

# Dataset Integrity Check for The Nephrotic Syndrome Study Network (NEPTUNE)

Prepared by NIDDK-CR  
February 3, 2023

## 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 results provided by the DCC.....	4
Table B1: Comparison of values computed in integrity check to DCC results (Total and Focal Segmental Glomerulosclerosis) .....	5
Table B2: Comparison of values computed in integrity check to DCC results (Minimal Change Disease and Membranous Nephropathy) .....	7
Table B3: Comparison of values computed in integrity check to DCC results (Non-Biopsied Childhood Onset Nephrotic Syndrome and Other).....	9
Table B4: Comparison of values computed in integrity check to DCC results (Unknown) .....	11
Attachment A: SAS Code .....	13

## **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**

Pro non-inflammatory glomerular diseases, including minimal change disease (MCD), focal and segmental glomerulosclerosis (FSGS) and membranous nephropathy (MN), account for approximately 15% of prevalent end-stage renal disease (ESRD) cases in the United States. Development of successful therapeutic interventions is hindered by a limited knowledge of underlying glomerular disease mechanisms. In response to the need for research concerning these conditions, the Nephrotic Syndrome Study Network (NEPTUNE) was established to investigate the underlying disease mechanisms, elucidate pathogenesis, and identify therapeutic targets for clinical trials.

The NEPTUNE cohort study is a prospective, observational study that enrolls children and adults with FSGS, MCD, and MN. Participants have a proteinuria  $\geq 500$  mg/day, estimated from a 24-hour or protein to creatinine ratio urine collection. Participants with sub-nephrotic proteinuria are also included to capture the broad spectrum of clinical presentation.

## **3 Archived Datasets**

All data files, as provided by the Data Coordinating Center (DCC), are located in the NEPTUNE folder in the data package. For this replication, variables were taken from the “patient\_ds.sas7bdat” dataset.

## **4 Statistical Methods**

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

## **5 Results**

Table A lists the variables that were used in the replication, and Tables B1 through B4 compare the results calculated from the archived data files to the results provided by the DCC. The results of the replication are an exact match to the DCC results.

## **6 Conclusions**

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

**Table A:** Variables used to replicate results provided by the DCC

<b>Table Variable</b>	<b>dataset.variable</b>
Age (years) at baseline visit	patient_ds.PAT_Cohort patient_ds.PAT_AgeV2
Sex	patient_ds.PAT_Cohort patient_ds.PAT_Sex
Race	patient_ds.PAT_Cohort patient_ds.PAT_Race
Ethnicity	patient_ds.PAT_Cohort patient_ds.PAT_Hispanic
Gross annual household income at baseline	patient_ds.PAT_Cohort patient_ds.AHHIncome
Urine protein: Creatinine ratio at screening/enrollment	patient_ds.PAT_Cohort patient_ds.upcr_v1
Birth weight (grams)	patient_ds.PAT_Cohort patient_ds.BirthWgt
End stage renal disease or 40% loss of eGFR reached after baseline	patient_ds.PAT_Cohort patient_ds.ESRDorEGFR40

**Table B1:** Comparison of values computed in integrity check to DCC results (Total and Focal Segmental Glomerulosclerosis)

Characteristic	DCC: Total (n=770)	DSIC: Total (n=770)	Diff. (n=0)	DCC: Focal Segmental Glomerulosclerosis (n=194)	DSIC: Focal Segmental Glomerulosclerosis (n=194)	Diff. (n=0)
Age (years) at baseline visit						
Mean (SD)	30.1 (22.56)	30.1 (22.56)	0 (0)	34.8 (21.0)	34.8 (21.0)	0 (0)
Median (IQR)	26.0 (10.0, 49.0)	26.0 (10.0, 49.0)	0 (0, 0)	33.0 (16.0 to 53.0)	33.0 (16.0, 53.0)	0 (0, 0)
Range	0.0 to 84.0	0.0 to 84.0	0, 0	2.0 to 80.0	2.0 to 80.0	0, 0
N	770	770	0	194	194	0
N missing	0	0	0	0	0	0
Sex, n (%)						
Male	453 (59)	453 (59)	0 (0)	116 (60)	116 (60)	0 (0)
Female	317 (41)	317 (41)	0 (0)	78 (40)	78 (40)	0 (0)
Race, n (%)						
Multi-Racial	40 (5)	40 (5)	0 (0)	6 (3)	6 (3)	0 (0)
Asian/Asian American	80 (10)	80 (10)	0 (0)	12 (6)	12 (6)	0 (0)
Black/African American	189 (25)	189 (25)	0 (0)	70 (36)	70 (36)	0 (0)
Native Hawaiian/Other Pacific Islander	6 (1)	6 (1)	0 (0)	3 (2)	3 (2)	0 (0)
White/Caucasian	417 (54)	417 (54)	0 (0)	94 (48)	94 (48)	0 (0)
Unknown	38 (5)	38 (5)	0 (0)	9 (5)	9 (5)	0 (0)
Ethnicity, n (%)						
Unknown	12 (2)	12 (2)	0 (0)	2 (1)	2 (1)	0 (0)
Hispanic or Latino	164 (21)	164 (21)	0 (0)	46 (24)	46 (24)	0 (0)
Not Hispanic or Latino	594 (77)	594 (77)	0 (0)	146 (75)	146 (75)	0 (0)
Gross Annual Household Income at baseline, n (%)						
\$0 - \$19,999	155 (20)	155 (20)	0 (0)	46 (24)	46 (24)	0 (0)
\$20,000 - \$39,999	121 (16)	121 (16)	0 (0)	27 (14)	27 (14)	0 (0)
\$40,000 - \$59,999	83 (11)	83 (11)	0 (0)	25 (13)	25 (13)	0 (0)
\$60,000 - \$79,999	48 (6)	48 (6)	0 (0)	12 (6)	12 (6)	0 (0)
\$80,000 - \$99,999	50 (6)	50 (6)	0 (0)	9 (5)	9 (5)	0 (0)
\$100,000 +	108 (14)	108 (14)	0 (0)	22 (11)	22 (11)	0 (0)
Don't wish to answer	144 (19)	144 (19)	0 (0)	35 (18)	35 (18)	0 (0)
Don't know	61 (8)	61 (8)	0 (0)	18 (9)	18 (9)	0 (0)

Characteristic	DCC: Total (n=770)	DSIC: Total (n=770)	Diff. (n=0)	DCC: Focal Segmental Glomerulosclerosis (n=194)	DSIC: Focal Segmental Glomerulosclerosis (n=194)	Diff. (n=0)
Urine protein: Creatinine ratio at screening/enrollment						
Mean (SD)	6.8 (7.95)	6.8 (7.95)	0 (0)	5.4 (5.96)	5.4 (5.96)	0 (0)
Median (IQR)	4.7 (2.0, 8.9)	4.7 (2.0, 8.9)	0 (0, 0)	3.6 (1.9, 7.1)	3.6 (1.9, 7.1)	0 (0, 0)
Range	0.0 to 59.8	0.0 to 59.8	0, 0	0.0 to 48.3	0.0 to 48.3	0, 0
N	761	761	0	194	194	0
N missing	9	9	0	0	0	0
Birth weight (grams)						
Mean (SD)	3354.8 (721.71)	3354.8 (721.71)	0 (0)	3417.9 (700.91)	3417.9 (700.91)	0 (0)
Median (IQR)	3345.2 (2976.7, 3630.0)	3345.2 (2976.7, 3630.0)	0 (0, 0)	3487.0 (2920.0, 3855.5)	3487.0 (2920.0, 3855.5)	0 (0, 0)
Range	1644.3 to 8164.7	1644.3 to 8164.7	0, 0	2041.2 to 4989.5	2041.2 to 4989.5	0, 0
N	153	153	0	15	15	0
N missing	617	617	0	179	179	0
End stage renal disease or 40% loss of eGFR reached after baseline, n (%)						
Unknown	50 (6)	50 (6)	0 (0)	13 (7)	13 (7)	0 (0)
No	564 (73)	564 (73)	0 (0)	128 (66)	128 (66)	0 (0)
Yes	156 (20)	156 (20)	0 (0)	53 (27)	53 (27)	0 (0)

**Table B2:** Comparison of values computed in integrity check to DCC results (Minimal Change Disease and Membranous Nephropathy)

Characteristic	DCC: Minimal Change Disease (n=172)	DSIC: Minimal Change Disease (n=172)	Diff. (n=0)	DCC: Membranous Nephropathy (n=106)	DSIC: Membranous Nephropathy (n=106)	Diff. (n=0)
Age (years) at baseline visit						
Mean (SD)	22.4 (19.77)	22.4 (19.77)	0 (0)	51.0 (14.46)	51.0 (14.46)	0 (0)
Median (IQR)	15.0 (7.0, 36.0)	15.0 (7.0, 36.0)	0 (0, 0)	54.0 (41.0, 62.0)	54.0 (41.0, 62.0)	0 (0, 0)
Range	2.0 to 80.0	2.0 to 80.0	0, 0	13.0 to 80.0	13.0 to 80.0	0, 0
N	172	172	0	106	106	0
N missing	0	0	0	0	0	0
Sex, n (%)						
Male	93 (54)	93 (54)	0 (0)	67 (63)	67 (63)	0 (0)
Female	79 (46)	79 (46)	0 (0)	39 (37)	39 (37)	0 (0)
Race, n (%)						
Multi-Racial	13 (8)	13 (8)	0 (0)	3 (3)	3 (3)	0 (0)
Asian/Asian American	22 (13)	22 (13)	0 (0)	6 (6)	6 (6)	0 (0)
Black/African American	41 (24)	41 (24)	0 (0)	15 (14)	15 (14)	0 (0)
Native Hawaiian/Other Pacific Islander	1 (1)	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)
White/Caucasian	89 (52)	89 (52)	0 (0)	78 (74)	78 (74)	0 (0)
Unknown	6 (3)	6 (3)	0 (0)	4 (4)	4 (4)	0 (0)
Ethnicity, n (%)						
Unknown	3 (2)	3 (2)	0 (0)	0 (0)	0 (0)	0 (0)
Hispanic or Latino	34 (20)	34 (20)	0 (0)	22 (21)	22 (21)	0 (0)
Not Hispanic or Latino	135 (78)	135 (78)	0 (0)	84 (79)	84 (79)	0 (0)
Gross Annual Household Income at baseline, n (%)						
\$0 - \$19,999	34 (20)	34 (20)	0 (0)	26 (25)	26 (25)	0 (0)
\$20,000 - \$39,999	29 (17)	29 (17)	0 (0)	20 (19)	20 (19)	0 (0)
\$40,000 - \$59,999	18 (10)	18 (10)	0 (0)	12 (11)	12 (11)	0 (0)
\$60,000 - \$79,999	7 (4)	7 (4)	0 (0)	7 (7)	7 (7)	0 (0)
\$80,000 - \$99,999	10 (6)	10 (6)	0 (0)	9 (8)	9 (8)	0 (0)
\$100,000 +	24 (14)	24 (14)	0 (0)	17 (16)	17 (16)	0 (0)
Don't wish to answer	33 (19)	33 (19)	0 (0)	11 (10)	11 (10)	0 (0)
Don't know	17 (10)	17 (10)	0 (0)	4 (4)	4 (4)	0 (0)



Characteristic	DCC: Minimal Change Disease (n=172)	DSIC: Minimal Change Disease (n=172)	Diff. (n=0)	DCC: Membranous Nephropathy (n=106)	DSIC: Membranous Nephropathy (n=106)	Diff. (n=0)
Urine protein: Creatinine ratio at screening/enrollment						
Mean (SD)	8.3 (10.18)	8.3 (10.18)	0 (0)	6.7 (4.83)	6.7 (4.83)	0 (0)
Median (IQR)	5.7 (1.9, 10.3)	5.7 (1.9, 10.3)	0 (0, 0)	5.6 (3.6, 8.6)	5.6 (3.6, 8.6)	0 (0, 0)
Range	0.0 to 59.8	0.0 to 59.8	0, 0	0.0 to 29.0	0.0 to 29.0	0, 0
N	172	172	0	106	106	0
N missing	0	0	0	0	0	0
Birth weight (grams)						
Mean (SD)	3259.4 (585.21)	3259.4 (585.21)	0 (0)	3827.9 (1030.4)	3827.9 (1030.4)	0 (0)
Median (IQR)	3345.2 (3146.8, 3515.3)	3345.2 (3146.8, 3515.3)	0 (0, 0)	3667.7 (3045.0, 4167.4)	3667.7 (3045.0, 4167.4)	0 (0, 0)
Range	1644.3 to 4195.7	1644.3 to 4195.7	0, 0	2863.3 to 6000.0	2863.3 to 6000.0	0, 0
N	27	27	0	8	8	0
N missing	145	145	0	98	98	0
End stage renal disease or 40% loss of eGFR reached after baseline, n (%)						
Unknown	10 (6)	10 (6)	0 (0)	4 (4)	4 (4)	0 (0)
No	136 (79)	136 (79)	0 (0)	77 (73)	77 (73)	0 (0)
Yes	26 (15)	26 (15)	0 (0)	25 (24)	25 (24)	0 (0)

**Table B3:** Comparison of values computed in integrity check to DCC results (Non-Biopsied Childhood Onset Nephrotic Syndrome and Other)

Characteristic	DCC: Non-Biopsied Childhood Onset Nephrotic Syndrome (n=110)	DSIC: Non-Biopsied Childhood Onset Nephrotic Syndrome (n=110)	Diff. (n=0)	DCC: Other (n=162)	DSIC: Other (n=162)	Diff. (n=0)
Age (years) at baseline visit						
Mean (SD)	4.8 (3.60)	4.8 (3.60)	0 (0)	36.1 (20.66)	36.1 (20.66)	0 (0)
Median (IQR)	4.0 (2.0, 6.0)	4.0 (2.0, 6.0)	0 (0, 0)	37.0 (18.0, 52.0)	37.0 (18.0, 52.0)	0 (0, 0)
Range	0.0 to 16.0	0.0 to 16.0	0, 0	1.0 to 84.0	1.0 to 84.0	0, 0
N	110	110	0	162	162	0
N missing	0	0	0	0	0	0
Sex, n (%)						
Male	67 (61)	67 (61)	0 (0)	97 (60)	97 (60)	0 (0)
Female	43 (39)	43 (39)	0 (0)	65 (40)	65 (40)	0 (0)
Race, n (%)						
Multi-Racial	11 (10)	11 (10)	0 (0)	5 (3)	5 (3)	0 (0)
Asian/Asian American	9 (8)	9 (8)	0 (0)	28 (17)	28 (17)	0 (0)
Black/African American	19 (17)	19 (17)	0 (0)	35 (22)	35 (22)	0 (0)
Native Hawaiian/Other Pacific Islander	1 (1)	1 (1)	0 (0)	1 (1)	1 (1)	0 (0)
White/Caucasian	59 (54)	59 (54)	0 (0)	87 (54)	87 (54)	0 (0)
Unknown	11 (10)	11 (10)	0 (0)	6 (4)	6 (4)	0 (0)
Ethnicity, n (%)						
Unknown	5 (5)	5 (5)	0 (0)	1 (1)	1 (1)	0 (0)
Hispanic or Latino	23 (21)	23 (21)	0 (0)	36 (22)	36 (22)	0 (0)
Not Hispanic or Latino	82 (75)	82 (75)	0 (0)	125 (77)	125 (77)	0 (0)
Gross Annual Household Income at baseline, n (%)						
\$0 - \$19,999	8 (7)	8 (7)	0 (0)	36 (22)	36 (22)	0 (0)
\$20,000 - \$39,999	22 (20)	22 (20)	0 (0)	23 (14)	23 (14)	0 (0)
\$40,000 - \$59,999	7 (6)	7 (6)	0 (0)	18 (11)	18 (11)	0 (0)
\$60,000 - \$79,999	8 (7)	8 (7)	0 (0)	11 (7)	11 (7)	0 (0)
\$80,000 - \$99,999	8 (7)	8 (7)	0 (0)	13 (8)	13 (8)	0 (0)
\$100,000 +	20 (18)	20 (18)	0 (0)	22 (14)	22 (14)	0 (0)
Don't wish to answer	28 (25)	28 (25)	0 (0)	28 (17)	28 (17)	0 (0)
Don't know	9 (8)	9 (8)	0 (0)	11 (7)	11 (7)	0 (0)

Characteristic	DCC: Non-Biopsied Childhood Onset Nephrotic Syndrome (n=110)	DSIC: Non-Biopsied Childhood Onset Nephrotic Syndrome (n=110)	Diff. (n=0)	DCC: Other (n=162)	DSIC: Other (n=162)	Diff. (n=0)
Urine protein: creatinine ratio at screening/enrollment						
Mean (SD)	10.8 (10.99)	10.8 (10.99)	0 (0)	4.8 (5.65)	4.8 (5.65)	0 (0)
Median (IQR)	8.3 (2.8, 14.8)	8.3 (2.8, 14.8)	0 (0, 0)	3.3 (1.8, 5.6)	3.3 (1.8, 5.6)	0 (0, 0)
Range	0.0 to 56.4	0.0 to 56.4	0, 0	0.0 to 54.9	0.0 to 54.9	0, 0
N	103	103	0	162	162	0
N missing	7	7	0	0	0	0
Birth weight (grams)						
Mean (SD)	3321.3 (494.65)	3321.3 (494.65)	0 (0)	2842.1 (628.80)	2842.1 (628.80)	0 (0)
Median (IQR)	3345.2 (2990.7, 3628.7)	3345.2 (2990.7, 3628.7)	0 (0, 0)	2862.5 (2454.9, 3331.1)	2862.5 (2454.9, 3331.1)	0 (0, 0)
Range	2211.3 to 5046.2	2211.3 to 5046.2	0, 0	1740.0 to 3700.0	1740.0 to 3700.0	0, 0
N	88	88	0	8	8	0
N missing	22	22	0	154	154	0
End stage renal disease or 40% loss of eGFR reached after baseline, n (%)						
Unknown	4 (4)	4 (4)	0 (0)	11 (7)	11 (7)	0 (0)
No	85 (77)	85 (77)	0 (0)	120 (74)	120 (74)	0 (0)
Yes	21 (19)	21 (19)	0 (0)	31 (19)	31 (19)	0 (0)

**Table B4:** Comparison of values computed in integrity check to DCC results (Unknown)

Characteristic	DCC: Unknown (n=26)	DSIC: Unknown (n=26)	Diff. (n=0)
Age (years) at baseline visit			
Mean (SD)	28.8 (19.91)	28.8 (19.91)	0 (0)
Median (IQR)	20.5 (14.0, 41.0)	20.5 (14.0, 41.0)	0 (0, 0)
Range	2.0 to 69.0	2.0 to 69.0	0, 0
N	26	26	0
N missing	0	0	0
Sex, n (%)			
Male	13 (50)	13 (50)	0 (0)
Female	13 (50)	13 (50)	0 (0)
Race, n (%)			
Multi-Racial	2 (8)	2 (8)	0 (0)
Asian/Asian American	3 (12)	3 (12)	0 (0)
Black/African American	9 (35)	9 (35)	0 (0)
Native Hawaiian/Other Pacific Islander	0 (0)	0 (0)	0 (0)
White/Caucasian	10 (38)	10 (38)	0 (0)
Unknown	2 (8)	2 (8)	0 (0)
Ethnicity, n (%)			
Unknown	1 (4)	1 (4)	0 (0)
Hispanic or Latino	3 (12)	3 (12)	0 (0)
Not Hispanic or Latino	22 (85)	22 (85)	0 (0)
Gross Annual Household Income at baseline, n (%)			
\$0 - \$19,999	5 (19)	5 (19)	0 (0)
\$20,000 - \$39,999	0 (0)	0 (0)	0 (0)
\$40,000 - \$59,999	3 (12)	3 (12)	0 (0)
\$60,000 - \$79,999	3 (12)	3 (12)	0 (0)
\$80,000 - \$99,999	1 (4)	1 (4)	0 (0)
\$100,000 +	3 (12)	3 (12)	0 (0)
Don't wish to answer	9 (35)	9 (35)	0 (0)
Don't know	2 (8)	2 (8)	0 (0)

Characteristic	DCC: Unknown (n=26)	DSIC: Unknown (n=26)	Diff. (n=0)
Urine protein: creatinine ratio at screening/enrollment			
Mean (SD)	5.9 (4.43)	5.9 (4.43)	0 (0)
Median (IQR)	4.1 (2.4, 8.9)	4.1 (2.4, 8.9)	0 (0, 0)
Range	0.4 to 15.0	0.4 to 15.0	0, 0
N	24	24	0
N missing	2	2	0
Birth weight (grams)			
Mean (SD)	4054.0 (2007.6)	4054.0 (2007.6)	0 (0)
Median (IQR)	3628.7 (2976.7, 4422.5)	3628.7 (2976.7, 4422.5)	0 (0, 0)
Range	1814.4 to 8164.7	1814.4 to 8164.7	0, 0
N	7	7	0
N missing	19	19	0
End stage renal disease or 40% loss of eGFR reached after baseline, n (%)			
Unknown	8 (31)	8 (31)	0 (0)
No	18 (69)	18 (69)	0 (0)
Yes	0 (0)	0 (0)	0 (0)

## Attachment A: SAS Code

```
libname nep "X:\NIDDK\niddk-dr_studies6\NEPTUNE\private_created_data\Data for DSIC";
```

```
/******  
/* DSIC for NEPTUNE */  
/* DCC Created Table */  
/******
```

```
*temp datasets;  
data long; set nep.longitudinal_ds;  
run;
```

```
data pt; set nep.patient_ds;  
run;
```

```
*Age;  
proc means data=pt mean std median q1 q3 min max N missing;  
var PAT_AgeV2;  
run;
```

```
proc means data=pt mean std median q1 q3 min max N missing;  
var PAT_AgeV2;  
class PAT_Cohort;  
run;
```

```
*Sex;  
proc freq data=pt;  
tables PAT_Sex*PAT_Cohort/norow missing;  
run;
```

```
*Race;  
proc freq data=pt;  
tables PAT_Race*PAT_Cohort/norow missing;  
run;
```

```
*Ethnicity;  
proc freq data=pt;  
tables PAT_Hispanic*PAT_Cohort/norow missing;  
run;
```

```
*Gross annual household income;  
proc freq data=pt;  
tables AHHIncome*PAT_Cohort/norow missing;  
run;
```

```
*Urine Protein creatinine ration;
```

```
proc means data=pt mean std median q1 q3 min max n missing;  
var upcr_v1/  
run;
```

```
proc means data=pt mean std median q1 q3 min max n missing;  
var upcr_v1;  
class PAT_Cohort;  
run;
```

```
*birth weight;  
proc means data=pt mean std median q1 q3 min max n missing;  
var BirthWgt;  
run;
```

```
proc means data=pt mean std median q1 q3 min max n missing;  
var BirthWgt;  
class PAT_Cohort;  
run;
```

```
*end stage renal disease;  
proc freq data=pt;  
tables ESRDorEGFR40*PAT_Cohort/norow missing;  
run;
```