PharmaSUG2011 – Paper HS07

Adjusting Analyses of Survey Results using a Predicted Probability of Response

Rob Gately, INNOVUS: Epidemiology, Waltham, MA

ABSTRACT

When conducting surveys, an issue that is typically encountered is a differential response rate for different categories of respondents. For example, women may respond at a higher rate than men or vice-versa. The same could occur across different age groups, individuals in different geographic regions, or any of many possible characteristics describing the population. In this paper I use proc logistic in SAS®, along with a number of other data steps, to demonstrate the creation and use of a predicted probability of response (PPR). The PPR is used to generate an inverse probability of response (IPR) which becomes the weight in the adjusted analyses.

The study for which this analysis takes place is a component of a larger Risk Evaluation and Mitigation Strategies (REMS) examining patients' understanding of the risks associated with the use of a particular medication. The survey asks questions about patients' receipt and understanding of the medication guide associated with the medication. Individuals are 18 years of age or older and recent users of the medication.

INTRODUCTION

Differential response rates to surveys can lead to potential bias of results. Whether or not any bias actually exists, needs to be examined empirically when surveys are conducted. Tate et. al (2007), for example, found differential response rates based on age, gender, and ethnic group, among others, yet did not find any appreciable response bias in their survey of veteran's health risks. One way to identify and account for potential response bias is through the use of an inverse probability of response weight to adjust survey responses (Cole and Hernán: 2008; Haneuse et. al: 2009). According to Hernán and Robins (2006), standardization and inverse weighting may generate similar estimates, but in practical application estimates derived from these two approaches may differ because they are based on different modeling assumptions.

In this paper I use logistic regression, including all subjects invited to participate in the survey, with the dependent variable being whether or not the subject responded to the survey, to estimate a predicted probability of response to the survey for each person. The predictor variables will include the covariates. All predictors will remain in the model regardless of statistical significance. The predicted probability of response will be used to inversely weight the responses to the individual questions.

DATA SETUP

Create dummy variables for each of the categorical variables. For example, there are 4 age categories – create 3 dummy variables for age. For the categorical variables, in most cases we will use the first level as the reference category. In some cases there are no observations within the first category, so another category is chosen.

Sample code creating dummy variables (not all covariates listed)…

```
Label
prior_drug1_yes = 'Baseline drug1, yes'
prior_drug1_no = 'Baseline drug1, no'

office1 = '0-1 outpatient office visits'
office2 = '2-3 outpatient office visits'
office3 = '>=4 outpatient office visits'

meds1 = '0-3 prescriptions'
meds2 = '4-6 prescriptions'
meds3 = '>=7 prescriptions'
```

```
cost1 = '<$2,000'
cost2 = '$2001-$5,000'
cost3 = '>$5,000'

surv_lag_cat1 = 'Survey lag of 0 to 91 days'
surv_lag_cat2 = 'Survey lag of 92 to 183 days'
surv_lag_cat3 = 'Survey lag of 184 to 365 days'
surv_lag_cat4 = 'Survey lag of 366 or more days'
;


if prior_drug1 = 1 then prior_drug1_yes = 1;
      else prior_drug1_yes = 0;
if prior_drug1 = 0 then prior_drug1_no = 1;
      else prior_drug1_no = 0;

if numvis_cat = 1 then office1 = 1;
      else office1 = 0;
if numvis_cat = 2 then office2 = 1;
      else office2 = 0;
if numvis_cat = 3 then office3 = 1;
      else office3 = 0;

if numrx_cat = 1 then meds1 = 1;
      else meds1 = 0;
if numrx_cat = 2 then meds2 = 1;
      else meds2 = 0;
if numrx_cat = 3 then meds3 = 1;
      else meds3 = 0;

if cost_cat = 1 then cost1 = 1;
      else cost1 = 0;
if cost_cat = 2 then cost2 = 1;
      else cost2 = 0;
if cost_cat = 3 then cost3 = 1;
      else cost3 = 0;

if surv_lag_cat = 1 then surv_lag_cat1 = 1;
      else surv_lag_cat1 = 0;
if surv_lag_cat = 2 then surv_lag_cat2 = 1;
      else surv_lag_cat2 = 0;
if surv_lag_cat = 3 then surv_lag_cat3 = 1;
      else surv_lag_cat3 = 0;
if surv_lag_cat = 4 then surv_lag_cat4 = 1;
      else surv_lag_cat4 = 0;
```

MODEL

1. Run logistic regression on the entire dataset of subjects who were invited to participate in the survey.
    a. Let y = 1 if responded to the survey; y = 0 if did not respond to the survey.
    b. Let y be the dependent variable in the logistic regression.
    c. Include the covariate variables for the independent variables.
    d. There may be problems with the model converging when running the logistic regression…
        i) may need to exclude or collapse some of the categories

2. Using the final logistic regression model developed in Step 1, calculate the predicted probability of response to the survey for each individual.

    Predicted probability of response = b0 + b1*age group 1 + b2*age group 2 + ...+ b5*sex + b6+educ1 + b7*educ2
    + …+ b9*package1 + b10*package2 + b11*package3

Sample code showing logistic regression…

```
proc logistic data=tablefile1;
  model response_survey(event='1') =

  /*agegrp1 - ref*/  agegrp2 agegrp3 agegrp4
  /*male - ref*/     female
/*  African_American  - ref*/
/* African_American_Hispanic - N = 0 */
Hispanic_Only
Caucasian
Caucasian_Hispanic
Asian
/* Asian_Hispanic  - N = 0 */
Multiracial_Other
Race_Unavailable
/*  nohsgrad - ref */ hsgrad colgrad Educ_Unavailable
/*region1 - ref*/   region2 region3 region4
/* timehealth1 - N=0  timehealth2 - ref*/ timehealth3 timehealth4
/*packagetype1 - ref*/    packagetype2 packagetype3 packagetype4
/*prior_drug1_yes - ref*/  prior_drug1_no
/*timesincedrug11 - ref*/ timesincedrug12 timesincedrug13 /* timesincedrug14 */ /*omit
category 4 since it is the same as prior_drug1_no */
heartdis
lungcan
othercan
/*office1 - ref*/ office2 office3
/*meds1 - ref*/   meds2 meds3
 /* surv_lag_cat1 - N=0  surv_lag_cat2 - ref*/ surv_lag_cat3 surv_lag_cat4      ;

output out=t2 p=prob;
run;
```

The LOGISTIC Procedure

Model Information

Data Set	WORK.TABLEFILE1	
Response Variable	response_survey	Response to Survey
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	3705
Number of Observations Used	3705

Response Profile

Ordered Value	response_survey	Total Frequency
1	0	3253
2	1	452

Probability modeled is response_survey='1'.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	2750.263	2685.720
SC	2756.480	2897.113
-2 Log L	2748.263	2617.720

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	130.5426	33	<.0001
Score	130.4102	33	<.0001
Wald	121.5729	33	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-2.7578	0.8953	9.4884	0.0021
agegrp2	1	0.4609	0.2363	3.8036	0.0511
agegrp3	1	0.7610	0.2409	9.9783	0.0016
agegrp4	1	1.4124	0.3607	15.3304	<.0001
female	1	0.4107	0.1062	14.9641	0.0001
Hispanic_Only	1	-0.3426	0.4629	0.5478	0.4592
Caucasian	1	0.3028	0.2416	1.5713	0.2100
Caucasian_Hispanic	1	-0.1090	1.0781	0.0102	0.9195
Asian	1	0.4178	0.4827	0.7491	0.3868
Multiracial_Other	1	0.3553	0.3012	1.3911	0.2382
Race_Unavailable	1	0.6542	0.6821	0.9200	0.3375
hsgrad	1	-0.1595	0.7745	0.0424	0.8368
colgrad	1	-0.6154	0.7855	0.6138	0.4334
Educ_Unavailable	1	-0.6609	0.9951	0.4412	0.5066
region2	1	0.1189	0.2386	0.2483	0.6183
region3	1	0.00915	0.1371	0.0045	0.9468
region4	1	0.00957	0.1796	0.0028	0.9575
timehealth3	1	0.0847	0.1797	0.2221	0.6375
timehealth4	1	0.1279	0.1536	0.6927	0.4053
packagetype2	1	0.3869	0.1619	5.7128	0.0168
packagetype3	1	0.2438	0.1819	1.7964	0.1801
packagetype4	1	0.5860	0.3057	3.6745	0.0552
prior_drug1_no	1	-0.3360	0.2924	1.3201	0.2506
timesincedrug12	1	-0.0670	0.2958	0.0513	0.8208
timesincedrug13	1	0.1349	0.2720	0.2459	0.6200
heartdis	1	-0.0747	0.2219	0.1133	0.7365
lungcan	1	-12.6984	348.3	0.0013	0.9709
othercan	1	0.6788	0.2412	7.9213	0.0049
office2	1	0.00277	0.1250	0.0005	0.9823
office3	1	-0.1116	0.1512	0.5447	0.4605
meds2	1	-0.1022	0.1313	0.6055	0.4365
meds3	1	-0.1095	0.1481	0.5461	0.4599
surv_lag_cat3	1	-0.1600	0.1204	1.7655	0.1839
surv_lag_cat4	1	-0.3596	0.1389	6.7053	0.0096

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
agegrp2	1.586	0.998	2.520
agegrp3	2.140	1.335	3.432
agegrp4	4.106	2.025	8.326
female	1.508	1.225	1.857
Hispanic_Only	0.710	0.287	1.759
Caucasian	1.354	0.843	2.173
Caucasian_Hispanic	0.897	0.108	7.418
Asian	1.519	0.590	3.911
Multiracial_Other	1.427	0.790	2.574
Race_Unavailable	1.924	0.505	7.323
hsgrad	0.853	0.187	3.890
colgrad	0.540	0.116	2.520
Educ_Unavailable	0.516	0.073	3.630
region2	1.126	0.706	1.798
region3	1.009	0.771	1.320
region4	1.010	0.710	1.436
timehealth3	1.088	0.765	1.548
timehealth4	1.136	0.841	1.536
packagetype2	1.472	1.072	2.022
packagetype3	1.276	0.893	1.823
packagetype4	1.797	0.987	3.271
prior_drug1_no	0.715	0.403	1.268
timesincedrug12	0.935	0.524	1.670
timesincedrug13	1.144	0.672	1.950
heartdis	0.928	0.601	1.434
lungcan	<0.001	<0.001	>999.999
othercan	1.971	1.229	3.163
office2	1.003	0.785	1.281
office3	0.894	0.665	1.203
meds2	0.903	0.698	1.168
meds3	0.896	0.670	1.198
surv_lag_cat3	0.852	0.673	1.079
surv_lag_cat4	0.698	0.532	0.916

Association of Predicted Probabilities and Observed Responses

Percent Concordant	66.0	Somers' D	0.329
Percent Discordant	33.1	Gamma	0.332
Percent Tied	0.9	Tau-a	0.070
Pairs	1470356	c	0.664

3. Use the predicted probability of response to inversely weight the responses to the individual questions. For example, if the predicted probability of responding to the questionnaire given the subject's characteristics is 0.50, then the subject's response to a given question will be weighted by a factor of (1/0.50) = 2.

Sample code showing calculation of the weight (inverse probability of response)…

```
proc sort data=t2;
  by indv_id;
run;

data t3;
  set t2;
wght = 1/prob;
drop _level_;
run;
```

```
proc print data=t3(obs=20);
title "*** file: t3.sas7bdat, first 20 obs, check predicted_prob and wght ***";
var indv_id prob wght;
run;
title;
```

Obs	indv_id	prob	wght
1	xxxxxxxx	0.17554	5.6965
2	xxxxxxxx	0.10018	9.9825
3	xxxxxxxx	0.07022	14.2401
4	xxxxxxxx	0.07037	14.2111
5	xxxxxxxx	0.06056	16.5113
6	xxxxxxxx	0.07874	12.7003
7	xxxxxxxx	0.07974	12.5410
8	xxxxxxxx	0.17563	5.6939
9	xxxxxxxx	0.09590	10.4270
10	xxxxxxxx	0.11813	8.4650
11	xxxxxxxx	0.14817	6.7490
12	xxxxxxxx	0.08430	11.8619
13	xxxxxxxx	0.15467	6.4652
14	xxxxxxxx	0.12415	8.0551
15	xxxxxxxx	0.14858	6.7302
16	xxxxxxxx	0.23727	4.2147
17	xxxxxxxx	0.10690	9.3550
18	xxxxxxxx	0.04039	24.7577
19	xxxxxxxx	0.19842	5.0399
20	xxxxxxxx	0.21068	4.7464

WEIGHT SURVEY RESPONSES

1. Calculate the weighted response for each survey question.

Sample code showing calculation of weighted responses…

```
data tablefile3;
  set tablefile2;
format Q: 7.0;

array aaa (*)
Q1a Q1b Q1c Q1d Q1e Q1f Q1g Q1h Q1z
Q2a Q2b Q2c Q2d Q2e Q2f Q2g Q2h Q2i Q2j Q2z
Q3a Q3b Q3c Q3d Q3e Q3f Q3g Q3h Q3i Q3j Q3z
Q4a Q4b Q4c Q4d Q4e Q4f Q4g Q4h Q4i Q4j Q4k Q4z
Q5a Q5b Q5c Q5d Q5e Q5f Q5g Q5z
Q6a Q6b Q6c Q6d Q6e Q6f Q6z
Q7a Q7b Q7z
Q8a Q8b Q8c Q8z
Q9a Q9b Q9c Q9d Q9e Q9z
Q10a Q10b Q10z;

do i=1 to dim(aaa);
if aaa(i) NE . then do;
       aaa(i)= aaa(i)*wght;
       end;
end;
drop i;
run;

proc print Data = tablefile3 (obs=30) width=min;
title "*** file: tablefile3, first 30 obs, select variables ***";
title2 "*** after applying weight ***";
```

```
var Q: wght;
run;
title;
title2;
```

Obs	q1_correct	q2_correct	q3_correct	Q1a	Q1b	Q1c	Q1d	Q1e	Q1f	Q1g	Q1h	Q2a	Q2b
1	3	2	3	0	0	8	8	8	0	0	0	0	8
2	2	1	3	0	0	0	8	0	8	0	0	0	8
3	2	3	3	0	0	19	0	19	0	0	0	0	19
4	4	2	3	4	0	4	4	4	0	0	0	0	4
5	0	0	0	0	0	0	0	0	0	0	29	0	0
6	.	.	.	.	.	.	.	.	.	.	.	.	.
7	1	1	3	0	0	0	0	7	0	0	0	0	7
8	0	1	1	0	0	0	0	0	0	6	0	0	6

Obs	Q2c	Q2d	Q2e	Q2f	Q2g	Q2h	Q2i	Q2j	Q3a	Q3b	Q3c	Q3d	Q3e	Q3f	Q3g	Q3h	Q3i	Q3j
1	8	8	0	0	8	0	0	0	8	8	8	8	8	8	0	0	0	0
2	0	0	0	0	0	0	0	0	0	8	0	8	0	8	8	0	0	0
3	0	19	0	19	0	0	0	0	0	19	0	19	0	19	0	0	0	0
4	0	4	0	0	0	0	0	0	0	4	0	4	0	4	4	0	0	0
5	0	0	0	0	0	0	0	29	0	0	0	0	0	0	0	0	0	29
6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
7	0	0	0	0	7	0	0	0	0	7	0	7	0	7	0	0	0	0
8	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0

Obs	Q4a	Q4b	Q4c	Q4d	Q4e	Q4f	Q4g	Q4h	Q4i	Q4i_ specify	Q4j	Q4k	Q5a	Q5b	Q5c	Q5d	Q5e
1	0	8	0	0	8	8	0	0	0		0	0	0	8	0	0	0
2	0	8	0	0	0	8	0	8	0		0	0	8	0	0	0	0
3	19	0	0	0	19	0	0	0	0		0	0	19	0	0	19	0
4	0	4	0	0	0	0	0	4	0		0	0	4	0	0	0	0
5	0	29	29	0	0	0	0	0	0		0	0	29	0	0	0	0
6	.	.	.	.	.	.	.	.	.		.	.	.	.	.	.	.
7	0	0	7	0	0	0	0	0	0		0	0	7	0	0	0	0
8	6	6	0	0	6	0	0	0	0		0	0	6	0	0	6	0

Obs	Q5f	Q5f_ specify	Q5g	Q6a	Q6b	Q6c	Q6d	Q6d_ specify	Q6e	Q6f	Q7a	Q7b	Q8a	Q8b	Q8c	Q9a
1	0		0	0	0	0	0		8	0	0	8	0	0	8	0
2	0		0	0	0	0	0		8	0	0	8	8	0	0	0
3	0		0	0	0	0	0		19	0	0	19	19	0	0	19
4	0		0	0	0	0	0		4	0	0	4	4	0	0	0
5	0		0	0	0	0	0		29	0	0	29	0	0	29	0
6	.		.	.	.	.	.		.	.	.	.	.	.	.	.
7	0		0	0	0	0	0		7	0	0	7	0	7	0	0
8	0		0	6	0	0	0		0	0	0	6	6	0	0	6

Obs	Q9b	Q9c	Q9d	Q9e	Q10a	Q10b	Q1z	Q2z	Q3z	Q4z	Q5z	Q6z	Q7z	Q8z	Q9z	Q10z	wght
1	0	0	0	8	0	8	0	0	0	0	0	0	0	0	0	0	8.0551
2	0	8	0	0	0	8	0	0	0	0	0	0	0	0	0	0	8.3460
3	0	0	0	0	0	19	0	0	0	0	0	0	0	0	0	0	19.4098
4	4	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	3.8931
5	0	0	0	29	0	29	0	0	0	0	0	0	0	0	0	0	28.6223
6	.	.	.	.	.	.	7	7	7	7	7	7	7	7	7	7	6.9778
7	0	7	0	0	0	7	0	0	0	0	0	0	0	0	0	0	6.9270
8	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	6.4353

RUN ADJUSTED ANALYSES

1. Set up empty dataset that will be used to append adjusted results for each question response.

```
data adjusted_1;
format name $10. value_adj 20.2 percent_adj 5.2;
name = '';
value_adj = .;
percent_adj = .;
run;
```

2. Use Macro to run adjusted analyses and output table.

```
%macro sumweights(q);

proc summary data=tablefile3 nway nmiss;
class response_survey;
where &q NE .;
var &q wght;
output out=&q.wdata sum=;
run;

proc print data=&q.wdata noobs;
*var response_survey &q. wght;
*sum wght;
title1 "Weighted Responses - &q.";
run;

data &q.wdata_2 (rename =(&q=value_adj));
name = "&q.";
format percent_adj 5.2;
set &q.wdata;
where response_survey = 1;
percent_adj = (&q / wght) * 100;
run;

proc print data=&q.wdata_2 noobs;
var name value_adj percent_adj;
format value_adj percent_adj 20.2;
title1 "Weighted Responses - &q. - reduced";
run;

proc append base = adjusted_1
             data = &q.wdata_2 force nowarn;
run;


%mend sumweights;
```

(Only question 1 shown here for space considerations, but this Macro is run for all questions)

```
%sumweights(Q1a);
%sumweights(Q1b);
%sumweights(Q1c);
%sumweights(Q1d);
%sumweights(Q1e);
%sumweights(Q1f);
%sumweights(Q1g);
%sumweights(Q1h);
%sumweights(Q1z);
```

```
                       Weighted Responses - Q1a
                  response_
```

```
 survey      _TYPE_    _FREQ_        Q1a      wght

    1           1        446        1162    3711.23
        Weighted Responses - Q1a - reduced
name               value_adj             percent_adj

Q1a                  1162.39                   31.32
        Weighted Responses - Q1b
response_
 survey      _TYPE_    _FREQ_        Q1b      wght

    1           1        446        1170    3711.23
        Weighted Responses - Q1b - reduced
name               value_adj             percent_adj

Q1b                  1170.40                   31.54
        Weighted Responses - Q1c
response_
 survey      _TYPE_    _FREQ_        Q1c      wght

    1           1        446        1185    3711.23
        Weighted Responses - Q1c - reduced
name               value_adj             percent_adj

Q1c                  1185.33                   31.94
```

(some data steps omitted here – just setting up and ordering data for final merge with the crude results)

```
proc print data=adjusted_2a;
title1 "Check adjusted data before final merge";
run;
```

Obs	varorder	name	value_adj	percent_adj
1	1	Q1a	1162.39	31.32
2	2	Q1b	1170.40	31.54
3	3	Q1c	1185.33	31.94
4	4	Q1d	1165.87	31.41
5	5	Q1e	2633.51	70.96
6	6	Q1f	1039.76	28.02
7	7	Q1g	325.86	8.78
8	8	Q1h	452.58	12.19
9	9	Q1z	40.67	1.08
10	10	Q2a	147.14	3.96
11	11	Q2b	3008.53	81.00
12	12	Q2c	414.43	11.16
13	13	Q2d	770.80	20.75
14	14	Q2e	26.01	0.70
15	15	Q2f	617.30	16.62
16	16	Q2g	760.54	20.48
17	17	Q2h	78.68	2.12
18	18	Q2i	378.75	10.20
19	19	Q2j	194.24	5.23
20	20	Q2z	37.58	1.00
21	21	Q3a	601.51	16.22
22	22	Q3b	3029.59	81.69
23	23	Q3c	929.55	25.06
24	24	Q3d	1725.16	46.52
25	25	Q3e	765.15	20.63
26	26	Q3f	1367.69	36.88
27	27	Q3g	416.54	11.23
28	28	Q3h	533.22	14.38
29	29	Q3i	186.78	5.04

```
            30       30       Q3j                          261.87      7.06
            31       31       Q3z                           43.08      1.15
```

```
proc print data = a.table5_temp (obs=80);
title 'Check table5_temp';
var name n1 num_all pcnt_all low hi value_adj percent_adj;
run;
title;
```

Crude and weighted responses to survey questions

Obs	NAME	n1	num_all	pcnt_all	low	hi	value_adj	percent_adj
1	Q1a	446	142	31.84	27.69	36.30	1162.39	31.32
2	Q1b	446	156	34.98	30.70	39.51	1170.40	31.54
3	Q1c	446	145	32.51	28.33	36.99	1185.33	31.94
4	Q1d	446	144	32.29	28.11	36.76	1165.87	31.41
5	Q1e	446	320	71.75	67.40	75.73	2633.51	70.96
6	Q1f	446	121	27.13	23.21	31.44	1039.76	28.02
7	Q1g	446	39	8.74	6.46	11.73	325.86	8.78
8	Q1h	446	45	10.09	7.63	13.23	452.58	12.19
9	Q1z	452	6	1.33	0.61	2.87	40.67	1.08
10	Q2a	446	23	5.16	3.46	7.62	147.14	3.96
11	Q2b	446	364	81.61	77.75	84.93	3008.53	81.00
12	Q2c	446	46	10.31	7.82	13.48	414.43	11.16
13	Q2d	446	99	22.20	18.59	26.28	770.80	20.75
14	Q2e	446	5	1.12	0.48	2.60	26.01	0.70
15	Q2f	446	85	19.06	15.68	22.96	617.30	16.62
16	Q2g	446	88	19.73	16.30	23.68	760.54	20.48
17	Q2h	446	11	2.47	1.38	4.36	78.68	2.12
18	Q2i	446	47	10.54	8.02	13.73	378.75	10.20
19	Q2j	446	22	4.93	3.28	7.36	194.24	5.23
20	Q2z	452	6	1.33	0.61	2.87	37.58	1.00
21	Q3a	444	74	16.67	13.49	20.42	601.51	16.22
22	Q3b	444	360	81.08	77.18	84.45	3029.59	81.69
23	Q3c	444	109	24.55	20.78	28.76	929.55	25.06
24	Q3d	444	220	49.55	44.92	54.18	1725.16	46.52
25	Q3e	444	91	20.50	17.00	24.50	765.15	20.63
26	Q3f	444	174	39.19	34.76	43.80	1367.69	36.88
27	Q3g	444	53	11.94	9.24	15.28	416.54	11.23
28	Q3h	444	62	13.96	11.05	17.50	533.22	14.38
29	Q3i	444	26	5.86	4.03	8.44	186.78	5.04
30	Q3j	444	27	6.08	4.21	8.70	261.87	7.06
31	Q3z	452	8	1.77	0.90	3.45	43.08	1.15
32	Q4a	446	223	50.00	45.38	54.62	1813.40	48.91
33	Q4b	446	317	71.08	66.70	75.09	2659.47	71.73
34	Q4c	446	87	19.51	16.10	23.44	634.00	17.10
35	Q4d	446	23	5.16	3.46	7.62	182.16	4.91
36	Q4e	446	122	27.35	23.42	31.67	1020.64	27.53
37	Q4f	446	116	26.01	22.15	30.27	984.72	26.56
38	Q4g	446	22	4.93	3.28	7.36	156.90	4.23
39	Q4h	446	124	27.80	23.85	32.14	1147.72	30.95
40	Q4i	446	10	2.24	1.22	4.08	134.96	3.64
41	Q4j	446	7	1.57	0.76	3.20	53.74	1.45
42	Q4k	446	3	0.67	0.23	1.96	21.35	0.58
43	Q4z	452	6	1.33	0.61	2.87	44.09	1.18
44	Q5a	448	380	84.82	81.20	87.85	3118.06	83.67
45	Q5b	448	70	15.63	12.56	19.28	524.30	14.07
46	Q5c	448	75	16.74	13.57	20.48	602.27	16.16
47	Q5d	448	55	12.28	9.55	15.64	478.50	12.84
48	Q5e	448	15	3.35	2.04	5.45	127.83	3.43
49	Q5f	448	10	2.23	1.22	4.06	86.89	2.33
50	Q5g	448	29	6.47	4.54	9.14	265.60	7.13
51	Q5z	452	4	0.88	0.34	2.25	25.32	0.67

52	Q6a	447	159	35.57	31.27	40.11	1350.14	36.28
53	Q6b	447	137	30.65	26.55	35.07	1110.35	29.84
54	Q6c	447	20	4.47	2.91	6.81	219.47	5.90
55	Q6d	447	8	1.79	0.91	3.49	55.41	1.49
56	Q6e	447	180	40.27	35.82	44.88	1498.37	40.27
57	Q6f	447	23	5.15	3.45	7.60	194.95	5.24
58	Q6z	452	5	1.11	0.47	2.56	30.83	0.82
59	Q7a	445	174	39.10	34.68	43.71	1445.24	39.16
60	Q7b	445	271	60.90	56.29	65.32	2245.21	60.84
61	Q7z	452	7	1.55	0.75	3.16	61.45	1.64
62	Q8a	447	210	46.98	42.40	51.61	1689.70	45.40
63	Q8b	447	181	40.49	36.04	45.11	1519.63	40.83
64	Q8c	447	56	12.53	9.77	15.92	512.21	13.76
65	Q8z	452	5	1.11	0.47	2.56	30.35	0.81
66	Q9a	443	200	45.15	40.57	49.80	1560.36	42.26
67	Q9b	443	167	37.70	33.31	42.30	1406.68	38.10
68	Q9c	443	34	7.67	5.54	10.53	287.96	7.80
69	Q9d	443	2	0.45	0.12	1.63	53.70	1.45
70	Q9e	443	40	9.03	6.70	12.06	383.62	10.39
71	Q9z	452	9	1.99	1.05	3.74	59.57	1.59
72	Q10a	440	18	4.09	2.60	6.37	145.13	3.96
73	Q10b	440	422	95.91	93.63	97.40	3521.19	96.04
74	Q10z	452	12	2.65	1.53	4.58	85.58	2.28

CONCLUSIONS

The results of the inverse probability of response weighting and question adjustments show that despite there being differential response rates for some categories of responders, i.e. older females, with some form of cancer other than lung cancer, continuing on the study drug, and with a shorter lag between their index dispensing and the time of survey administration were more likely to respond to the survey, there was not an appreciable bias in the survey results as seen by the similar crude and adjusted results.

REFERENCES

Stephen R. Cole and Miguel A. Hernán (2008). Constructing Inverse Probability Weights for Marginal Structural Models. *American Journal of Epidemiology*. 168(6): 656–664.

S. Haneuse, J. Schildcrout, P. Crane, J. Sonnen, J. Breitner and E. Larson (2009). Adjustment for Selection Bias in Observational Studies with Application to the Analysis of Autopsy Data. *Neuroepidemiology*. 32: 229-239.

Miguel A Hernán and James M Robins (2006). Estimating causal effects from epidemiological data. Journal of Epidemiology and Community Health. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2652882/?tool=pmcentrez. (accessed 21FEB2011).

A Rosemary Tate, Margaret Jones, Lisa Hull, Nicola T Fear, Roberto Rona, Simon Wessely, and Matthew Hotopf (2007). How many mailouts? Could attempts to increase the response rate in the Iraq war cohort study be counterproductive? *BMC Medical Research Methodology*. 7:51 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2140265/pdf/1471-2288-7-51.pdf. (accessed 21FEB2011).

ACKNOWLEDGEMENTS

The author would like to thank all of his colleagues who provided valuable input and comments on the programming and/or written content of this paper.

CONTACT INFORMATION

Rob Gately
950 Winter Street, Suite 3800
Waltham, MA 02451

robert.gately@i3drugsafety.com