

# **Dataset Integrity Check for the National Analgesic Nephropathy Study (NANS) 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 .....	8
	Table A: Variables Used to Replicate Table 3, Distribution of 221 ESRD patients according to CT classification and demographic factors .....	4
	Table B: Comparison of Values Computed in Integrity Check to Reference Article Table 3 Values .....	5
	Table C: Variables Used to Replicate Table 4, Regular use of analgesic products with various ingredients among 15 patients with SICK and 204 other patients with ESRD and Table 5, Regular use of aspirin- and acetaminophen-containing products among 12 patients who had SICK and 203 other patients who had ESRD and did not use phenacetin .....	6
	Table D: Comparison of Values Computed in Integrity Check to Reference Article Tables 4 and 5 Values .....	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 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

Previous studies have suggested that the CT scan is a highly reliable tool for the diagnosis of analgesic-associated renal disease. However, none of these studies have included the United States population [1].

NANS is an observational study of incident patients with ESRD, conducted at five centers across the United States. Patients underwent a helical CT scan and detailed questioning about drug history. Criteria were developed to determine whether CT findings (small indented calcified kidneys [SICK]) were linked to analgesic ingestion [1].

Henrich et al. found that the CT finding of SICK was present in only a minority of heavy analgesic users, yielding low sensitivity. Findings of SICK do not occur in a sufficient proportion of heavy analgesic users to render the CT scan a useful tool to detect analgesic-associated kidney injury [1].

### **3 Archived Datasets**

All SAS data files, as provided by the Data Coordinating Center (DCC), are located in the NANS Data folder in the Official Archive. For this replication, all variables were taken from the SAS data files tblCTScanData and tblInterviewData located in the Official Archive.

### **4 Statistical Methods**

Analyses were performed to duplicate results for the data published by Henrich et al [1] in the Journal of the American Society of Nephrology in May 2006.

To verify the integrity of the data files housed at the repository, descriptive statistics according to CT classification and demographic factors were computed, as well as use of analgesic products. The SAS code for our analysis is included in Attachment 1.

### **5 Results**

Table 3 in the publication [1], Distribution of 221 ESRD patients according to CT classification and demographic factors, reports demographic factors by CT classification and overall. 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 results published in Table 3. The results of the replication are similar to published results.

Table 4, Regular use of analgesic products with various ingredients among 15 patients with SICK and 204 other patients with ESRD and Table 5 in the publication [1], Regular use of aspirin- and acetaminophen-containing products among 12 patients who had SICK and 203 other patients who had ESRD and did not use phenacetin, report on analgesic use by phenacetin use. Our Table C lists the variables we used in our replication and Table D compares the results calculated from the archived data file to selected results published in Tables 4 and 5. Again, the results of the replication are similar to published results.

### **6 Conclusions**

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

## **7 References**

1. Henrich, William L. et al, Non-Contrast-Enhanced Computerized Tomography and Analgesic-Related Kidney Disease: Report of the National Analgesic Nephropathy Study; Journal of the American Society of Nephrology May 2006 17: 1472-1480.

**Table A: Variables Used to Replicate Table 3, Distribution of 221 ESRD patients according to CT classification and demographic factors**

<b>Table Variable</b>	<b>Variables Used in Replication</b>
CT classification	tblCTScanData: SubjType; tblInterviewData: sick, debroe
Age	tblInterviewData: age
Gender	tblInterviewData: sex
Race	tblInterviewData: ethnic
Educational level	tblInterviewData: educat
Region	tblInterviewData: region

**Table B: Comparison of Values Computed in Integrity Check to Reference Article Table 3 Values**

Characteristic	SICK			Other Patients			Total		
	Henrich	Integrity Check	Diff	Henrich	Integrity Check	Diff	Henrich	Integrity Check	Diff
Sample size	15	15	0	204	204	0	221*	221	0
Age (yr), n (%)									
35 to 49	1 (7)	1 (7)	0	57 (28)	57 (28)	0	58 (26)	58 (26)	0
50 to 64	3 (20)	3 (20)	0	51 (25)	51 (25)	0	55 (25)	55 (25)	0
65 to 74	5 (33)	5 (33)	0	51 (25)	51 (25)	0	57 (26)	57 (26)	0
≥75	6 (40)	6 (40)	0	45 (22)	45 (22)	0	51 (23)	51 (23)	0
Gender, n (%)									
Male	10 (67)	10 (67)	0	113 (55)	113 (55)	0	124 (56)	124 (56)	0
Female	5 (33)	5 (33)	0	91 (45)	91 (45)	0	97 (44)	97 (44)	0
Race, n (%)									
Black	3 (20)	3 (20)	0	108 (53)	108 (53)	0	111 (50)	111 (50)	0
Other	12 (80)	12 (80)	0	96 (47)	96 (47)	0	110 (50)	110 (50)	0
Educational Level, n (%)									
<8 <sup>th</sup> grade	2 (13)	2 (13)	0	26 (13)	26 (13)	0	28 (13)	28 (13)	0
8 <sup>th</sup> to 11 <sup>th</sup> grade	2 (13)	2 (13)	0	49 (24)	49 (24)	0	52 (24)	52 (24)	0
Completed high school	3 (20)	3 (20)	0	46 (23)	46 (23)	0	49 (22)	49 (22)	0
Vocational college	0	0	0	9 (4)	9 (4)	0	9 (4)	9 (4)	0
Some college	5 (33)	5 (33)	0	49 (24)	49 (24)	0	54 (24)	55 (25)	+1
Completed college	3 (20)	3 (20)	0	25 (12)	25 (12)	0	29 (13)	28 (13)	-1
Region, n (%)									
Chapel Hill, NC (UNC)	3 (20)	3 (20)	0	51 (25)	51 (25)	0	54 (24)	54 (24)	0
Winston-Salem, NC (WFU)	8 (53)	8 (53)	0	53 (26)	53 (26)	0	61 (28)	61 (28)	0
Toledo, OH / Detroit, MI (MCO)	2 (13)	2 (13)	0	36 (18)	36 (18)	0	39 (18)	39 (18)	0
Dallas, TX (DNA)	1 (7)	1 (7)	0	49 (24)	49 (24)	0	51 (23)	51 (23)	0
Portland, OR (OHSU)	1 (7)	1 (7)	0	15 (7)	15 (7)	0	16 (7)	16 (7)	0

\*Includes two patients who met the DeBroe criteria but not SICK.



**Table C: Variables Used to Replicate Table 4, Regular use of analgesic products with various ingredients among 15 patients with SICK and 204 other patients with ESRD and Table 5, Regular use of aspirin- and acetaminophen-containing products among 12 patients who had SICK and 203 other patients who had ESRD and did not use phenacetin**

<b>Table Variable</b>	<b>Variables Used in Replication</b>
CT classification	tblCTScanData: SubjType; tblInterviewData: sick, debroe
Phenacetin status	tblInterviewData: pctdose
Aspirin before index year	tblInterviewData: anyasa
Acetaminophen before index year	tblInterviewData: anyapap
Phenacetin before index year	tblInterviewData: if pctdose>0
All analgesics before index year	tblInterviewData: anydrug
Ibuprofen before index year	tblInterviewData: anyibu

**Table D: Comparison of Values Computed in Integrity Check to Reference Article Tables 4 and 5 Values**

Characteristic	SICK			Other Patients		
	Henrich	Integrity Check	Diff	Henrich	Integrity Check	Diff
*Aspirin before index year, n (%)	11 (73)	11 (73)	0	37 (18)	37 (18)	0
*Acetaminophen before index year, n (%)	5 (33)	5 (33)	0	26 (13)	26 (13)	0
*Phenacetin before index year, n (%)	3 (20)	2 (13)	-1	1 (0.5)	1 (0.5)	0
*All analgesics before index year, n (%)	11 (73)	11 (73)	0	51 (25)	51 (25)	0
*Ibuprofen before index year, n (%)						
Patients who did not use phenacetin						
**Aspirin before index year, n (%)	8 (67)	9 (69)	+1	36 (18)	36 (18)	0
**Acetaminophen before index year, n (%)	2 (17)	3 (23)	+1	25 (12)	25 (12)	0
**All analgesics before index year, n (%)	8 (67)	9 (69)	+1	50 (25)	50 (25)	0

\*Percentages based on 15 patients with SICK and 204 other patients with ESRD.

\*\*Percentages based on 12 patients with SICK and 203 other patients with ESRD, who did not use phenacetin.

## Attachment A: SAS Code

```

options nofmterr;
/*****
/*
/* Program: R:\05_Users\Norma\NANS\table3.sas
/* Author: Norma Pugh
/* Date: February 2013
/* Purpose: Replicate table 3 results.
/*
*****/
/* DATA SOURCE */
libname data
'\\samba1.rtp.rti.org\NIDDK\03_Data_And_Tools\Studies\NANS\Official_Archive\Official_Archive_v2\DATA\SASFi
les';

/*****/
/* FORMATS */
/*****/
/* NOTE: Formats are taken from \\samba1.rtp.rti.org\NIDDK\03_Data_And_Tools\Studies\
NANS\Official_Archive\Official_Archive_v2\DATA\SASFiles\procformat.sas. This format
program, as provided by the DCC, is not valid due to format names that end with
numbers and white space.
*/

proc format;
Value sick

1='sick'
2='nonsick';

Value Gender

1='Male'
2='Female';

Value Race

1='White'
2='Black/African-American'
3='Asian/Pacific Islander'
4='Native American'
5='Something else'
6='Hispanic';

Value Educat

1='Less than 8th grade'
2='8th through 11th grade'
3='12th grade/high school graduate (include GED)'
4='Vocational school'
5='Some college (include 2 year college degree)'
6='4 year college degree'
7='School beyond college'
8='Unknown';

Value region

1='ohsu'
2='unc'
3='wfu'
4='mco'
5='dna';
run;

```

## NANS

```
/******  
/* GET DATA */  
/******  
/* ESRD subjects */  
proc sort data=data.tblCTScanData(where=(SubjType=1)) out=popn; by intvwid; run; /* ESRD population */  
  
/* Demography data */  
proc sort data=data.tblInterviewData out=demog; by intvwid; run;  
  
/* Merge data */  
data table3; merge popn(in=popn drop=sick) demog; by intvwid; if popn; /* Only keep ESRD population */  
if debroe=1 & sick=2 then exclude=1; /* Subjects who met DeBroe  
criteria but not sick */  
if 35<=age<=49 then agegrp='35 to 49'; /* Create age groups */  
  
else if 50<=age<=64 then agegrp='50 to 64';  
else if 65<=age<=74 then agegrp='65 to 74';  
else if age>=75 then agegrp='>=75';  
run;  
  
/******  
/* REPLICATE ANALYSIS RESULTS */  
/******  
/* By group */  
title'By group';  
proc freq data=table3(where=(exclude^=1)); tables sick*(agegrp sex ethnic educat region) / list missing;  
format sick sick. sex gender. ethnic race. educat educat. region region.;  
run;  
  
/* Total */  
title'Total';  
proc freq data=table3; tables agegrp sex ethnic educat region / list missing;  
format sick sick. sex gender. ethnic race. educat educat. region region.;  
run;
```

## NANS

```

options nofmterr;
/*****
/*
/* Program: R:\05_Users\Norma\NANS\table4_5.sas
/* Author: Norma Pugh
/* Date: February 2013
/* Purpose: Replicate table 4 & 5 selected results.
/*
*****/
/* DATA SOURCE */
libname data
'\\samba1.rtp.rti.org\NIDDK\03_Data_And_Tools\Studies\NANS\Official_Archive\Official_Archive_v2\DATA\SASFiles';

/*****/
/* FORMATS */
/*****/
/* NOTE: Formats are taken from \\samba1.rtp.rti.org\NIDDK\03_Data_And_Tools\Studies\
NANS\Official_Archive\Official_Archive_v2\DATA\SASFiles\procformat.sas. This format
program, as provided by the DCC, is not valid due to format names that end with
numbers and white space.
*/

proc format;
Value sick
1='sick'
2='nonsick';
run;

/*****/
/* GET DATA */
/*****/
/* ESRD subjects */
proc sort data=data.tblCTScanData(where=(SubjType=1)) out=popn; by intvwid; run; /* ESRD population */

/* Demography data */
proc sort data=data.tblInterviewData out=demog; by intvwid; run;

/* Merge data */
data table4; merge popn(in=popn drop=sick) demog; by intvwid; if popn; /* Only keep ESRD population */
if debroe=1 & sick=2 then exclude=1; /* Subjects who met DeBroe

criteria but not sick */
if pctdose>0 then anyphen=1; else anyphen=0; /* Create y/n var for phenacetin use */

run;

/*****/
/* REPLICATE ANALYSIS RESULTS */
/*****/
title'Analgesic Use';
proc freq data=table4(where=(exclude^=1)); tables sick*(anyasa anyapap anyphen anydrug anyibu) / list
missing;
format sick sick.;
run;

title'Analgesic Use for those who did not use Phenacetin';
proc freq data=table4(where=(exclude^=1 & anyphen=0)); tables sick*(anyasa anyapap anydrug) / list
missing;
format sick sick.;
run;

```