ALLEN INSTITUTE for BRAIN SCIENCE

Aging, Dementia and TBI Study

TECHNICAL WHITE PAPER: WEIGHTED ANALYSES

OVERVIEW

This Technical White Paper describes how to conduct analyses that examine risk factors other than traumatic brain injury (TBI) with loss of consciousness using the cohort described in this study, given that the specimens here were selected from the Adult Changes in Thought (ACT) study specifically for TBI exposure with age-matched controls. Specifically, the document details the generation of sampling weights that reflect selection into the Allen TBI group from the larger ACT cohort, and provides syntax or advice for several common statistical packages. These weights make it possible to generalize findings to the entire ACT cohort, a community-based sample.

Why Use Weights

Individuals were carefully chosen for this TBI study from the larger ACT cohort on the basis of history of exposure to TBI with loss of consciousness (LOC). This sample of individuals can be considered a matched cohort[1] for purposes of TBI exposure, and no special statistical techniques are needed to analyze TBI with LOC as a risk factor for brain outcomes.

However, that is only one possible use of these data. One may wish to consider associations between brain outcomes and some exposure other than TBI with LOC or some other characteristic entirely, and to generalize these findings to the entire ACT cohort, a community-based sample. Weights will enable this, despite the fact that subjects were chosen for inclusion on the basis of their TBI with LOC exposure status.

Individuals who have autopsy data differ in two ways from the entire ACT cohort: 1) They have all died, and 2) They were willing to consent to have an autopsy. All of the present samples are drawn from the autopsy cohort, so those factors make the members of the group evaluated in this project unrepresentative in terms of the ACT sample as a whole.

Furthermore, the present sample over-represents individuals with TBI with LOC compared to the entire ACT sample (50% in the present sample vs. 17% in the entire ACT sample, see **Table 1**). Furthermore, males have a higher rate of TBI with LOC than females, and sex-matched individuals without exposure to TBI with LOC were selected. Thus, while males make up only 41% of the overall ACT sample, they represent 58% of the present sample.

For the Allen TBI Group, a subset of 55 people with a history of TBI with LOC (TBI-LOC) and rapid autopsy from the ACT autopsy cohort were matched with 55 people without TBI-LOC based on sex, rapid autopsy, age, date of death, and post-mortem interval (in that order). In **Table 1**, the Allen TBI Group is compared with the rest of the cohort and important differences can be seen between the two sample sets.

Table 1. Comparison of Allen TBI Group to the rest of ACT Cohort.

Characteristic	Allen TBI Group Mean (SD) or n (%)	Rest of ACT Cohort Mean (SD) or n (%)	All of ACT Cohort Mean (SD) or n (%)	P-value
Age	86.3 (6.3)	81.2 (7.7)	81.3 (7.7)	<0.001
Education	14.3 (3.2)	14.8 (3.2)	14.8 (3.2)	0.073
Female	46 (42%)	2926 (59%)	2972 (59%)	<0.001
Cohort				<0.001
Expansion	26 (24%)	784 (16%)	810 (16%)	
Replacement	12 (11%)	1,678 (34%)	1,690 (33%)	
Lifetime TBI-LOC	55 (50%)	793 (16%)	848 (17%)	<0.001
Dementia Diagnosis	51 (46%)	1,016 (20%)	1,067 (21%)	<0.001

If one wants to make any inferences that apply to the entire ACT cohort and not just to the 108 in the Allen TBI Group, we need to account for this selection bias[2]. This can be accomplished using an inverse probability weighting approach.

The first step of the inverse probability weighting approach is to determine how probable it is that each individual person would have been selected from the whole of the ACT cohort in terms of demographic and exposure characteristics – this first step addresses how "representative" of the whole cohort each person selected is. In the second step, weights as the inverse of that probability are generated, such that the weighted sample again has characteristics similar to the overall cohort.

The probability of Allen TBI Group membership was modeled using a logistic regression model with the following covariates: TBI-LOC status, age at most recent study visit, sex, education, study entry cohort, and diagnosis of dementia. Model results were used to determine each person's probability of being in the Allen TBI Group cohort. The inverse of this probability is available to be used as the weight in subsequent statistical modeling.

Because the weights are derived from a model that estimates coefficients with error, ideally one would incorporate uncertainty in the weights into subsequent statistical modeling to ensure adequately conservative confidence intervals and tests of significance. This can be accomplished using a bootstrapping procedure[2]. This would require a data request to ACT to obtain the covariates on the full ACT cohort. Data can be requested from ACT by contacting ACTproposals@ghc.org.

Derivation Details

In the ACT cohort, data were available on 5,081 people as of 10/01/15. One person was omitted due to missing data on education, leaving n=5,080.

Weights were computed in Stata (version 14.0) based on the following ACT study variables:

```
allen          In Allen TBI Group
age            Age at last study visit
gender         Sex
education      Number of years of education, with >21 recoded to 21
cohort         Study entry cohort
any_ever       Lifetime history of TBI with LOC
anydementia    DSM-IV diagnosis of dementia as of last study visit
```

Here is the code that was used, where age4 was age categorized as 65-74, 75-82, 83-87 and 88+ years and ed3 was education categorized as 3-12, 13-16 and 17+ years.

```
logistic allen i.age4 gender i.ed3 i.cohort any_ever anydementia
predict w1
replace w1=1/w1
```

The resulting weights range from 3.9 to 232.8, with a mean of 37.2 (SD 38.7). The percentiles are shown in **Table 2**.

Table 2. Sampling weights and percentiles.

Percentiles	Sampling Weight
1%	4.1
5%	4.7
10%	7.8
25%	10.3
50%	23.0
75%	54.9
90%	82.9
95%	112.5
99%	179.7

The participants with the ten largest probability weights are listed in **Table 3**. These individuals are weighted most heavily because their characteristics more closely match those of the rest of the cohort; they have not reported a TBI with LOC, most were not diagnosed with dementia, and those with the highest probability weights are some of the youngest members of the Allen TBI Group.

Table 3. Subjects with the ten largest probability weights.

Probability Weight	Age Range	Sex	Education (years)	Study Entry Cohort	Any TBI-LOC	Dementia Diagnosis
84.2	75 - 82	Female	16+	Expansion	No	No
94.5	83 - 86	Female	13 - 16	Original	No	No
94.5	83 - 86	Female	13 - 16	Original	No	No
98.2	88+	Female	13 - 16	Replacement	No	Yes
112.5	83 - 86	Female	3 - 12	Original	No	No
116.9	75 - 82	Male	13 - 16	Replacement	No	No
121.1	75 - 82	Female	13 - 16	Replacement	No	No
133.4	75 - 82	Female	13 - 16	Original	No	No
179.7	75 - 82	Male	17+	Replacement	No	No
232.8	75 - 82	Female	13 - 16	Replacement	No	No

Participants with the ten smallest probability weights are shown in **Table 4**. They are quite old, all have reported a TBI with LOC, and all were diagnosed with dementia. Their probability weights are low because they are not typical ACT cohort members.

Table 4. Subjects with the ten smallest analytic weights.

Probability Weight	Age Range	Sex	Education (years)	Study Entry Cohort	Any TBI-LOC	Dementia Diagnosis
3.9	88+	Male	13 - 16	Original	Yes	Yes
4.1	83 - 87	Male	13 - 16	Expansion	Yes	Yes
4.4	88+	Male	3 - 12	Original	Yes	Yes
4.4	88+	Male	3 - 12	Original	Yes	Yes
4.5	88+	Male	17+	Expansion	Yes	Yes
4.7	83 - 87	Male	3 - 12	Expansion	Yes	Yes
5.4	88+	Male	17+	Original	Yes	Yes
5.7	83 - 87	Male	3 - 12	Original	Yes	Yes
5.7	83 - 87	Male	3 - 12	Original	Yes	Yes
6.7	88+	Female	13 - 16	Original	Yes	Yes

EXAMPLES OF HOW TO USE THE WEIGHTS

Stata

Here is a sham example showing how to use the probability weights. In this example, diabetes information is missing for 11 people, leaving n=99. First is the unweighted analysis using education, age and diabetes status to predict dementia diagnosis, controlling for study cohort.

```
. logistic anydem education age diabetes i.cohort

Logistic regression                               Number of obs     =         99
                                                  LR chi2(5)        =       8.44
                                                  Prob > chi2       =     0.1335
Log likelihood = -63.260135                       Pseudo R2         =     0.0625

------------------------------------------------------------------------------
 anydementia | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   education |   .8685601   .0621764    -1.97   0.049     .7548595     .999387
         age |   .9747055   .0356907    -0.70   0.484     .9072042    1.047229
    diabetes |   1.710862   1.259649     0.73   0.466      .404115    7.243112
             |
      cohort |
  Expansion  |   2.424595   1.360544     1.58   0.114     .8072227    7.282574
Replacement  |    .359085   .3123212    -1.18   0.239     .0652902    1.974906
             |
       _cons |   41.89113   145.8683     1.07   0.283     .0455166    38554.41
------------------------------------------------------------------------------
```

Here is how to incorporate the probability weights, where w1 is the name of the weight.

```
. logistic anydem education age diabetes i.cohort [pweight=w1]

Logistic regression                               Number of obs     =         99
                                                  Wald chi2(5)      =       8.11
```

```
                                                Prob > chi2       =     0.1501
Log pseudolikelihood = -1777.5103               Pseudo R2         =     0.1126

------------------------------------------------------------------------------
             |               Robust
 anydementia | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   education |   .8388218   .0877952    -1.68   0.093     .6832486    1.029818
         age |   .9988039   .0594018    -0.02   0.984     .8889079    1.122286
    diabetes |   3.576935    3.65719     1.25   0.213      .482182    26.53451
             |
      cohort |
  Expansion  |   2.239621   1.889038     0.96   0.339     .4287629    11.69854
Replacement  |   .3175121   .3396032    -1.07   0.283     .0390243    2.583359
             |
       _cons |    2.77862    16.2725     0.17   0.861     .0000288    268374.1
------------------------------------------------------------------------------
```

Of course this sham analysis would be done with much more power in the full cohort, but it demonstrates how much the sampling weight can affect the results.

SAS

In SAS the syntax would be

```
proc logist descending;
model anydementia=education age diabetes expansion replacement;
weight w1;run;
```

where expansion and replacement are indicator (dummy) variables created from the cohort.

R

In R, the packages srvydesign or ipw can be considered. In other R packages that allow weights, be careful that the weights used are probability weights, not analytic weights.

SPSS

SPSS Base and SPSS Advanced Models do not allow for probability weights, so either one of their other modules or other software needs to be used.

REFERENCES

1. Greenland, S. & Morgenstern, H. Matching and efficiency in cohort studies. *Am J Epidemiol.* **131**, 151-9. (1990).
2. Haneuse, S. et al. Adjustment for selection bias in observational studies with application to the analysis of autopsy data. *Neuroepidemiology.* **32**, 229-39. doi: 10.1159/000197389. Epub 2009 Jan 29. (2009).