Insuring the Integrity of Clinical Registries: An Example of Managing a Large Multicenter Neurologically Ill Patient Database

Aardhra Venkatachalam *MPH*, **Anjali Perera** *BSN RN*, **Sonja Stutzman** *PhD*, **DaiWai Olson** *PhD RN CCRN*, **Venkatesh Aiyagari** *MD* , **Folefac Atem** *PhD*

INTRODUCTION

- All data is subject to error and missing values; Large databases are prone to greater errors and more missing data
- Errors have implications for patient treatments and clinical practice
- Database used : Establishing Normative Data for Pupillometer Assessments in Neuroscience Intensive Care (END-PANIC)
- Multicenter database houses over >30,000 pupil readings, which are used to provide clinical info in neurocritically ill patients
- This poster highlights the methods used to ongoing data quality assurance.

METHODS

I. Screening

- Enrollment of participants; Inclusion/Exclusion Criteria defined
- What type of Data was collected? (Ex: Baseline, Pupil, Daily)
- Where did we get the data? (Texas, California, Ohio)

II. Data Organization

-Enrollment of participants; deidentification of data
-segregated data into groups with identifier for all locations
-treated blank cells, extra columns, highlighted errors
-standardized variable columns and variable values

III. Diagnostic

-Did the data we collected make sense? If not, what was the issue?
-Made sure data was biologically possible with related points

IV. Treatment

-What did we do to fix the problems identified?
-made changes in spreadsheet and rerun descriptive statistics
**Checks were done as new data was added to the database

V. Missing Data

-What does the missing data tell us about the quality of our data?
-Was the missing data informative or non-informative?

UT Southwestern
O'Donnell Brain Institute

MAGNET RECOGNIZED
AMERICAN NURSES CREDENTIALING CENTER

An inclusive, **interdisciplinary** approach to managing large datasets leads to **higher** data integrity. **Missing data analysis** can further direct data collection and use of resources.

Take a picture to **download** the **full poster**

Aardhra.Venkatachalam@UTSouthwestern.edu

RESULTS

Table 1: T-test/Chi-squared to determine type of missingness

Variable	Gender	Age	Race	Ethnicity
NPIL	0.6542	0.0019	<.0001	<.0001
NPIR	0.0936	0.2870	0.2351	<.0001
CVL	0.6542	<.0001	<.0001	<.0001
CVR	0.0936	<.0001	0.2351	<.0001
DVL	0.6542	0.0019	<.0001	<.0001
DVR	0.0936	0.2870	0.2351	<.0001
CataL	0.6542	0.0019	<.0001	<.0001
CataR	0.0936	0.2870	0.2351	<.0001
LatL	0.6542	0.0019	<.0001	<.0001
LatR	0.0936	0.2870	0.2351	<.0001
MCVL	0.6542	0.0019	<.0001	<.0001
MCVR	0.0936	0.2870	0.2351	<.0001
SizeL	0.6542	0.0019	0.0076	<.0001
SizeR	0.0936	0.2870	0.0009	<.0001

Missing data was informative for age, race and ethnicity. Missing data was not informative fore gender for any of the pupillometer measurements used.

Table 2: Pearson's Correlation: Standardized coefficient

Variable(s)	Texas	Ohio	California	Combined
NPIL/NPIR	0.80	0.91	0.77	0.81
CVL/CVR	0.88	0.87	0.90	0.87
DVL/DVR	0.87	0.86	0.88	0.86
CataL/CataR	0.62	0.97	0.96	0.63
LatL/LatR	0.63	0.68	0.63	0.65
MCVL/MCVR	0.88	0.88	0.91	0.89
SizeL/SizeR	0.92	0.92	0.90	0.92

Npi, Dilation velocity, constriction velocity, pupil latency all showed a high correlation between the Texas, California and Ohio locations. Prese4nce/absence of cataract shoed as a low correlation, possibly because of the difference in age distribution between sites.

CONCLUSIONS

- Methods used to clean and manage database allowed for easy identification and correction of errors
- Process was double checked by a faculty level statistician
- Informative and non-informative missing data was identified, which helped the research team plan future steps and directions.
- We encourage others to utilize these techniques on large datasets with neurocritically ill patients

NEUROCRITICAL CARE SOCIETY