

Dataset Integrity Check for Validation of Bladder Health Instrument for Evaluation in Women (VIEW)

Prepared by NIDDK-CR
September 10, 2024

Contents

1 Standard Disclaimer	2
2 Study Background	2
3 Archived Datasets	2
4 Statistical Methods	2
5 Results	3
6 Conclusions	3
Table A: Variables used to replicate the DCC provided descriptive statistics table	4
Table B: Comparison of values computed in integrity check to DCC provided descriptive statistics table .	5
Attachment A: SAS Code	7

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 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

The Prevention of Lower Urinary Symptoms (PLUS) Research Consortium sought to optimize prevention of lower urinary tract symptoms (LUTS) in women and adolescent females across their life spans. The ability to measure bladder health and key risk and protective factors was crucial to the PLUS mission. To describe and measure the spectrum of bladder health in diverse populations, researchers needed a valid and reliable instrument. Prior to the Validation of Bladder Health Instrument for Evaluation in Women (VIEW) study, the PLUS Consortium's work on the design of a bladder health instrument was a culmination of expert opinion, information from focus groups, and incorporation of previously validated items and language where appropriate, along with cognitive interviews of participants from the general public. The VIEW study prospectively collected data to test and validate bladder health instrument (BHI) items for inclusion in a final bladder health scale (BHS) that assessed the full range of bladder health of women. The study collected data from the general population using mailed surveys, from the clinical population using surveys and in-person evaluations, and from postpartum women.

3 Archived Datasets

A full listing of the archived datasets included in the package can be found in the Roadmap document. All data files, as provided by the Data Coordinating Center (DCC), are located in the VIEW folder in the data package. For this replication, variables were taken from the "survey_forms.sas7bdat" dataset.

4 Statistical Methods

Analyses were performed to replicate descriptive statistics provided by the DCC. To verify the integrity of the data, only descriptive statistics were computed.

5 Results

For the DCC provided descriptive statistics, Table A lists the variables that were used in the replication, and Table B compares the results calculated from the archived data files to the results in the DCC provided table. The results of the replication are within expected variation to the DCC provided results.

6 Conclusions

The NIDDK Central Repository is confident that the VIEW data files to be distributed are a true copy of the study data.

Table A: Variables used to replicate the DCC provided descriptive statistics table

Table Variable	dataset.variable
PAPI	survey_forms.papi_comp
CASI	survey_forms.papi_comp
Age	survey_forms.n2
Gender identity	survey_forms.o14
Highest education completed	survey_forms.o10a survey_forms.o10b survey_forms.o10b2 survey_forms.o10c survey_forms.o10d survey_forms.o10e
Latina, Hispanic, or Spanish origin	survey_forms.o12
Race	survey_forms.o13__1 survey_forms.o13__2 survey_forms.o13__3 survey_forms.o13__4 survey_forms.o13__5 survey_forms.o13__6 survey_forms.o13__7
English as primary language	survey_forms.o11
Sought care for bladder	survey_forms.o7
Parous	survey_forms.t12b
History of any pelvic surgery	survey_forms.t7__2
BMI	survey_forms.t1 survey_forms.t1a survey_forms.t2

Table B: Comparison of values computed in integrity check to DCC provided descriptive statistics table

Characteristics	DCC Provided (n=794)	DSIC (n=794)	Diff. (n=0)
PAPI, n (%)	428 (53.9)	428 (53.9)	0 (0)
CASI, n (%)	366 (46.1)	366 (46.1)	0 (0)
Age, mean (SD)	48.5 (18.4)	48.5 (18.4)	0 (0)
Age by category, n (%)			
18-25 years	89 (11.6)	89 (11.6)	0 (0)
26-44 years	266 (34.7)	266 (34.7)	0 (0)
45-64 years	225 (29.3)	225 (29.3)	0 (0)
≥ 65 years	187 (24.4)	187 (24.4)	0 (0)
Missing	27	27	0
Gender identity, n (%)			
Female/woman	779 (99.4)	779 (99.4)	0 (0)
Trans male/trans man	2 (0.3)	2 (0.3)	0 (0)
Genderqueer/gender nonconforming	1 (0.1)	1 (0.1)	0 (0)
Identify in a different way	2 (0.3)	2 (0.3)	0 (0)
Missing	10	10	0
Highest education completed, n (%)			
High school or less	101 (12.9)	103 (13.1)	2 (0.2)
Some college or Associate degree	217 (27.7)	217 (27.6)	0 (0.1)
Bachelor's degree	226 (28.8)	226 (28.7)	0 (0.1)
Graduate degree	240 (30.6)	240 (30.5)	0 (0.1)
Missing	10	8	2
Latina, Hispanic, or Spanish origin, n (%)			
Not of Latino, Hispanic, or Spanish origin	700 (93.8)	700 (93.8)	0 (0)
Mexican or Mexican American	26 (3.5)	26 (3.5)	0 (0)
Puerto Rican	2 (0.3)	2 (0.3)	0 (0)
Cuban	2 (0.3)	2 (0.3)	0 (0)
Another Latino, Hispanic, or Spanish origin	16 (2.1)	16 (2.1)	0 (0)
Missing	48	48	0
Race, n (%)			
American Indian or Alaska Native	2 (0.3)	2 (0.3)	0 (0)
Another race, ethnicity, or origin	13 (1.7)	13 (1.7)	0 (0)
Asian	31 (4.0)	31 (4.0)	0 (0)
Black or African American	108 (14.0)	108 (14.0)	0 (0)
Middle Eastern or North African	0 (0)	0 (0)	0 (0)
Multiracial	53 (6.9)	53 (6.9)	0 (0)
Native Hawaiian or other Pacific Islander	2 (0.3)	2 (0.3)	0 (0)
White	562 (72.9)	562 (72.9)	0 (0)
Missing	23	23	0
English as primary language, n (%)	763 (97.1)	763 (97.1)	0 (0)
Sought care for bladder, n (%)	236 (30.0)	236 (30.0)	0 (0)
Parous, n (%)	564 (71.0)	562 (70.8)	2 (0.2)
History of any pelvic surgery, n (%)			
Hysterectomy	134 (16.9)	134 (16.9)	0 (0)

Characteristics	DCC Provided (n=794)	DSIC (n=794)	Diff. (n=0)
BMI, mean (SD)	28.5 (7.17)	28.5 (7.18)	0 (0.1)
Healthy weight, n (%)	274 (35.5)	271 (35.3)	3 (0.2)
Obese, n (%)	267 (34.6)	267 (34.8)	0 (0.2)
Overweight, n (%)	223 (28.9)	222 (28.9)	1 (0.0)
Underweight, n (%)	7 (0.9)	7 (0.9)	0 (0)
Missing, n (%)	23	27	4

Attachment A: SAS Code

```
libname view "X:\NIDDK\niddk-dr_studies6\VIEW\private_created_data\Data for DEID\new";
```

```
/******  
/* VIEW DSIC: Constantine et al. */  
/******
```

```
*demographics;  
*papi and gender identity;  
proc freq data=view.survey_forms;  
tables papi_comp o14 ;  
run;
```

```
*age is n2, create categorical age var;  
proc means data=view.survey_forms n mean std;  
var n2;  
run;
```

```
*creating categorical variable;  
data one; set view.survey_forms;  
age_cat = .;  
if n2 >= 18 AND n2 <= 25 then age_cat = 1;  
if n2 >= 26 AND n2 <= 44 then age_cat = 2;  
if n2 >= 45 AND n2 <= 64 then age_cat = 3;  
if n2 >= 65 then age_cat = 4;  
run;
```

```
proc freq data=one;  
tables age_cat;  
run;
```

```
*education is in multiple variables;  
*o10a-o10e;
```

```
proc freq data=view.survey_forms;  
tables o10a o10b o10b2 o10c o10d o10e;  
run;
```

```
data two; set view.survey_forms;  
educ_cat = .;  
if o10a ^= . OR o10b ^= . OR o10b2 ^= . or o10c ^= . then educ_cat = 1;  
if o10d = 1 OR o10d = 2 then educ_cat = 2;  
if o10d = 3 then educ_cat = 3;  
if o10e ^= . then educ_cat = 4;  
run;
```



```

proc freq data=two;
tables educ_cat;
run;

*ethnicity;
proc freq data=two;
tables o12;
run;

*race;
proc freq data=two;
tables o13__1 o13__2 o13__3 o13__3 o13__4 o13__5 o13__6 o13__7;
run;

*creating new race variable;
data three;
set two;
length race_cat $20;
array races[7] o13__1-o13__7;
race_count = 0;
do i = 1 to 7;
if races[i] = 1 then race_count + 1;
end;
if race_count = 1 then do;
do i = 1 to 7;
if races[i] = 1 then race_cat = vname(races[i]);
end;
else if race_count > 1 then race_cat = 'Multiracial';
drop i race_count;
run;

proc freq data=three;
tables race_cat;
run;

*language;
proc freq data=three;
tables o11;
run;

*bladder care;
proc freq data=three;
tables o7;
run;

*parous;
proc freq data=three;
tables t12 t12a t12b t12c t12d t12e t12;

```

```

run;

data four; set three;
parous = .;
if t12b > 0 then parous = 1;
run;

proc freq data=four;
tables parous/missing;
run;

*surgery;
proc freq data=four;
tables t7__2;
run;

*need to convert weight and height to metric;
data five; set four;

*convert height to meters;
height_m = (t1 * 0.3048) + (t1a * 0.0254);

*convert weight to kg;
weight_kg = (t2 * 0.453592);

*cacl BMI;
bmi = weight_kg / (height_m**2);
bmi = round(bmi, 0.1);
run;

proc means data=five n mean std ;
var bmi;
run;

*BMI cat;
data six; set five;
bmi_cat = " ";
if BMI < 18.5 AND BMI > 0 then BMI_cat = "under";
if BMI >=18.5 AND BMI < 25 then BMI_cat = "healthy";
if BMI >=25 AND BMI < 30 then BMI_Cat = "over";
if BMI >=30 then BMI_Cat = "obese";
run;

proc freq data=six;
tables bmi_cat;
run;

```