol* 1991;82:239–259.
- White L, Petrovitch H, Hardman J, Nelson J, Davis DG, Ross GW, Masaki K, Launer L, Markesbery WR: Cerebrovascular pathology and dementia in autopsied Honolulu-Asia Aging Study participants. *Ann NY Acad Sci* 2002;977:9–23.
- Braak H, Braak E: Evolution of the neuropathology of Alzheimer's disease. *Acta Neurol Scand Suppl* 1996;165:3–12.
- Small BJ, Fratiglioni L, Viitanen M, Winblad B, Backman L: The course of cognitive impairment in preclinical Alzheimer disease: three- and 6-year follow-up of a population-based sample. *Arch Neurol* 2000;57:839–844.
- Silverman DH, Small GW, Chang CY, Lu CS, Kung De Aburto MA, Chen W, Czernin J, Rapoport SI, Pietrini P, Alexander GE, Schapiro MB, Jagust WJ, Hoffman JM, Welsh-Bohmer KA, Alavi A, Clark CM, Salmon E, de Leon MJ, Mielke R, Cummings JL, Kowell AP, Gambhir SS, Hoh CK, Phelps ME: Positron emission tomography in evaluation of dementia: regional brain metabolism and long-term outcome. *JAMA* 2001;286:2120–2127.
- Kemppainen NM, Aalto S, Wilson IA, Nagren K, Helin S, Bruck A, Oikonen V, Kailajarvi M, Scheinin M, Viitanen M, Parkkola R, Rinne JO: PET amyloid ligand [11C]PIB uptake is increased in mild cognitive impairment. *Neurology* 2007;68:1603–1606.
- Mintun MA, Larossa GN, Sheline YI, Dence CS, Lee SY, Mach RH, Klunk WE, Mathis CA, DeKosky ST, Morris JC: [11C]PIB in a nondemented population: potential antecedent marker of Alzheimer disease. *Neurology* 2006;67:446–452.
- Ellenberg JH: Differential postmorbidity mortality in observational studies of risk factors for neurologic disorders. *Neuroepidemiology* 1994;13:187–194.
- Ellenberg JH: Observational data bases in neurological disorders: selection bias and generalization of results. *Neuroepidemiology* 1994;13:268–274.
- Riedel-Heller SG, Schork A, Matschinger H, Angermeyer MC: Recruitment procedures and their impact on the prevalence of dementia. Results from the Leipzig Longitudinal Study of the Aged (LEILA75+). *Neuroepidemiology* 2000;19:130–140.