

Dataset Integrity Check for the RAND Interstitial Cystitis Epidemiology (RICE) Data Files



Prepared by

RTI International
3040 Cornwallis Road
Research Triangle Park, NC 27709-2194
February, 2013

Revision History

Version	Author/Title	Date	Comments
1.0	Norma Pugh	February, 2013	Original

Table of Contents

Contents

1	Standard Disclaimer	1
2	Study Background	1
3	Archived Datasets.....	2
4	Statistical Methods	2
5	Results	2
6	Conclusions	2
7	References	3
	Attachment A: SAS Code	6
	Attachment B: SAS Output.....	8
	Table A: Variables Used to Replicate Selected Statistics from Results Section of Publication.....	4
	Table B: Comparison of Values Computed in Integrity Check to Reference Article Selected Statistical Values	5

1 Standard Disclaimer

The intent of this DSIC is to provide confidence that the data distributed by the NIDDK repository is a true copy of the study data. Our intent is not to assess the integrity of the statistical analyses reported by study investigators. As with all statistical analyses of complex datasets, complete replication of a set of statistical results should not be expected on a first (or second) exercise in secondary analysis. This occurs for a number of reasons including differences in the handling of missing data, restrictions on cases included in samples for a particular analysis, software coding used to define complex variables, etc. Experience suggests that most discrepancies can ordinarily be resolved by consultation with the study data coordinating center (DCC), however this process is labor-intensive for both DCC and Repository staff. It is thus not our policy to resolve every discrepancy that is observed in an integrity check. Specifically, we do not attempt to resolve minor or inconsequential discrepancies with published results or discrepancies that involve complex analyses, *unless NIDDK Repository staff suspect that the observed discrepancy suggests that the dataset may have been corrupted in storage, transmission, or processing by repository staff*. We do, however, document in footnotes to the integrity check those instances in which our secondary analyses produced results that were not fully consistent with those reported in the target publication.

2 Study Background

Bladder pain syndrome/interstitial cystitis (BPS/IC) is a poorly understood condition that can cause serious disability. As there are no objective signs or tests to confirm the diagnosis, BPS/IC is a clinical diagnosis based on patient reported symptoms [1].

The RICE Study was a telephone survey of United State households, which sought adult women with bladder symptoms or a BPS/IC diagnosis. A two-stage screening process was used to identify subjects who met case definition criteria, and each subject completed a 60-minute interview on the severity and impact of bladder symptoms, health care seeking and demographics. Data collection began in August 2007 and ended in April 2009 [1].

Berry et al. provide the first population based symptom prevalence estimate among United States adult females. They found that BPS/IC “symptoms are widespread among United States women and associated with considerable disability”. Further, their results suggest BPS/IC may be underdiagnosed [1, p. 540].

3 Archived Datasets

Both SAS data files, as provided by the Data Coordinating Center (DCC), are located in the RICE Data folder in the Official Archive. For this replication, all variables were taken from the SAS data file BASELINE_DATA located in the Official Archive. This particular publication does not contain results from FU12MTH_DATA.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Berry et al [1] in The Journal of Urology in August 2011.

To verify the integrity of the BASELINE_DATA data file housed at the repository, selected statistics from the Results section of the paper were computed (Table B). The SAS code for our replication is included in Attachment 1.

5 Results

Selected statistics cover a variety of scores for both high sensitivity and high specificity definitions. Our Table A lists the variables we used in our replication and Table B compares the results calculated from the archived data file to the published results. The results of the replication are similar to published results. Please note, the weighting variable documented as BASE_WT is actually named NEW_WT. Also of note, the REGION variable is not included in this data release.

6 Conclusions

The NIDDK repository is confident that the RICE data files to be distributed are a true copy of the study data.

7 References

1. Berry, S.H., Elliot, M.N., Suttorp, M., Bogart, L.M., Stoto, M.A., Eggers, P., Nyberg, L., Clemens, J. Q. (2011) Prevalence of symptoms of bladder pain syndrome interstitial cystitis among adult females in the United States. *Journal of Urology*. 186(2), 540-544.

Table A: Variables Used to Replicate Selected Statistics from Results Section of Publication

Publication Variable	Variables Used in Replication
Met high sensitivity criteria	randid (all unique values)
Met high specificity criteria	new_hispec = 1
Did not meet high specificity criteria	new_hispec ^ = 1
SF-36 physical functioning score	pf_tag
SF-36 mental health score	agg_mentag
ICSI score	ols_symp
ICPI score	ols_prob
Consulted a physician regarding symptoms	rice4vs5at_n
Number of physicians seen	rice4vs6at_num
Saw a urologist	rice4vs6a
Received a specific diagnosis	rice4vs7
Received a diagnosis of BPS/IC	rice4vs11

*All variables taken from dataset BASELINE_DATA.

Table B: Comparison of Values Computed in Integrity Check to Reference Article Selected Statistical Values

Characteristic	Berry	Integrity Check	Difference
Met high sensitivity criteria, n	3,397	3,397	0
Met high specificity criteria, n	1,469	1,469	0
Did not meet high specificity criteria, n	1,928	1,928	0
Met high specificity criteria:			
SF-36 physical functioning score, mean (stderr)	39.3 (0.3-0.4)	40.7 (0.4)	+1.4 (0)
SF-36 mental health score, mean (stderr)	44.8 (0.3-0.4)	44.7 (0.3)	-0.1 (0)
ICSI score, mean (stderr)	11.5 (0.1)	11.5 (0.1)	0
ICPI score, mean (stderr)	14.2 (0.1)	14.4 (0.1)	+0.2 (0)
Consulted a physician, %	87.1	87.1	0
Number of physicians seen, mean	3.5	3.6	+0.1
Saw a urologist, %	40.4	42.4	+2
Received a specific diagnosis, %	45.8	46.8	+1
Received a diagnosis of BPS/IC, %	9.9	9.3	-0.6
Met high sensitivity criteria but not the high specificity criteria:			
SF-36 physical functioning score, mean (stderr)	42.1 (0.3-0.4)	43.4 (0.3)	+1.3 (0)
SF-36 mental health score, mean (stderr)	45.3 (0.3-0.4)	45.3 (0.2)	0 (-0.1)
ICSI score, mean (stderr)	10.1 (0.1)	10.3 (0.1)	+0.2 (0)
ICPI score, mean (stderr)	12.5 (0.1)	12.9 (0.1)	+0.4 (0)

Attachment A: SAS Code

```

options errorabend nofmterr;
/*****
/*
/* Program: R:\05_Users\Norma\RICE\results.sas
/* Author:  Norma Pugh
/* Date:    February 2013
/* Purpose: Replicate select results.
/*
*****/
/* DATA SOURCE */
libname data '\\samba1.rtp.rti.org\NIDDK\03_Data_And_Tools\Studies\RICE\Official_Archive\Data';

/*****
/* GET DATA */
*****/
data baseline; set data.baseline_data; run;

/*****
/* REPLICATE ANALYSIS RESULTS */
*****/
/* Sequence of population screening: # hi sensitivity, # hi specificity, # not hi specificity */
title'Sequence of population screening: # hi sensitivity, # hi specificity, # not hi
specificity';
proc freq data=baseline; tables new_hispec / list missing; run;

/* Hi Specificity Popn: SF-36 phys func score, SF-36 mental health score, ICSI score, ICPI score
*/
title'Hi Specificity Popn: SF-36 phys func score, SF-36 mental health score, ICSI score, ICPI
score';
proc means data=baseline(where=(new_hispec=1)) n mean stderr;
  var pf_tag agg_mentag ols_symp ols_prob;
run;

/* Hi Specificity Popn: % consult physician */
title'Hi Specificity Popn: % consult physician';
proc freq data=baseline(where=(new_hispec=1));
  tables rice4vs5at_n / list missing;
run;

/* Hi Specificity Popn: # physicians seen */
title'Hi Specificity Popn: # physicians seen';
proc means data=baseline(where=(new_hispec=1)) n mean stderr;
  var rice4vs6at_num;
run;

/* Hi Specificity Popn: % consult urologist, % specific dx, % BPS/IC dx */
title'Hi Specificity Popn: % consult urologist, % specific dx, % BPS/IC dx';
proc freq data=baseline(where=(new_hispec=1));
  tables rice4vs6a rice4vs7 rice4vs11 / list missing;
run;

/* Hi Sensitivity but not Hi Specificity Popn: SF-36 phys func score, SF-36 mental health score,
ICSI score, ICPI score */
title'Hi Sensitivity but not Hi Specificity Popn: SF-36 phys func score, SF-36 mental health
score, ICSI score, ICPI score';

```

RICE

```
proc means data=baseline(where=(new_hispec^=1)) n mean stderr;  
  var pf_tag agg_mentag ols_symp ols_prob;  
run;
```

Attachment B: SAS Output

Sequence of population screening: # hi sensitivity, # hi specificity, # not hi specificity 1
12:10 Thursday, February 14, 2013

The FREQ Procedure

new_hispec: New criteria for High Specificity met/not met

NEW_HISPEC	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	1928	56.76	1928	56.76
1	1469	43.24	3397	100.00

RICE

Hi Specificity Popn: SF-36 phys func score, SF-36 mental health score, ICSI score, ICPI score 2
12:10 Thursday, February 14, 2013

The MEANS Procedure

Variable	Label	N	Mean	Std Error
PF_TAG	NEMC physical functioning T-score (age adjusted)	1469	40.6903766	0.3533996
AGG_MENTAG	NEMC mental health T-score (age adjusted)	1469	44.6634562	0.2895136
OLS_SYMP	D:OLS symptom index	1469	11.5248468	0.0899390
OLS_PROB	D:OLS problem index	1469	14.4120717	0.0932080

RICE

Hi Specificity Popn: % consult physician

3

12:10 Thursday, February 14, 2013

The FREQ Procedure

RICE4VS5at_n: NEVER GONE TO DOCTOR FOR SYMPT

RICE4VS5AT_N	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	1279	87.07	1279	87.07
0	190	12.93	1469	100.00

RICE

Hi Specificity Popn: # physicians seen 4
12:10 Thursday, February 14, 2013

The MEANS Procedure

Analysis Variable : RICE4VS6AT_NUM RICE4VS6at_num: # Drs seen for bladn sympt dx-treatmnt

N	Mean	Std Error
1276	3.5869906	0.1148331

RICE

Hi Specificity Popn: % consult urologist, % specific dx, % BPS/IC dx 5
 12:10 Thursday, February 14, 2013

The FREQ Procedure

RICE4VS6A: Ever visit urologist

RICE4VS6A	Frequency	Percent	Cumulative Frequency	Cumulative Percent
D	4	0.27	4	0.27
0	190	12.93	194	13.21
1	623	42.41	817	55.62
2	652	44.38	1469	100.00

RICE4VS7: Ever specific dx from doctor re bladrr sympt

RICE4VS7	Frequency	Percent	Cumulative Frequency	Cumulative Percent
D	7	0.48	7	0.48
0	190	12.93	197	13.41
1	687	46.77	884	60.18
2	585	39.82	1469	100.00

RICE4VS11: Ever IC dx

RICE4VS11	Frequency	Percent	Cumulative Frequency	Cumulative Percent
D	21	1.43	21	1.43
1	137	9.33	158	10.76
2	1311	89.24	1469	100.00

RICE

Hi Sensitivity but not Hi Specificity Popn: SF-36 phys func score, SF-36 mental health score, 6
12:10 Thursday, February 14, 2013

The MEANS Procedure

Variable	Label	N	Mean	Std Error
PF_TAG	NEMC physical functioning T-score (age adjusted)	1928	43.3702670	0.2986647
AGG_MENTAG	NEMC mental health T-score (age adjusted)	1928	45.2766492	0.2352221
OLS_SYMP	D:OLS symptom index	1928	10.3393845	0.0847939
OLS_PROB	D:OLS problem index	1928	12.8526971	0.0916809