A COMPARATIVE STUDY: THE RASCH PARAMETER ESTIMATES AS A RESULT OF IMPUTING MISSING DATA IN WINSTEPS AND R

Yasmin Hatt & Eleni Pissaridou

21.03.2019

MISSING DATA: WHY?

Item Bank

Time Constraint

Losing part of a test

Skipping a question (to come back to later?)

Survey: choosing to not answer some parts

…

MISSING DATA: TYPES

Missing Completely At Random (MCAR)

Missing At Random (MAR)

Missing Not At Random (MNAR)

MISSING DATA: EFFECTS

Produce bias estimates

Reduce statistical power of the test

May not accurately represent some members of the cohort

…

MISSING DATA: WHAT CAN WE DO?

Minimize missing data

- Test construct
- Delivery method
- Cohort composition
- Sample size
- …

But… often unavoidable!!

- How do we handle it?
- Imputing?
- How accurate?

INVESTIGATION: THE DATASET

Complete dataset

- 1000 candidates
- 53 items

2 types of items:

- Dichotomous
- Polytomous (4 score categories)

Initial calibration in Winsteps:

- Ability range: -2.08, 4.89
- Item difficulty range: -1.62 to 3.07

INVESTIGATION: THE METHOD

Complete dataset

Randomly deleted varying percentages:

- 1
- 3
- 5
- 10
- 20
- 30
- 40
- 50
- 60

- Repeated 50 times for each percentage

Imputed missing data values

1. Winsteps XFILE
2. R Amelia Package

Calculated:

- % imputed data that matched the original
- and % of agreement between Rasch estimates

WINSTEPS XFILE

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Person	Item	OBSERVED	ORDERED	EXPECTATION	VARIANCE	ZSCORE	RESIDUAL	PERSON MEASURE	ITEM MEASURE	MEASURE DIFFERENCE	LOG (PROBABILITY)	PREDICTED PERSON MEASURE	PREDICTED ITEM MEASURE	KURTOSIS	MORE PROBABLE RESPONSE	RESPONSE WEIGHT	RESPONSE STATUS	CODE	PERSON LABEL	ITEM LABEL
2	1	1	1	1	1.37523	0.80595	-0.41797	-0.37523	-0.62844	-0.12911	-0.49933	-1.56114	-1.07557	0.31801	1.28025	2	1	1	1	0001	I0001
3	1	2	2	2	1.97277	0.11774	0.07936	0.02723	-0.62844	-0.39443	-0.23401	-0.12611	-0.39443	-0.62844	0.11604	2	1	1	2	0001	I0002
4	1	3	-1	-1	2.02299	0.02247	0	0	-0.62844	3.12081	-3.74925	0	0	0	0.02095	2	1	0	.	0001	I0003
5	1	4	2	2	2.06535	0.22078	-0.13908	-0.06535	-0.62844	-0.92741	0.29897	-0.25496	-0.92741	-0.62844	0.21368	2	1	1	2	0001	I0004
6	1	5	-1	-1	2.21164	0.2512	0	0	-0.62844	-1.52582	0.89738	0	0	0	0.19036	2	1	0	.	0001	I0005
7	1	6	2	2	1.96741	0.15388	0.08309	0.03259	-0.62844	-0.41489	-0.21355	-0.16835	-0.41489	-0.62844	0.15168	2	1	1	2	0001	I0006
8	1	7	2	2	1.97144	0.05828	0.11833	0.02856	-0.62844	-0.10106	-0.52738	-0.06091	-0.10108	-0.62842	0.05612	2	1	1	2	0001	I0007
9	1	8	2	2	1.90805	0.15049	0.23704	0.09195	-0.62844	0.03182	-0.66026	-0.1731	0.03182	-0.62844	0.13297	2	1	1	2	0001	I0008
10	1	9	-1	-1	1.7489	0.5194	0	0	-0.62844	-0.4824	-0.14604	0	0	0	1.14023	2	1	0	.	0001	I0009
11	1	10	-1	-1	1.56197	0.51039	0	0	-0.62844	-0.50156	-0.12688	0	0	0	0.74951	2	1	0	.	0001	I0010
12	1	11	2	2	1.98389	0.22434	0.03401	0.01611	-0.62844	-0.78478	0.15634	-0.11009	-0.55386	-0.85937	0.6898	2	1	1	2	0001	I0011
13	1	12	-1	-1	2.06319	0.21625	0	0	-0.62844	-0.92363	0.29519	0	0	0	0.2095	2	1	0	.	0001	I0012
14	1	13	2	2	2.07269	0.16431	-0.17934	-0.07269	-0.62844	-1.08667	0.45823	-0.18584	-1.08668	-0.62844	0.15375	2	1	1	2	0001	I0013
15	1	14	3	3	2.25623	0.45218	1.10606	0.74377	-0.62844	-1.4684	0.83996	-1.02663	0.97544	-3.07228	0.96944	2	1	1	3	0001	I0014
16	1	15	-1	-1	1.77329	0.73971	0	0	-0.62844	-0.65167	0.02323	0	0	0	1.71479	2	1	0	.	0001	I0015
17	1	16	2	2	1.89743	0.65921	0.12633	0.10257	-0.62844	-0.79736	0.16892	-0.43061	-0.46203	-0.96376	1.68649	2	1	1	2	0001	I0016
18	1	17	2	2	1.70787	0.36039	0.48663	0.29213	-0.62844	-0.83726	0.20882	-0.45489	0.28363	-1.74933	0.4442	2	1	1	2	0001	I0017
19	1	18	-1	-1	2.05357	0.24132	0	0	-0.62844	-0.85144	0.223	0	0	0	0.23689	2	1	0	.	0001	I0018
20	1	19	-1	-1	1.9941	0.21154	0	0	-0.62844	-0.60057	-0.02787	0	0	0	0.21148	2	1	0	.	0001	I0019
21	1	20	2	2	1.55809	0.54782	0.59706	0.44191	-0.62844	-0.19612	-0.43232	-0.49152	0.50764	-1.33221	0.88313	2	1	1	2	0001	I0020
22	1	21	2	2	1.2531	0.38674	1.20102	0.7469	-0.62844	0.67156	-1.3	-1.06111	2.20451	-2.16139	0.36714	1	1	1	2	0001	I0021
23	1	22	1	1	1.23552	0.2916	-0.43616	-0.23552	-0.62844	0.37276	-1.0012	-0.41425	-1.46943	1.21375	0.2482	1	1	1	1	0001	I0022
24	1	23	-1	-1	1.31446	0.33722	0	0	-0.62844	0.13026	-0.7587	0	0	0	0.29519	1	1	0	.	0001	I0023
25	1	24	1	1	1.16092	0.22158	-0.34185	-0.16092	-0.62844	1.01804	-1.64648	-0.28309	-1.40613	1.79573	0.18284	1	1	1	1	0001	I0024
26	1	25	0	0	0.43425	0.24568	-0.8761	-0.43425	-0.62844	-0.36389	-0.26455	-0.56959	-1.4625	0.47017	0.06461	0	1	1	0	0001	I0025
27	1	26	-1	-1	0.63273	0.23238	0	0	-0.62844	-1.17239	0.54395	0	0	0	0.07038	1	1	0	.	0001	I0026
28	1	27	-1	-1	0.25764	0.19126	0	0	-0.62844	0.42981	-1.05825	0	0	0	0.08152	0	1	0	.	0001	I0027
29	1	28	1	1	0.2773	0.20041	1.61436	0.7227	-0.62844	0.32944	-0.95788	-1.28264	1.42805	-1.72705	0.07992	0	1	1	1	0001	I0028
30	1	29	1	1	1.21581	0.4278	-0.32995	-0.21581	-0.62844	0.69451	-1.32295	-0.62957	-1.12249	1.18855	0.43048	1	1	1	1	0001	I0029
31	1	30	-1	-1	1.00297	0.44897	0	0	-0.62844	0.61278	-1.24122	0	0	0	0.51168	1	1	0	.	0001	I0030
32	1	31	2	2	1.79954	0.21379	0.43355	0.20046	-0.62844	0.44115	-1.06959	-0.29299	0.44115	-0.62844	0.14962	2	1	1	2	0001	I0031
33	1	32	2	2	1.71557	0.48491	0.40846	0.28443	-0.62844	-0.04725	-0.58119	-0.23139	0.36964	-1.04534	1.11673	2	1	1	2	0001	I0032

R PACKAGE AMELIA

Honaker J., King G., and Blackwell M. (2011)

Imputation using bootstrap-based Expectation Maximization (EM)

Method overview:

1. Bootstrap on incomplete data (this is the full data set)
2. Estimation of sufficient statistics by EM
3. Fill in missing values using these estimated sufficient statistics

R AMELIA PACKAGE

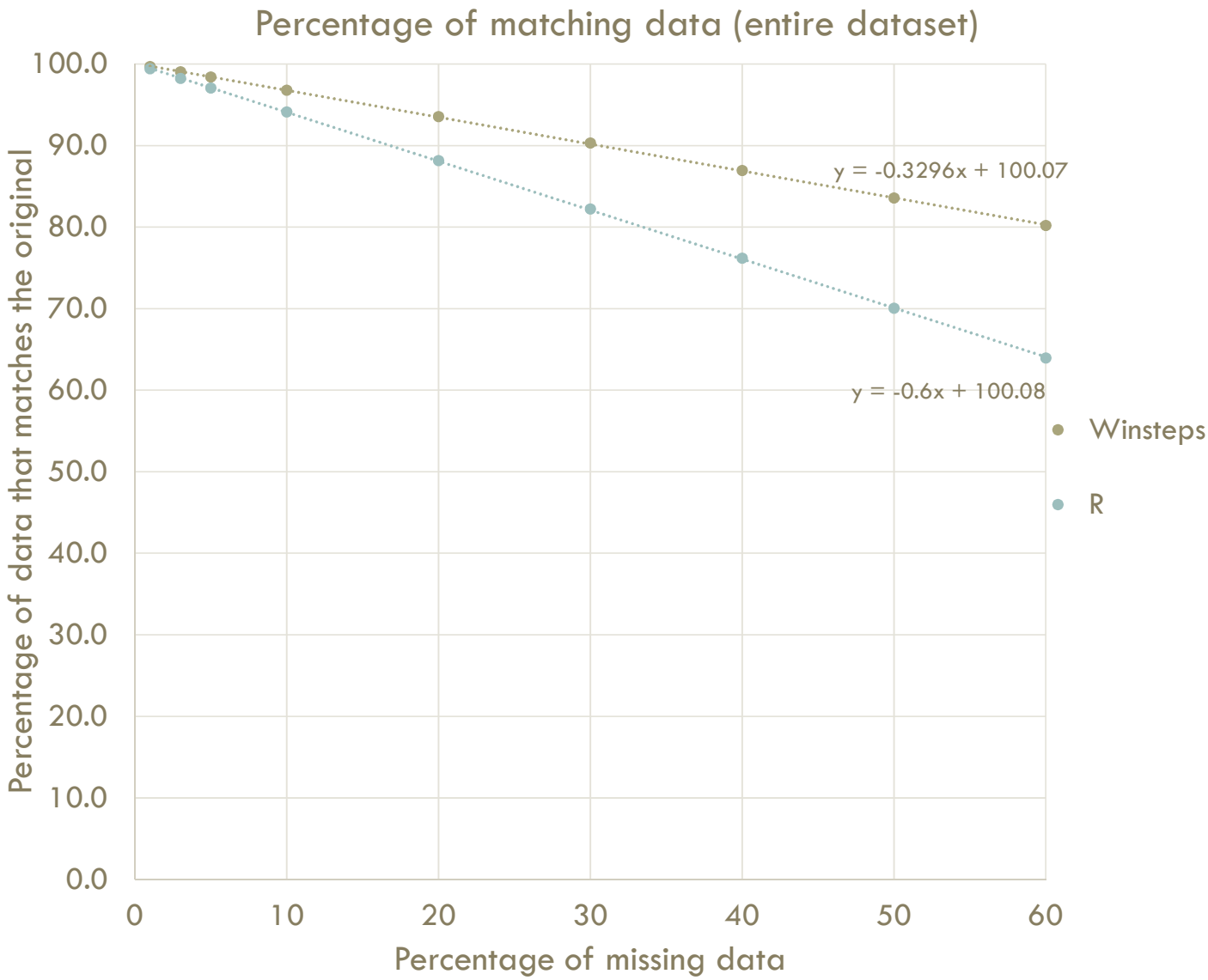
```
amelia(x=dat, m=1, ords=1:53, ...)
```

Parameters used:

- `x:` our incomplete data set
- `m:` number of imputed sets to create
- `ords:` indicator of ordinal variables
- `boot.type:` non-parametric bootstrap

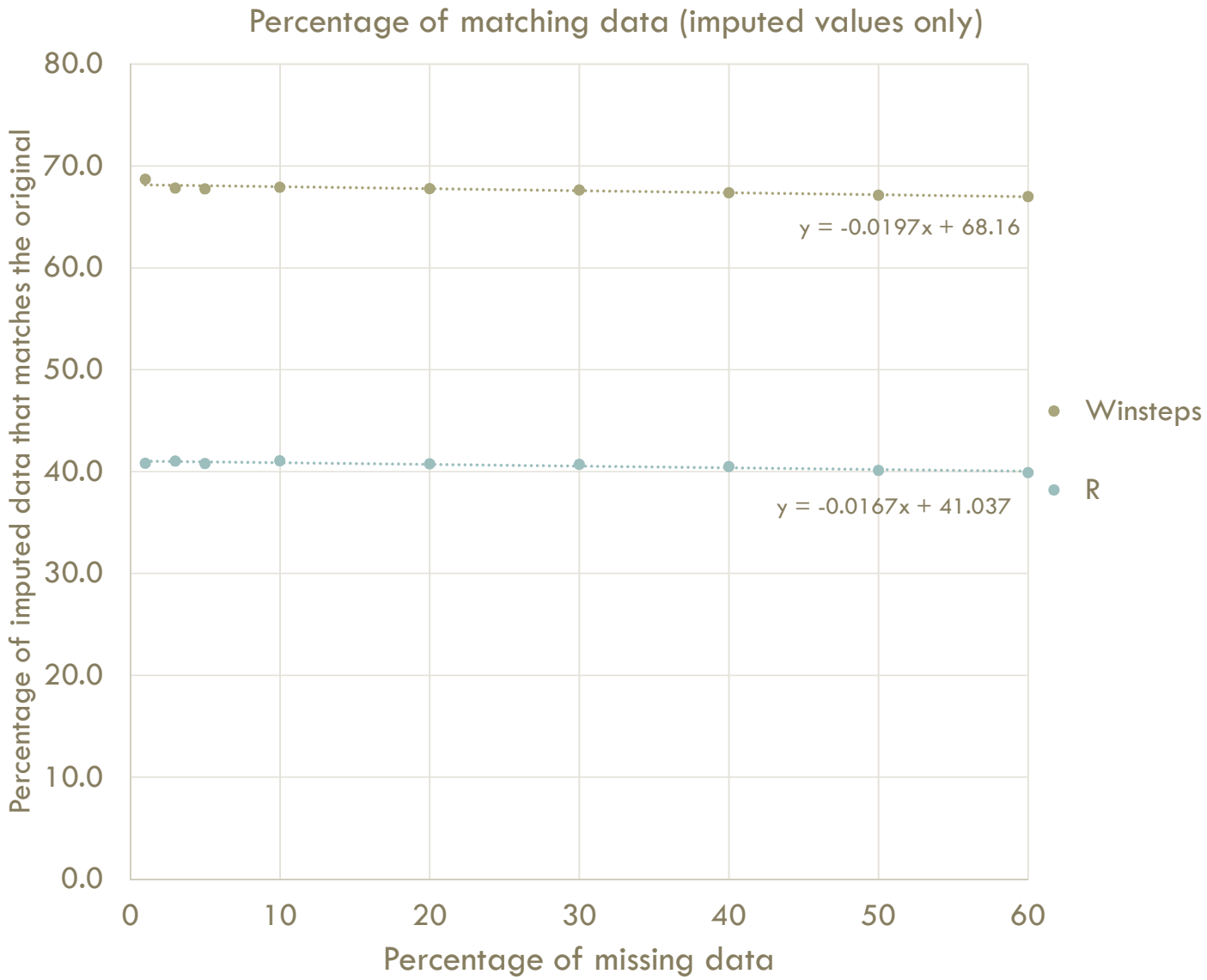
THE RESULTS:

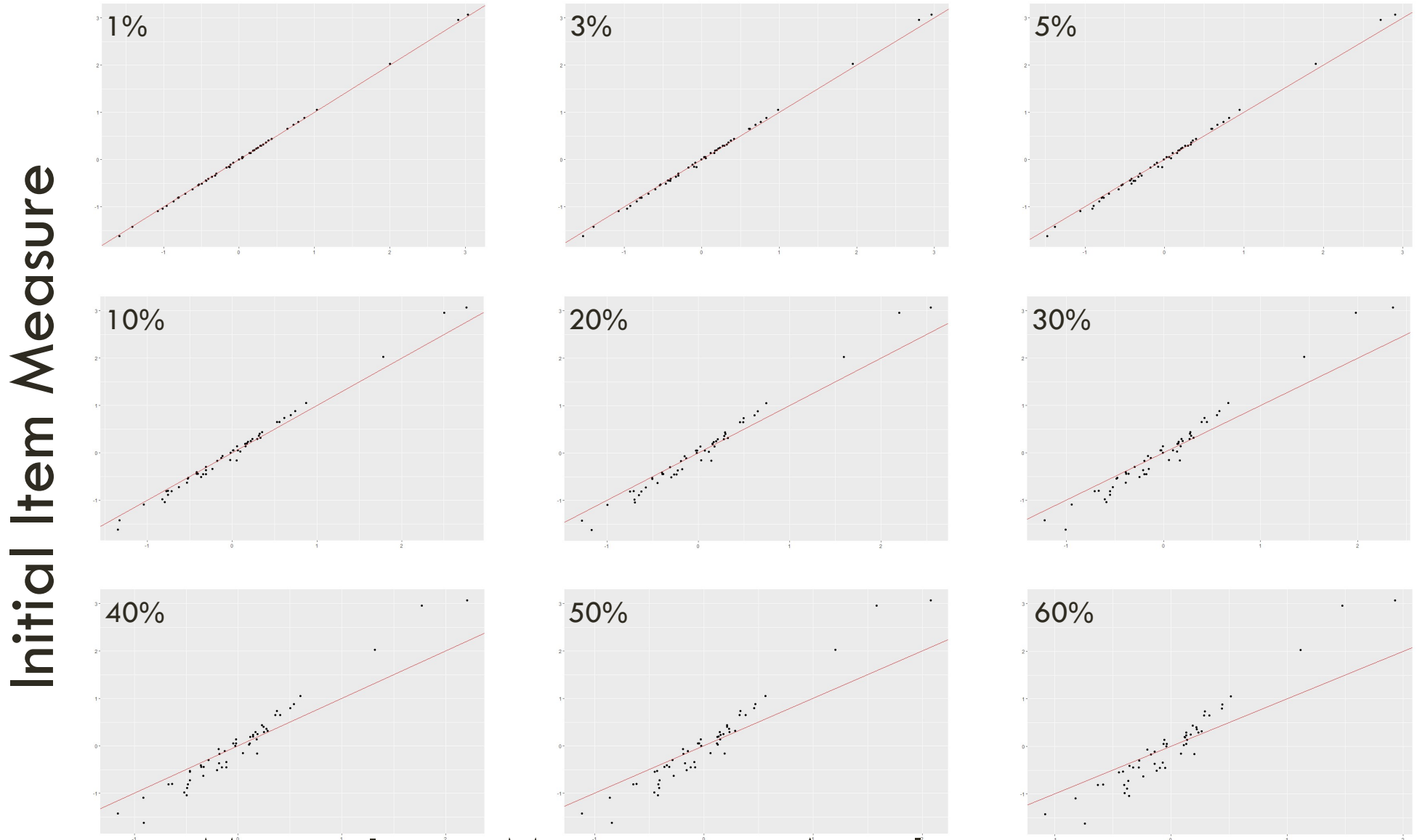
	Mean		Variance	
Percentage of data deleted	Winsteps	R	Winsteps	R
1	99.7	99.4	0.0004	0.0004
3	99.0	98.2	0.0013	0.0009
5	98.4	97.0	0.0020	0.0015
10	96.8	94.1	0.0041	0.0025
20	93.6	88.2	0.0085	0.0082
30	90.3	82.2	0.0243	0.0127
40	87.0	76.2	0.0377	0.0181
50	83.6	70.1	0.0401	0.0303
60	80.2	63.9	0.0000	0.0371

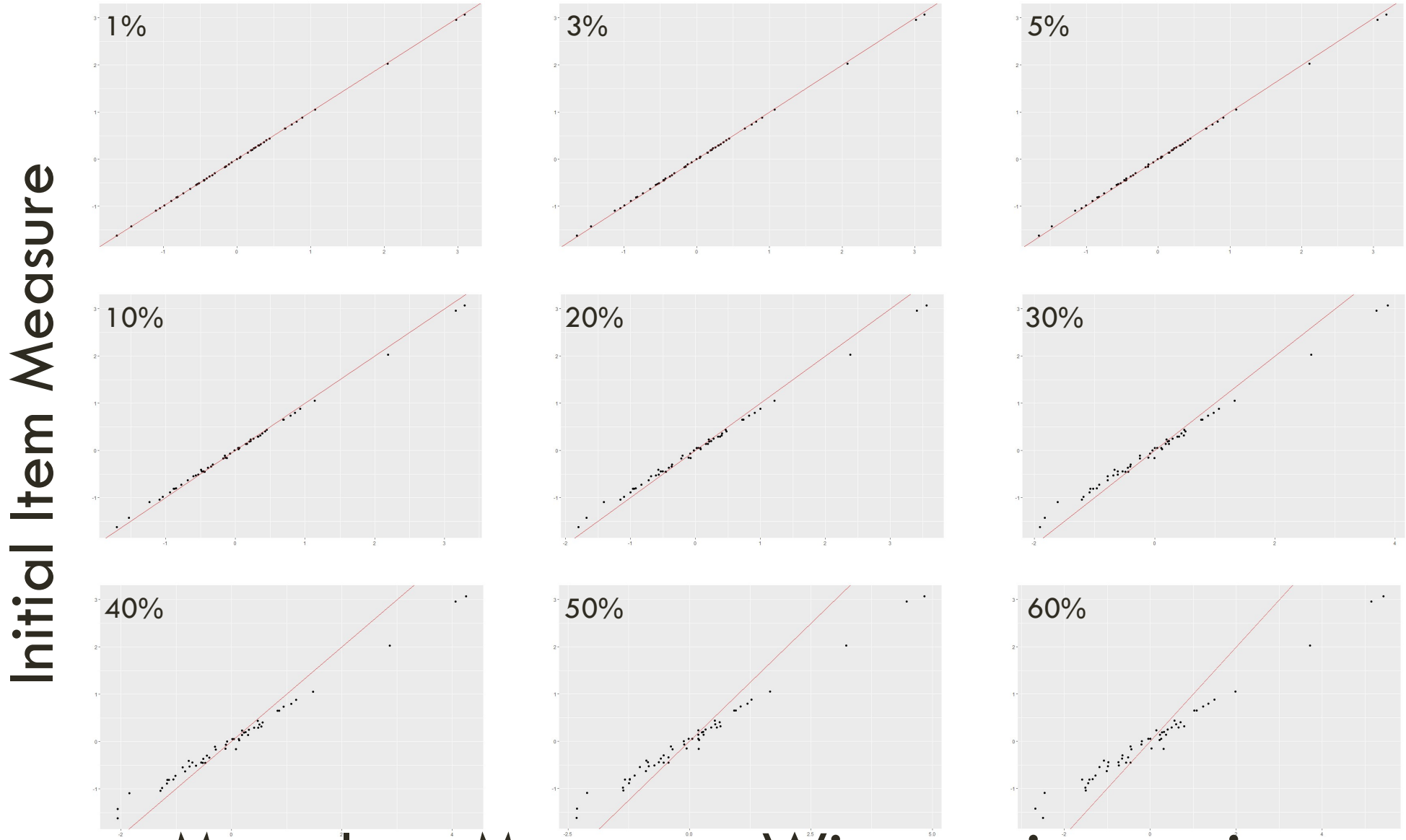


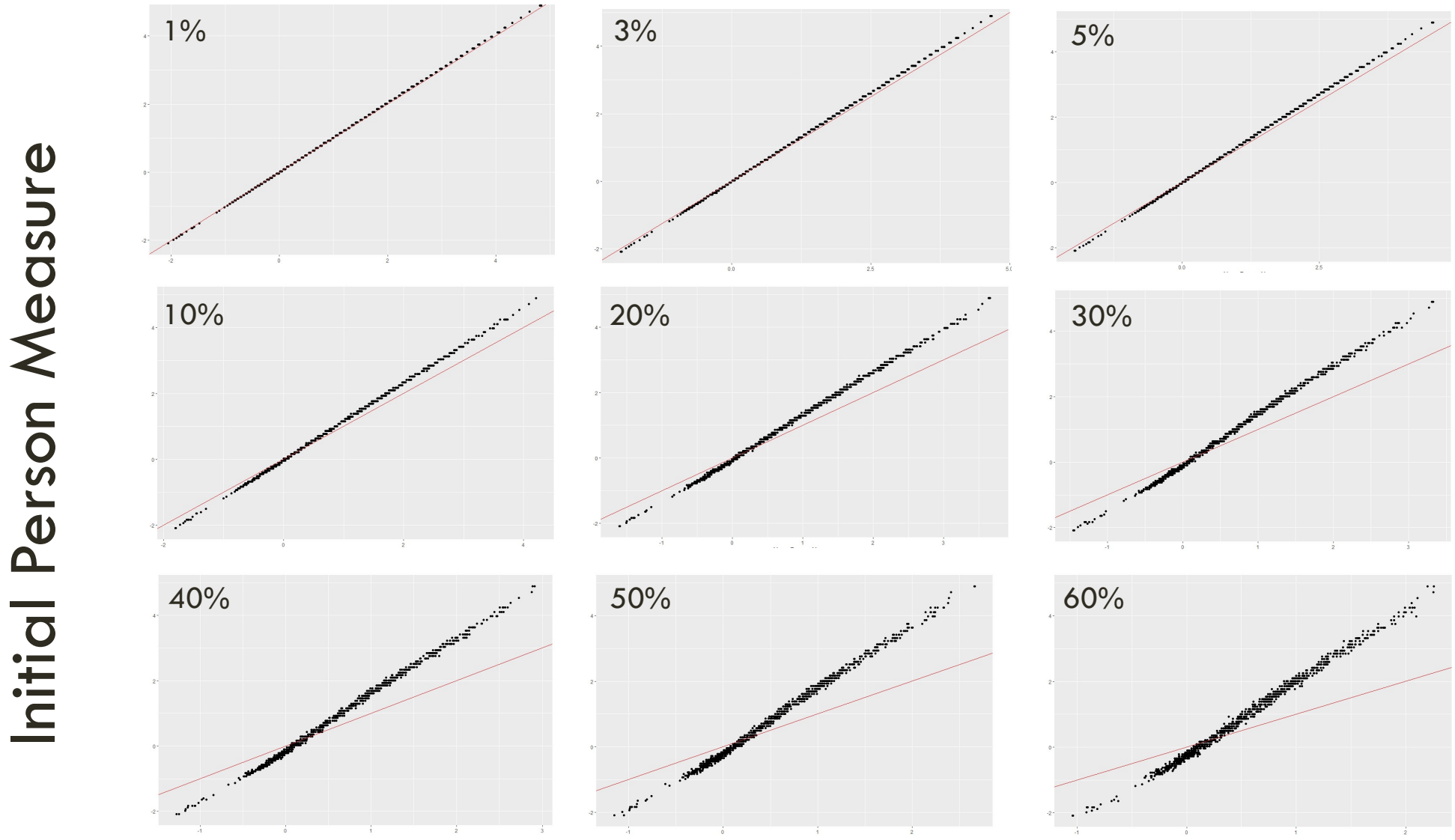
THE RESULTS:

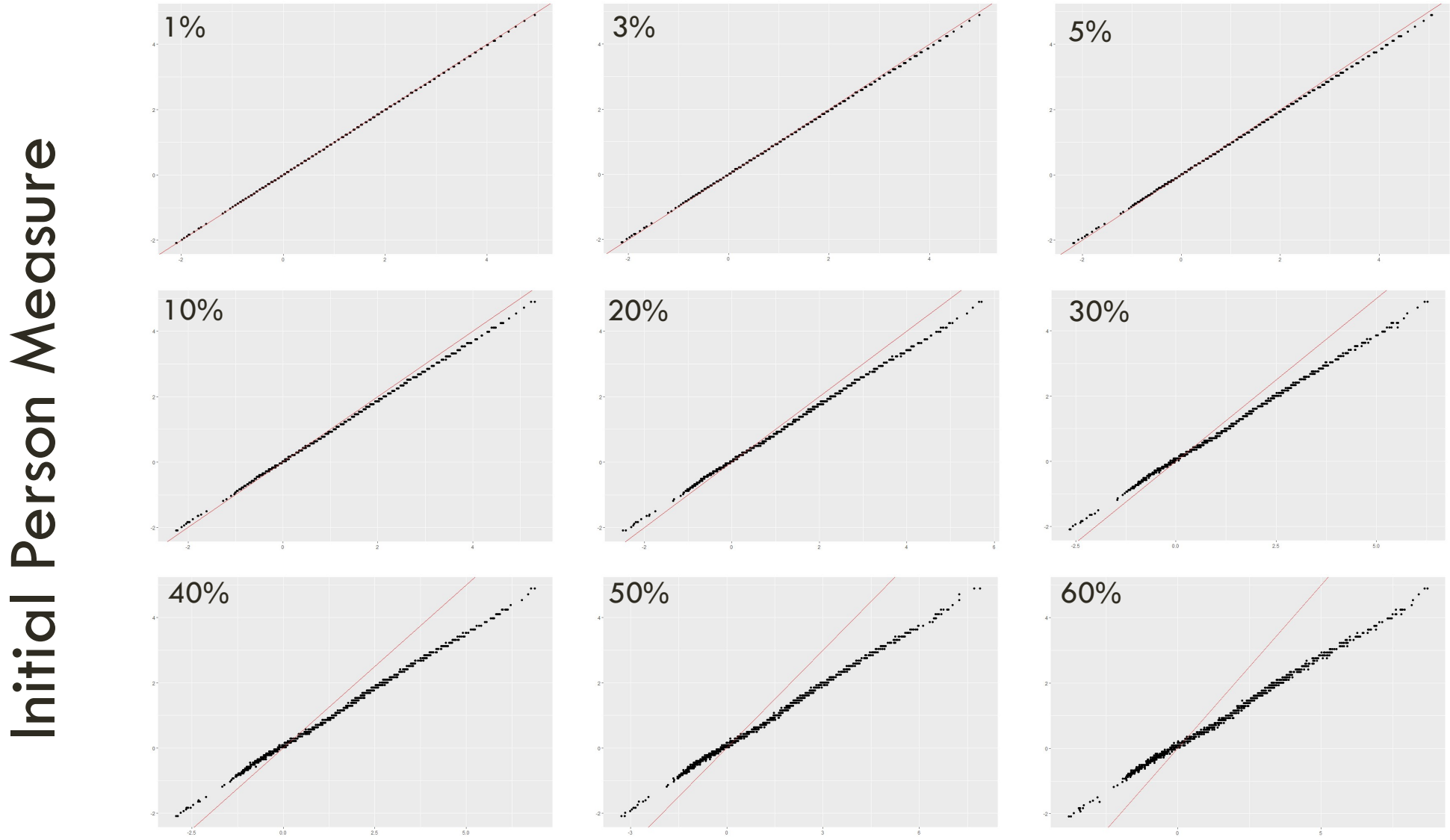
	Mean		Variance	
Percentage of data deleted	Winsteps	R	Winsteps	R
1	68.7	40.8	3.7	4.2
3	67.8	41.0	1.5	1.0
5	67.7	40.8	0.8	0.6
10	67.9	41.1	0.4	0.3
20	67.8	40.8	0.2	0.2
30	67.6	40.7	0.3	0.1
40	67.4	40.5	0.2	0.1
50	67.1	40.1	0.2	0.1
60	67.0	39.9	0.0	0.1




Initial Item Measure
1%
3%
5%
10%
20%
30%
40%
50%
60%
Mean Item Measure - Amelia imputations


1%
3%
5%
10%
20%
30%
40%
50%
60%
Initial Item Measure
Mean Item Measure - Winsteps imputations


1%
3%
5%
10%
20%
30%
40%
50%
60%
Initial Person Measure
Mean Person Measure - Amelia imputations


1%
3%
5%
10%
20%
30%
40%
50%
60%
Initial Person Measure
Mean Person Measure - Winsteps imputations

FURTHER STUDY

Sample size

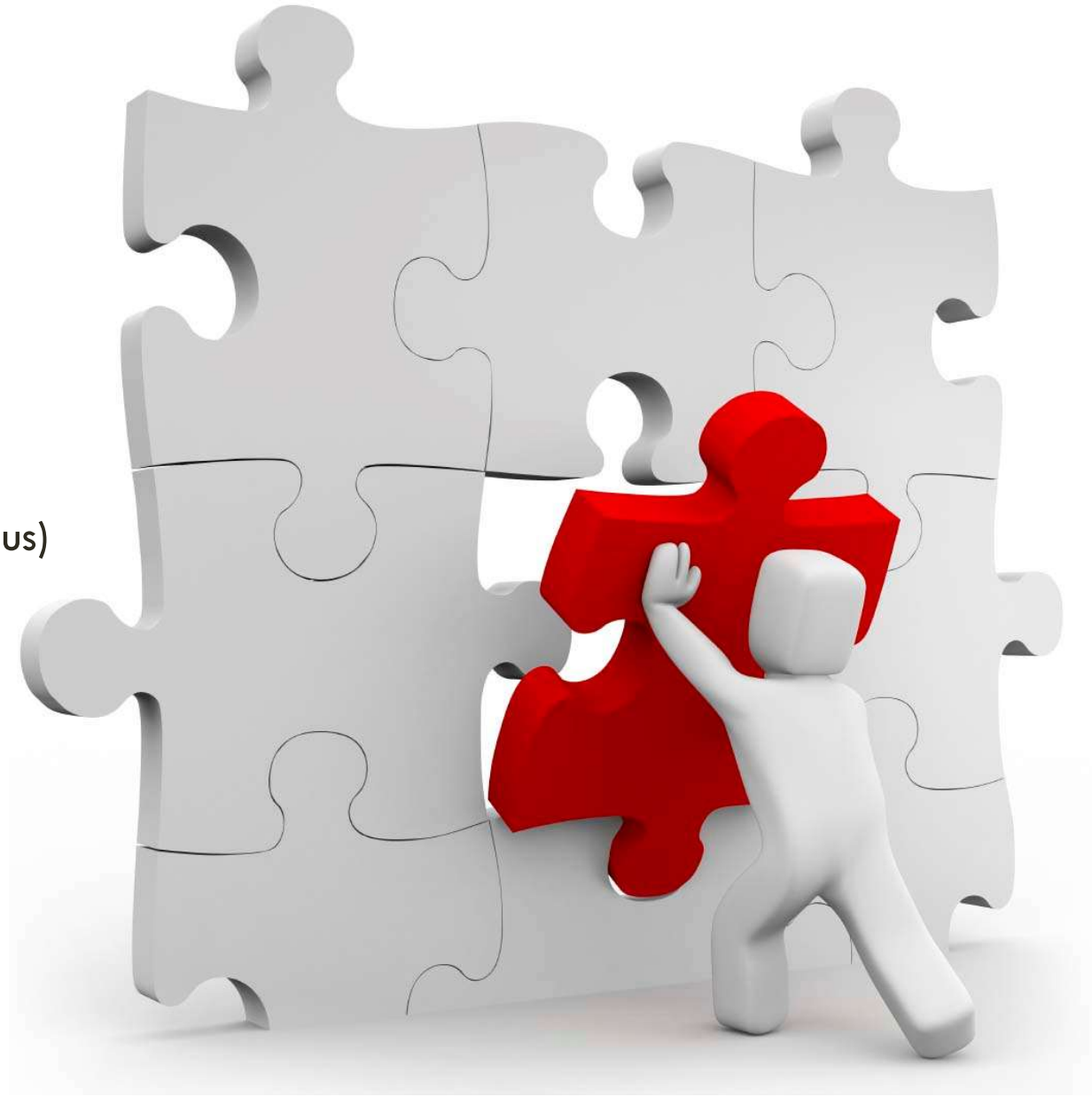
Number of items

Types of items (dichotomous/polytomous)

- If polytomous: number of score categories

Percentages missing

Cohort composition



REFERENCES

Payne, Nicole (2014). *Missing Puzzle Piece (Image)*. Available at: http://www.thesevencandles.com/2014/04/missing-puzzle-piece.html (Accessed: 19.03.2019).

Greene, Avalon (2018). *Missing Piece (Image)*. Available at: https://weheartit.com/articles/318419976-missing-piece (Accessed: 19.03.2019).

Wikipedia (2019). *R (programming language) (Image)*. Available at: https://en.wikipedia.org/wiki/R_(programming_language) (Accessed: 19.03.2019).

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

James Honaker, Gary King, Matthew Blackwell (2011). Amelia II: A Program for Missing Data. Journal of Statistical Software, 45(7), 1-47. URL http://www.jstatsoft.org/v45/i07/

ggplot in R: H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Linacre, J. M. (2019). Winsteps® Rasch measurement computer program. Beaverton, Oregon: Winsteps.com