

# Inspecting and Curating MPOG Data

*before the statistical analysis*

Nicholas J. Douville, MD, PhD

Clinical Lecturer

University of Michigan

# Outline

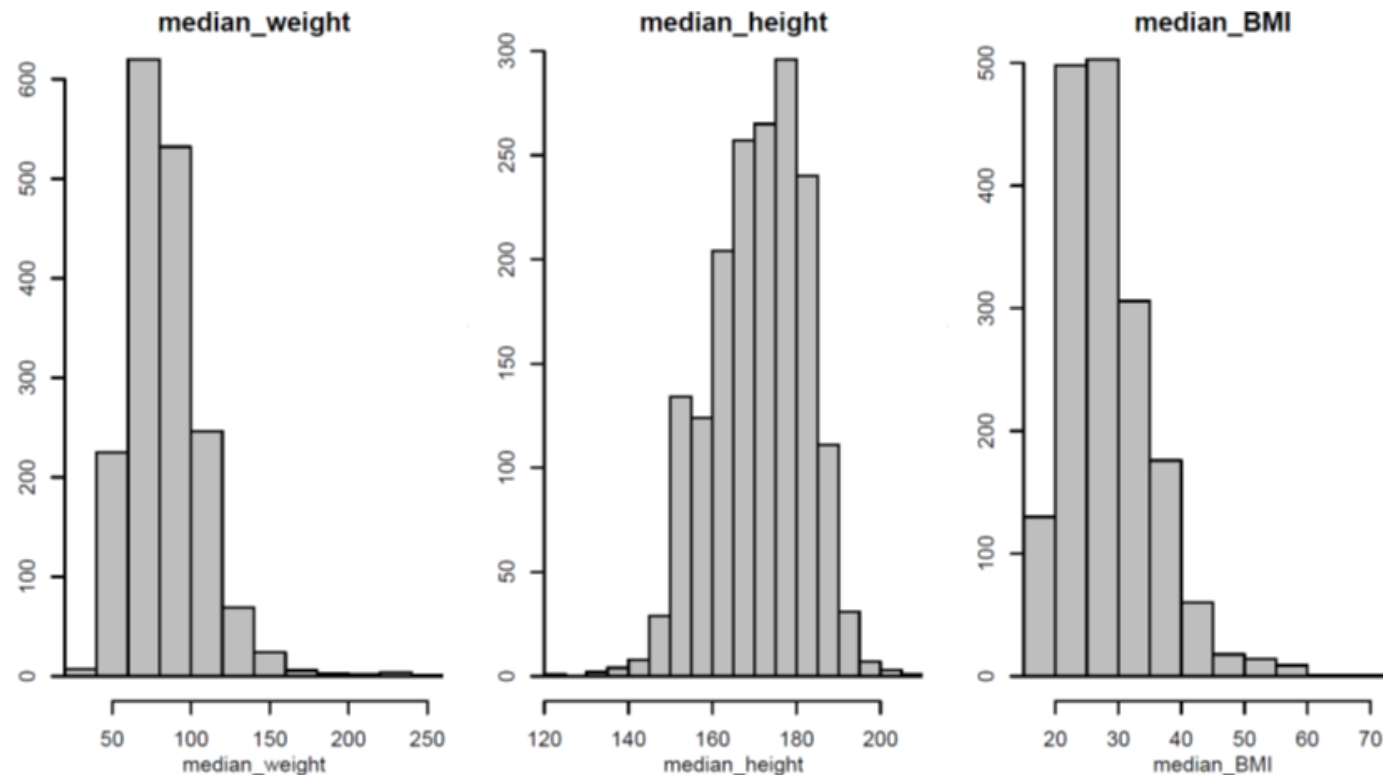
- Data Visualization
  - MPOG *Data Explorer*
- Data Cleaning
  - Case-by-Case *Audit Tool*
  - MPOG *Data Cleaning Tool*

# You finally have your data → now what?

- Generate Descriptive Statistics
  - This is a good way to spot “extreme” outliers that are likely to be a source of data error

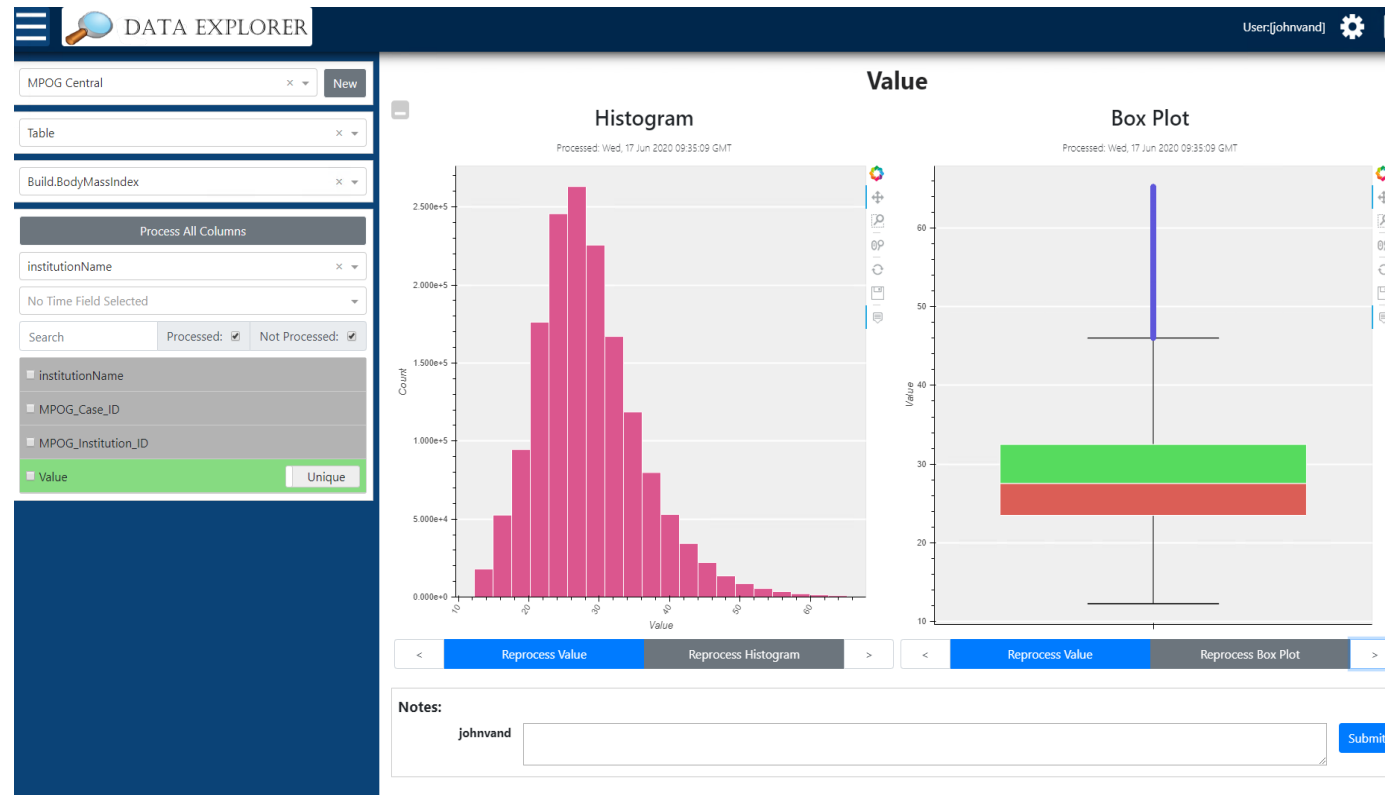
# You finally have your data → now what?

- Example: if the weight is > 225 kg, then potentially the user mistakenly entered lbs instead of kg (and the weight may be off by a factor of 2.2x)



# MPOG Data Explorer

- MPOG has created a tool to help streamline this process and identify systematic/institutional sources of error
- For the BMI example:



# MPOG Data Explorer

- We have the ability to then create histograms based on each institution in our Dataset

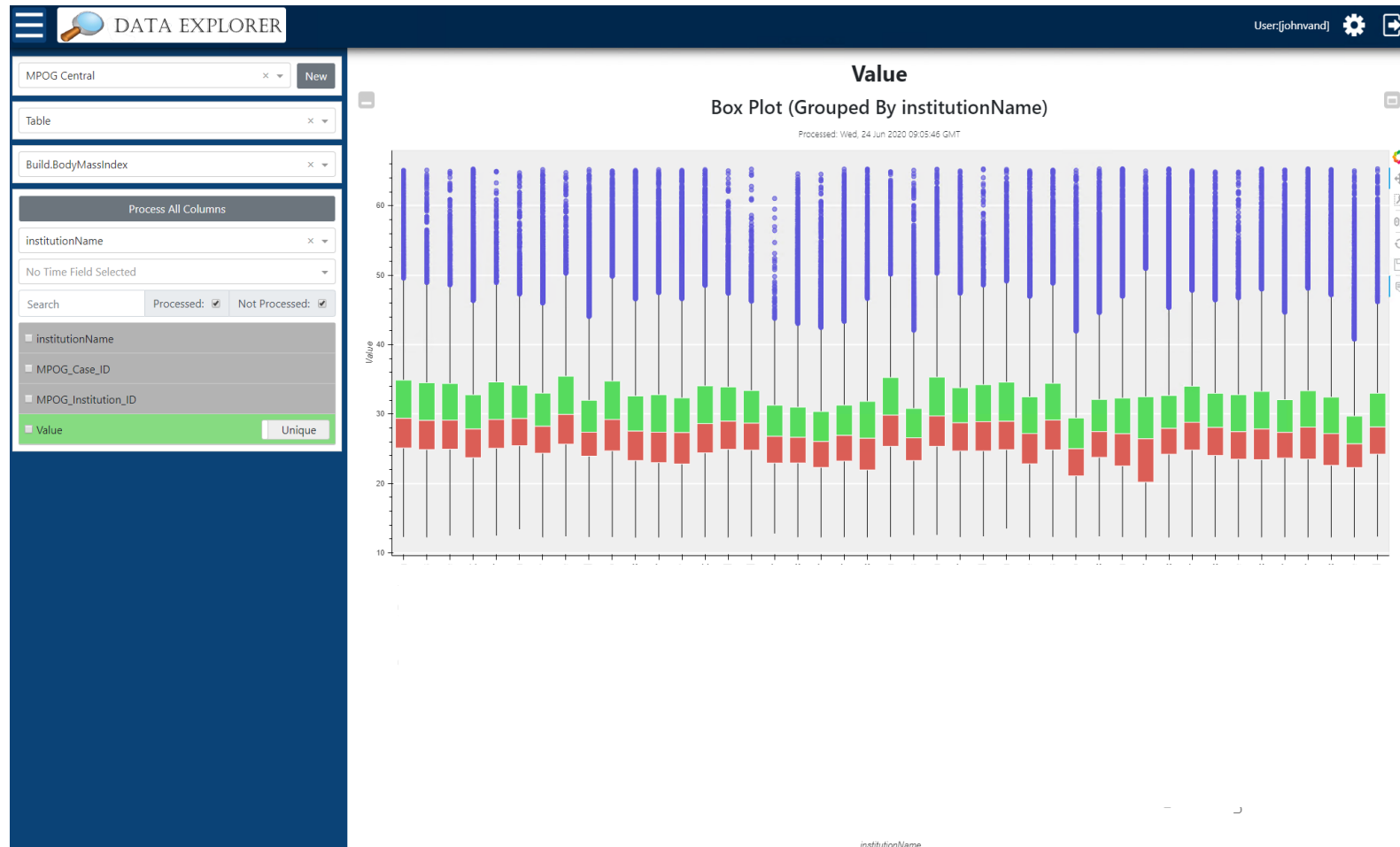


Your job is to assess discrepancies between the distributions:

- Appropriate (one center may do more ambulatory cases and another may be tertiary care center)
- Inappropriate (one center may document incorrectly frequently or have missing data)

# MPOG Data Explorer

- We can also graph box plots to visualize



# Data “Cleaning” – Case by Case Audit

- You need to consider all sources of error for your data.
- Examples:
  - Blood Transfusion: 250 units of packed red blood products at a single moment of time, they almost certainly meant to chart 250 mL of pRBC
  - Vasopressors/Inotropes: phenylephrine is charted at our institution in mcg/min.
    - If the provider accidentally enters mcg/kg/min ... then the factor could be off by a factor of 100.
    - This does not always reveal itself in histograms, but if you see a second “peak” within the distribution, it is worth investigating if there is a common “error” that a few different providers have made.



# MPOG Research Data Cleaning Tool

- For one project, I needed to know if a patient is on a beta-blocker
- The MPOG programmer, pulls the full medication “list” → clean to categorical variable

MPOG Research Categorization Utility

Prep Med -> 88

Filter Text:

Distinct Values Remaining: 0 Total Values Mapped (%): 100

Original Value	Count	Category
(None)	7245	No
None	4348	No
Prenatal Vitamin	916	No
Pre Natal Vitamin	857	No
Multi Vitamin	748	No
PNV	705	No
Prenatal Vitamin	387	No
Bupropion (Wellbutrin)	355	No
	288	Unknown
Omeprazole (Prilosec)	278	No
NORCO (ACETAMINOPHEN HYDROCODONE BITARTRATE)	252	No
Ibuprofen (SINUTRAC)	213	No
Bupropion (WYBUTIN)	205	No
Acetaminophen (Tylenol)	191	No
Synthroid (Levothyroxine)	183	No
Multivitamin Tablets	182	No
(None) Name (None)	170	No
ASA (Aspirin)	148	No
Name (None) (None)	143	No
Loxaprop (Proin, Zelnor)	136	No
Prilosec (Omeprazole)	124	No
Motrin (Ibuprofen)	121	No
Albuterol Aerosol	118	No
HYDROCODONE-ACETAMINOPHEN (NORCO)	103	No
peg LYSO electrolyte (LYSTEL)	103	No
WITAMIN D (ERGOCALCIFEROL)	87	No
(Unknown)	80	Unknown
Mefenamic (Anaprox)	73	No
PNV w/ Ca comb #12-4a-6a (PRENATAL PLUS CALCIUM CARB)	72	No
Sertraline (Zoloft)	69	No
Unknown	68	Unknown
oxyCODONE (ROBUCODONE)	68	No
AMITRIPTYLINE (AMITRIPT)	67	No
Zoloft (Sertraline)	66	No
simvastatin (ZOCOR)	63	No
Benzyd (ZESTRA)	62	No
Atorvastatin (Lipitor)	62	No
Iron Prenatal Vitamin	60	No
Indinavir (VIRAMIN II)	59	No
Amidopine (Norvasc)	58	No
Citalopram (Celexa)	57	No

Categorical Values

Unknown

No

Yes

# Approach to using the cleaning tool:

- Start with comprehensive list
- Systematic Search (and filter)

Beta Blockers	
Brand Name	Other Name
Bystolic	Nebivolol
	Timolol
Coreg	Carvedilol
Corgard	Nadolol
Inderal	Propranolol
Inderal LA	Propranolol
	Betaxolol
Levitol	Penbutolol
Lopressor	Metoprolol
Sectral	Acebutolol
Tenormin	Atenolol
Toprol XL	Metoprolol
Trandate	Labetalol
	Pindolol
Zebeta	Bisoprolol

5

MPOG Research Categorization Utility

Preop Med -> BB

Filter Text:  ? ☐ Hide Mapped Values

Distinct Values Remaining: 0 Total Values Mapped (%): 100

Original Value	Count	Category
carvedilol (COREG)	10	Yes
oxyCODONE-acetaminophen (PERCOCET) carvedilol (COREG)	6	Yes
carvedilol (COREG) lisinopril (ZESTRIL)	4	Yes
carvedilol (COREG) oxyCODONE-acetaminophen (PERCOCET)	4	Yes
Norco Carvedilol (Coreg)	3	Yes
losartan (COZAAR) carvedilol (COREG)	2	Yes
ASA (Aspirin) Bumetanide (Bumex) Carvedilol (Coreg) Oxycodone (Roxicodone)	2	Yes

# Next Steps

- Then Trade Names: Coreg
- Consider misspellings: Labetolol ....Inderol....etc
- Consider negations: “patient not on a beta blocker”
- Consider “Hide Mapped Values” to remove ones you have already viewed and sorted
- **Remember:** *pristine data quality isn't a priority for busy, multi-tasking clinicians*

# Next Steps

- MPOG has “phenotypes” for some key variables (for example: *Tobacco Smoking Classification*)

## Tobacco Smoking Classification

### Description

Classification to determine non, former, or current smoker. Vaping is considered smoking.

### Limitation

This collation relies upon the SmokingNotes\_Cleaned collation which utilizes mapping. Mapping values and constraints are updated periodically and thus subject to change

Value	Value Code	Definition
Missing	-999	Patient didn't have any information on smoking
Invalid	-998	Patient has smoking notes, but the entries are not discernable to s smoking status
Conflicting	-997	Patient is simultaneously labeled as a non-smoker and a smoker for the same case
Non-Smoker	0	Patient is a non smoker defined as never smoking / no history of smoking
Former Smoker	1	Patient is a former smoker defined as having a smoking history, but quit smoking and is not currently smoking
Current Smoker	2	Patient is a current smoker

- For any study – you need to decide if the phenotype work best or if you need a more subtle/nuanced description of that particular variable
- Researchers can decide what categorical groupings make the most sense for their study and work with MPOG programmers to include those in *the Data Cleaning Tool*.

# Summary

- Data Visualization
  - MPOG *Data Explorer*
- Data Cleaning
  - Case-by-Case Audit tool
  - MPOG *Data Cleaning Tool*