Data Preprocessing and Analysis

Xueyao Li

1

Missing-data Imputation

IPD Dataset

Discard data

- **Complete-case analysis**: exclude all units for which the outcome or any of the inputs are missing.
- **Available-case analysis**: study with subsets of the data; simply exclude a variable or set of variables because of their missing-data rates.

Lead to biased estimates and large standard errors of estimates.

Impute data

Mean imputation

Replace each missing value with the mean of the observed data for this variable.

Distort the distribution for this variable and relationships between variables; Lead to complications with summary measures including underestimates of the standard deviation.

Random imputation

Randomly sample missing values from the observed data for this variable.

Regression prediction

Fit a regression to the observed data and then use that to predict the missing values.

Sample Size and Sites

Study	COMMET	EXERT	Move to Exercise	TRAIN	ENGAGE	Total
N	31	32	21	28	46	158
No. of sites	2	6	1	6	8	23

Note. This tables shows the sample size and number of sites for each of the individual studies.

Amount of Missing Data Imputed

Observed Variable	# of Observations Missing	Percentage of Data
mMS Pre	8	6%
EuroQoL-5D	17	13%
TFC	8	6%
UHDRStV1	1	1%
SDMT Pre	18	14%
TMS	45	34%

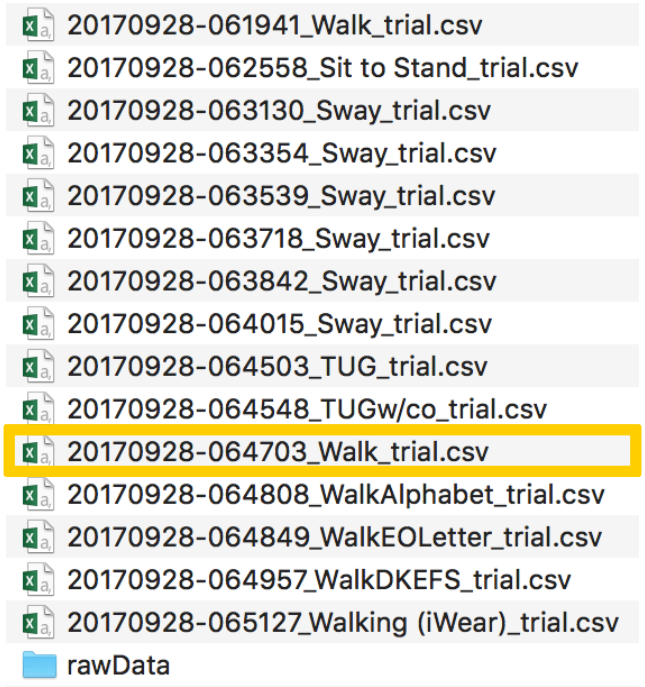
Note. TMS was not included in the analysis due to the amount of missingness of the data.

- Count missing data rates and exclude TMS
- Exclude participants with missing outcome data (modmotscpost)
- Random imputation
- Create 20 imputed datasets and run analysis on all the 20 datasets

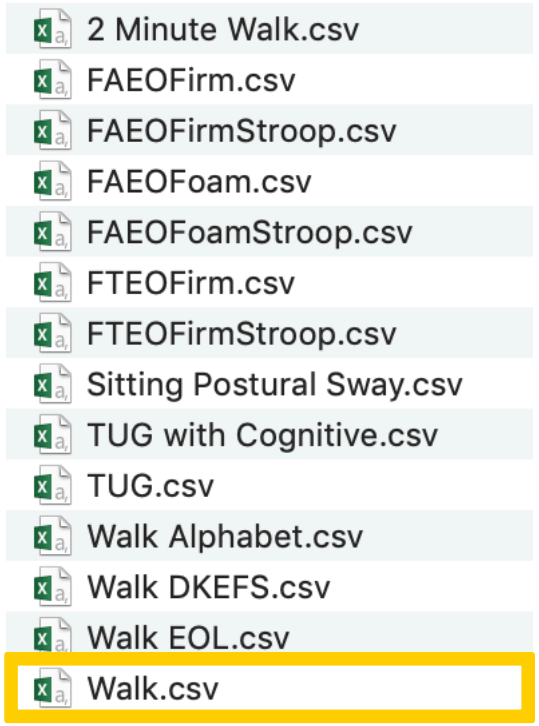
2

Data Integration

APDM Dataset



CSV.4 | Subject Group	iWear			
Subject Public ID	GH1			
Record Date	20170928-064703			
Condition	Open Ended			
File Name	20170928-124703_Walking__iWear_.h5			
Trial Notes				
Analysis Log	Warning::Algorithm for estimating sensor biases exp			
Analysis Version	3			
Measure	Normative M	Normative St	Mean	StDev
Anticipatory Postural Adjustment - APA Duration (s)	0.51	0.26		
Anticipatory Postural Adjustment - First Step Duration (s)	0.52	0.05		
Anticipatory Postural Adjustment - First Step Range of Motion (degrees)	33.17	9.49		
Anticipatory Postural Adjustment - Forward APA Peak (m/s^2)	0.52	0.28		
Anticipatory Postural Adjustment - Lateral APA Peak (m/s^2)	0.44	0.2		
Duration (s)			16.87	0
Gait - Lower Limb - Cadence L (steps/min)	116.71	9.14	109.88	1.04
Gait - Lower Limb - Cadence R (steps/min)	116.71	9.14	108.57	2.75
Gait - Lower Limb - Double Support L (%GCT)	18.31	3.56	20.8	2.29
Gait - Lower Limb - Double Support R (%GCT)	18.31	3.56	20.89	1.4
Gait - Lower Limb - Elevation at Midswing L (cm)	1.27	0.6	0.41	0.45
Gait - Lower Limb - Elevation at Midswing R (cm)	1.27	0.6	1.89	0.74
Gait - Lower Limb - Gait Cycle Duration L (s)	1.04	0.09	1.1	0.01
Gait - Lower Limb - Gait Cycle Duration R (s)	1.04	0.09	1.11	0.03
Gait - Lower Limb - Gait Speed L (m/s)	1.36	0.18	1.22	0.03
Gait - Lower Limb - Gait Speed R (m/s)	1.36	0.18	1.2	0.1
Gait - Lower Limb - Lateral Step Variability L (cm)	3.38	0.7	5.23	0
Gait - Lower Limb - Lateral Step Variability R (cm)	3.38	0.7	4.09	0
Gait - Lower Limb - Circumduction L (cm)	3.73	1.18	1.72	1.28
Gait - Lower Limb - Circumduction R (cm)	3.73	1.18	4.02	4.41
Gait - Lower Limb - N (#)			4	0



	Duration (s)	Gait - Lower Limb - Cadence L (steps/min)	Gait - Lower Limb - Cadence R (steps/min)
GH1	16.87	109.88	108.57
GH10	8.11	121.54	120.87
GH11	6.71	108.39	107.77
GH12	8.34	113.12	112.83
GH13	8.81	104.12	103.81
GH14	7.91	117.19	117.56
GH15	16.04	109.98	110.59
GH16	7.98	120.19	118.7
GH17	10.35	78.37	74.42
GH18	7.24	113.88	112.83
GH19	6.69	109.91	109.88
GH2	8.82	108.62	109.52
GH20	8.76	103.42	102.69
GH3	7.5	124.38	124.28
GH4	7.64	116.28	114.19
GH5	6.27	132.79	131.24
GH6	6.63	112.85	109.71
GH7	10.06	95.77	93.54
GH8	7.77	115.18	114.83

3

Correlation and Regression

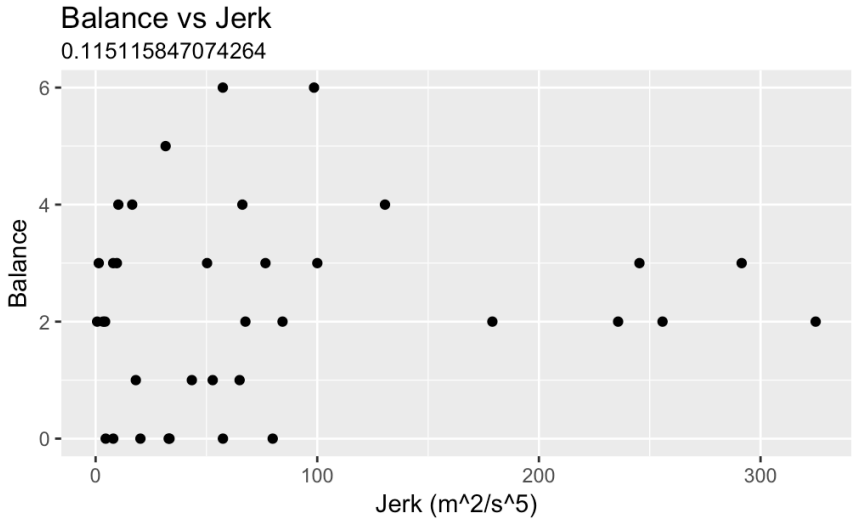
Dual Task Analysis on iWear and APDM Datasets

iWear Subscores

- **Balance Subscore**: Sum of Tandem Walk and Retropulsion Pull Tests from UHDRS Motor
- **Chorea Subscore**: Sum of Maximal Chorea Measures (i.e, Face, BOL, Trunk, RUE, LUE, RLE, LLE)
- **Dystonia Subscore**: Sum of Maximal Dystonia and Rigidity
- **Chorea Eye Subscore**: Sum of Ocular Pursuit, Saccade Initiation and Saccade Velocity
- **Chorea Total Subscore**: Sum of Chorea Subscore and Chorea Eye Subscore

Correlation

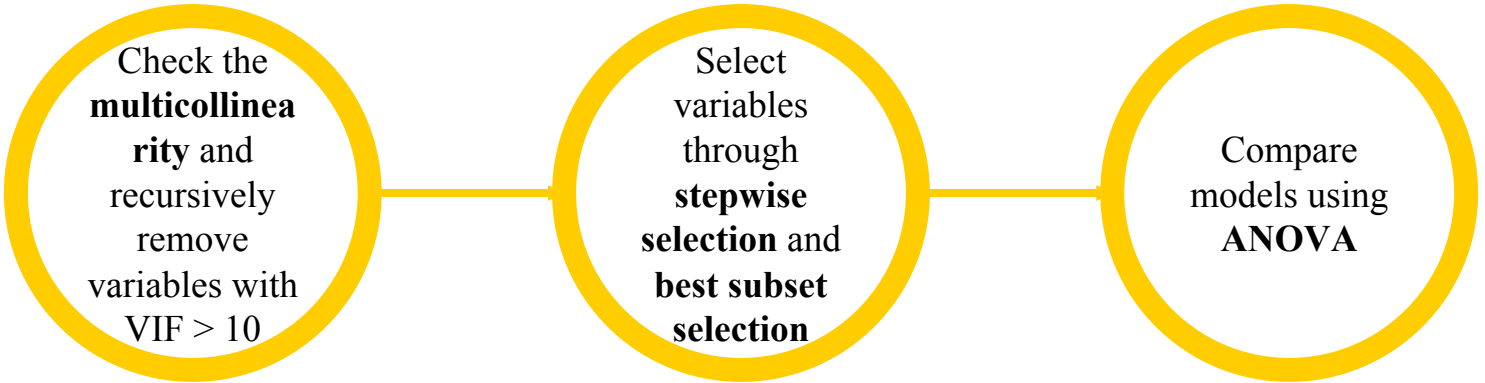
Plot these subscores against the standing (Feet Apart, Eyes Open, Firm Surface) and sitting (Sitting Postural Sway) task outcome measures of **Jerk, Mean Velocity, RMS Sway and Sway Area**.



Regression

- Subscore ~ FeetApart + FeetTogether + Sitting
- Subscore ~ FeetApartStroop + FeetTogetherStroop
- Subscore ~ FeetApart + FeetTogether + FeetApartStroop + FeetTogetherStroop
- Subscore ~ FeetApart + FeetTogether + FeetApartStroop + FeetTogetherStroop + Sitting

Regression


Check the **multicollinearity** and recursively remove variables with VIF > 10
Select variables through **stepwise selection** and **best subset selection**
Compare models using **ANOVA**

Thanks!

Any questions ?