

Run an initial reliability test of all the sicknesses

Unidimensional Reliability

Frequentist Scale Reliability Statistics

Coefficient	Estimate	Std. Error	95% CI	
			Lower	Upper
Coefficient α	0.892	0.010	0.871	0.912

Frequentist Individual Item Reliability Statistics

Item	Coefficient α (if item dropped)			Item-rest correlation		
	Estimate	Lower 95% CI	Upper 95% CI	Estimate	Lower 95% CI	Upper 95% CI
Allergies	0.882	0.859	0.904	0.625		
Allergy	0.872	0.848	0.896	0.795		
AnimalBite	0.872	0.847	0.897	0.812		
Asthma	0.879	0.857	0.901	0.668		
Bruises	0.876	0.854	0.899	0.712		
Diarrhea	0.891	0.870	0.912	0.444		
Dysmenorrhea	0.900	0.881	0.919	0.349		
Fainting	0.881	0.859	0.902	0.640		
Fever	0.885	0.865	0.906	0.553		
Fracture	0.878	0.855	0.902	0.677		
SoreThroat	0.888	0.868	0.908	0.496		
Trauma	0.887	0.865	0.908	0.535		

Note. The analytic confidence interval is not available for the item-rest correlation.

A coeff α 0.892 means very good internal consistency within the sickness variables

The individual reliability suggests that if the feature has been removed what will be the value of α if it increases then the feature is not align well with the other sickness (features)

Only Dysmenorrhea will cause α to drop or is the feature that is the least consistent (I've tested both with Dysmenorrhea and without and the PCA results (below) are better with so we will not drop this)

Item-rest correlation tells us if a variable behaves like the rest of scale in this case, only Dysmenorrhea (0.35) and Diarrhea (0.444) show poor alignment.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Test	
	MSA
Overall MSA	0.752
SoreThroat	0.694
Fever	0.621
Diarrhea	0.607
Dysmenorrhea	0.821
Allergies	0.752
Asthma	0.738
Allergy	0.870
AnimalBite	0.702
Fracture	0.815
Bruises	0.797
Fainting	0.822
Trauma	0.883

Bartlett's Test		
X ²	df	p
2577.899	66.000	< .001

KMO test the adequacy of our dataset for factor analysis a score of .75 means our dataset is above average or good enough for these types of analysis

With a Bartlett's Test p value is less than 0.05 means our variables are significantly correlated our data is good to use for PCA /FA.

Correlation Matrix

Pearson's Correlations													
Variable		SoreThroat	Fever	Diarrhea	Dysmenorrhea	Allergies	Asthma	Allergy	AnimalBite	Fracture	Bruises	Fainting	Trauma
1. SoreThroat	Pearson's r	—											
	p-value	—											
2. Fever	Pearson's r	0.840***	—										
	p-value	< .001	—										
3. Diarrhea	Pearson's r	0.607***	0.766***	—									
	p-value	< .001	< .001	—									
4. Dysmenorrhea	Pearson's r	0.184**	0.105	0.086	—								
	p-value	0.003	0.090	0.164	—								
5. Allergies	Pearson's r	0.156*	0.350***	0.288***	0.354***	—							
	p-value	0.011	< .001	< .001	< .001	—							
6. Asthma	Pearson's r	0.383***	0.384***	0.140*	0.435***	0.729***	—						
	p-value	< .001	< .001	0.023	< .001	< .001	—						
7. Allergy	Pearson's r	0.451***	0.510***	0.473***	0.192**	0.569***	0.512***	—					
	p-value	< .001	< .001	< .001	0.002	< .001	< .001	—					
8. AnimalBite	Pearson's r	0.214***	0.286***	0.319***	0.379***	0.766***	0.681***	0.813***	—				
	p-value	< .001	< .001	< .001	< .001	< .001	< .001	< .001	—				
9. Fracture	Pearson's r	0.154*	0.168**	0.065	0.301***	0.454***	0.500***	0.654***	0.759***	—			
	p-value	0.012	0.006	0.291	< .001	< .001	< .001	< .001	< .001	< .001	—		
10. Bruises	Pearson's r	0.449***	0.410***	0.304***	0.295***	0.352***	0.491***	0.595***	0.537***	0.622***	—		
	p-value	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001	—	
11. Fainting	Pearson's r	0.355***	0.463***	0.374***	0.137*	0.369***	0.347***	0.623***	0.555***	0.553***	0.676***	—	
	p-value	< .001	< .001	< .001	0.026	< .001	< .001	< .001	< .001	< .001	< .001	< .001	—
12. Trauma	Pearson's r	0.216***	0.203***	0.144*	0.192**	0.299***	0.307***	0.533***	0.562***	0.627***	0.503***	0.388***	—
	p-value	< .001	< .001	0.019	0.002	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001
* p < .05, ** p < .01, *** p < .001													

* p < .05, ** p < .01, *** p < .001

The following features Fever, SoreThroat, Allergy, AnimalBite, Fracture, Asthma, and Bruises are strongly intercorrelated

Dysmenorrhea shows weaker correlation – contributes less may not be essential

Factor Loadings

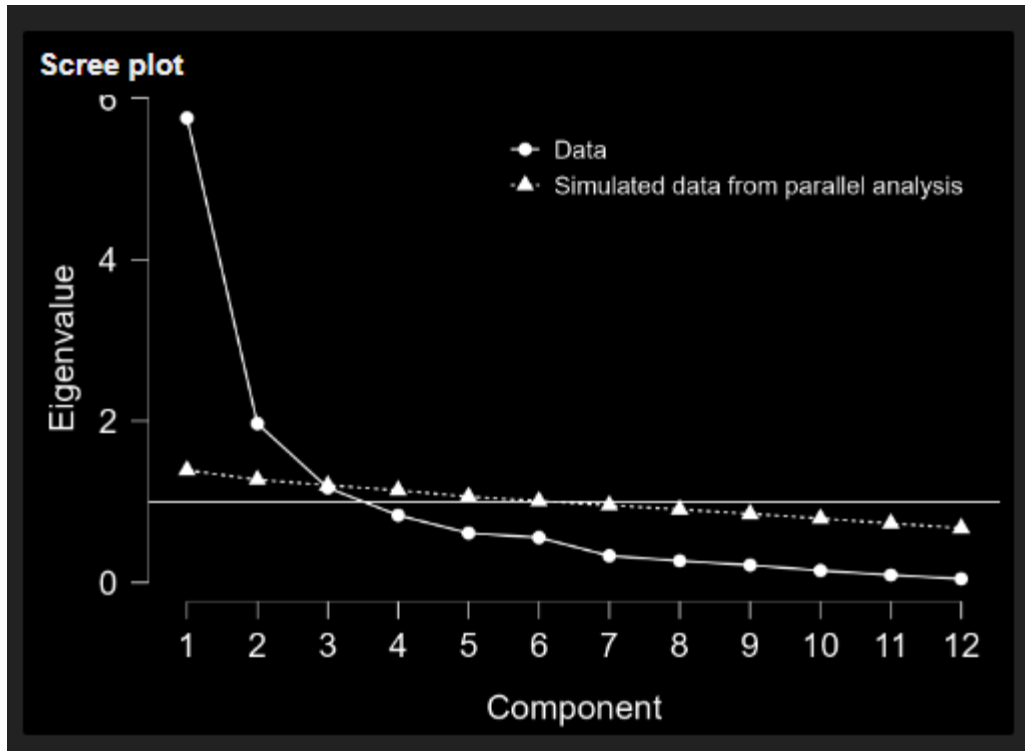
<i>Component Loadings</i>				
	RC1	RC2	RC3	Uniqueness
Fracture	0.914			0.176
Trauma	0.881			0.387
Fainting	0.750			0.337
Bruises	0.724			0.339
Allergy	0.669			0.227
AnimalBite	0.617		0.491	0.135
Fever		0.957		0.086
SoreThroat		0.899		0.212
Diarrhea		0.890		0.255
Asthma			0.799	0.220
Dysmenorrhea			0.790	0.480
Allergies			0.785	0.252

Note. Applied rotation method is promax.

We have 3 principal components with loadings greater than 0.4 means they are meaningful. The RC1 focus is on feelings based on physical injuries. RC2 focus is on germ/infectious related sickness while RC3 focuses on allergy induced or body reaction type of sickness.

The uniqueness tells us how the portion of the variance that is not explained by 3 components the lower the better. We can see here again that Dysmenorrhea doesn't align well with the features in this dataset.

Scree Plot



Looking at the scree plot our elbow value is 3, hence the 3 components.

The solid lines indicate our eigenvalues (stretch/shrunk)

Dashed line is for simulation of eigenvalues against our threshold of greater than 1 which also suggest that 3 is our best number of components

Total Variance Explained – how much info is preserved

<i>Component Characteristics</i>						
	Unrotated solution			Rotated solution		
	Eigenvalue	Proportion var.	Cumulative	SumSq. Loadings	Proportion var.	Cumulative
Component 1	5.752	0.479	0.479	3.723	0.310	0.310
Component 2	1.969	0.164	0.643	2.771	0.231	0.541
Component 3	1.173	0.098	0.741	2.399	0.200	0.741

With total variance of 0.741 the 3 PC solution captures a strong structure of the original dataset.

We can confidently reduce our data into 3 components without fear of losing much needed info

Principal Components Formed with profiling

Component Loadings

	RC1	RC2	RC3	Uniqueness
Fracture	0.914			0.176
Trauma	0.881			0.387
Fainting	0.750			0.337
Bruises	0.724			0.339
Allergy	0.669			0.227
AnimalBite	0.617		0.491	0.135
Fever		0.957		0.086
SoreThroat		0.899		0.212
Diarrhea		0.890		0.255
Asthma			0.799	0.220
Dysmenorrhea			0.790	0.480
Allergies			0.785	0.252

Note. Applied rotation method is promax.

PC1 – Physical/external Injury related reflects experience seen on physically active individuals or places that are not safe.

The symptoms included are physical (Fracture, Trauma, Bruises)

Aftermath of physical injuries (Fainting)

Physical irritations (Animal Bite/Allergies)

PC2 – Adenovirus types of symptoms, they are viruses that can cause both respiratory and gastrointestinal symptoms such as Fever, SoreThroat and Diarrhea.

PC3 – Symptoms influenced by Hormonal/inflammation, asthma, dysmenorrhea, allergies are body reaction from either hormonal imbalance or inflammation.

PCA Results without Dysmenorrhea (*scored lower than included*)

Without Dysmenorrhea Principal Component Analysis

Kaiser-Meyer-Olkin Test

	MSA
Overall MSA	0.743
SoreThroat	0.699
Fever	0.617
Diarrhea	0.603
Allergies	0.739
Asthma	0.712
Allergy	0.879
AnimalBite	0.687
Fracture	0.808
Bruises	0.789
Fainting	0.817
Trauma	0.881

Bartlett's Test

X ²	df	p
2491.857	55.000	< .001

Mardia's Test of Multivariate Normality

	Value	Statistic	df	p
Skewness	70.724	3111.851	286	< .001
Small Sample Skewness	70.724	3153.163	286	< .001
Kurtosis	149.961	3.344		< .001

Note. The statistic for skewness is assumed to be Chi² distributed and the statistic for kurtosis standard normal.

Chi-squared Test

	Value	df	p
Model	508.820	25	< .001

Component Loadings

	RC1	RC2	RC3	Uniqueness
Trauma	0.911			0.363
Fracture	0.894			0.172
Bruises	0.803			0.290
Fainting	0.732			0.351
Allergy	0.553			0.230
AnimalBite	0.481		0.622	0.105
Fever		0.954		0.086
SoreThroat		0.903		0.202
Diarrhea		0.891		0.255
Allergies			0.981	0.114
Asthma			0.852	0.226

Note. Applied rotation method is promax.

Component Characteristics

	Unrotated solution			Rotated solution		
	Eigenvalue	Proportion var.	Cumulative	SumSq. Loadings	Proportion var.	Cumulative
Component 1	5.610	0.510	0.510	3.462	0.315	0.315
Component 2	1.934	0.176	0.686	2.769	0.252	0.567
Component 3	1.062	0.097	0.782	2.375	0.216	0.782

Component Correlations

	Component 1	Component 2	Component 3
Component 1	1.000	0.417	0.564
Component 2	0.417	1.000	0.327
Component 3	0.564	0.327	1.000

Scree plot

