FUKUSHIMA
MEDICAL
UNIVERSITY

Basic biostatistics: Supplement

Aya Goto
Center for Integrated Science and Humanities

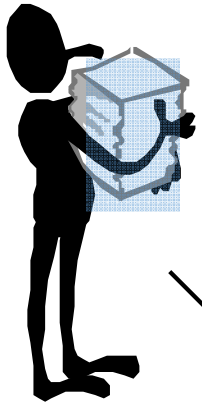# Research tool
Reliability and validity

Goto A, Yasumura S, Fukao A.
A reproductive health survey on unintended pregnancy in Yamagata, Japan: Feasibility of the survey and test-retest reliability and validity of a questionnaire.
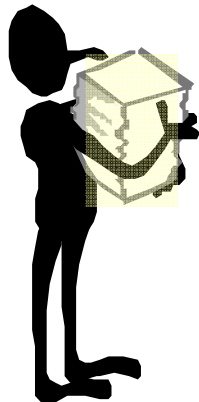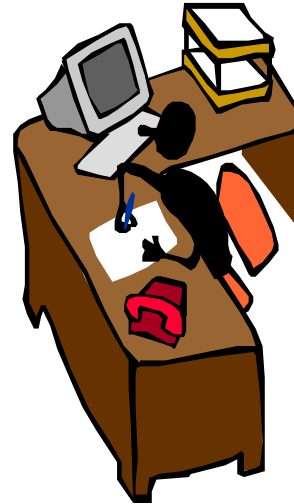Journal of Epidemiology. 2000; 10: 376-382.

Back-translation

Step 1
Translation
Eng. to VN

Step 2
Back-Translation
VN to Eng.

Step 3
Compare original and back-translated version, and address any differences in meaning.
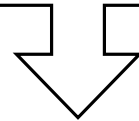
Step 4
Pilot Study

Reliability and validity

Developing a question to ask pregnancy intention in JPN

When you learned of your pregnancy, how did you feel?

(1) Pregnancy was at the right time

(2) Pregnancy was too soon

(3) I wanted a child but the pregnancy was too late

(4) I did not want to have a (any more) child even in the future

(1) and (3) = Intended Pregnancy
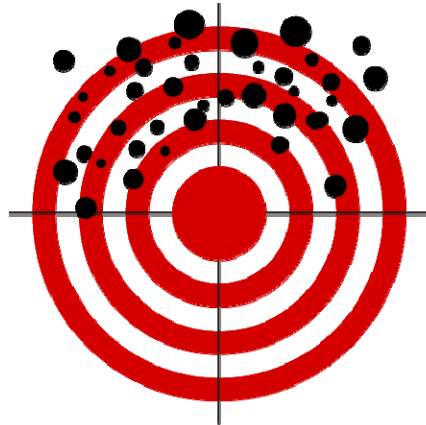
(2) = Mistimed Pregnancy

(4) = Unwanted Pregnancy

* Based on the definitions of intended, mistimed and unwanted pregnancies used in the National Survey of Family Growth (NSFG) in the US.
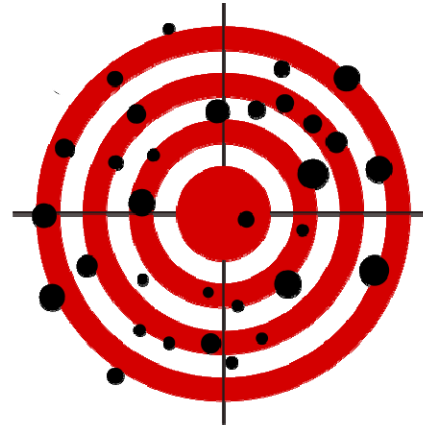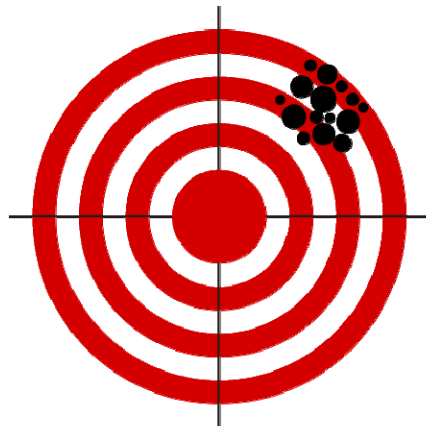
Not valid

Valid

Not reliable

Unreliable & Unvalid

Unreliable, But Valid
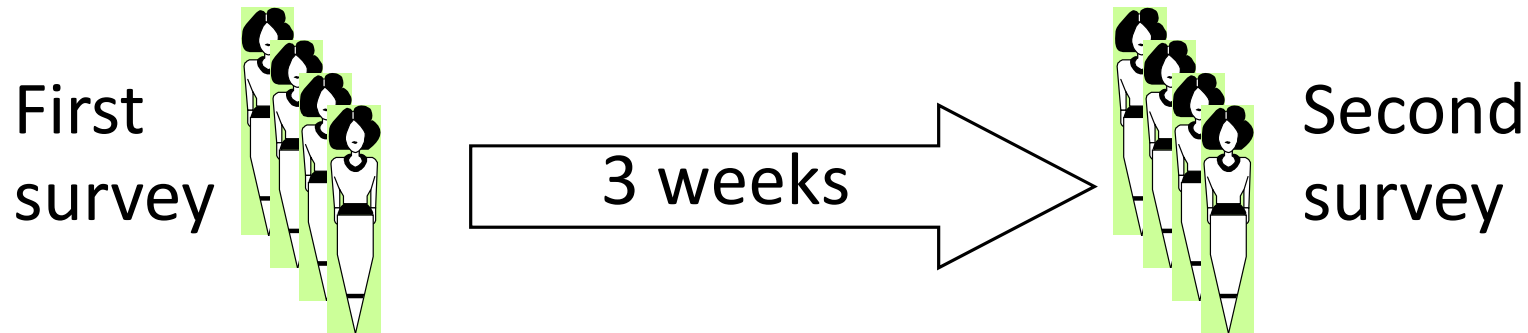
Reliable

Reliable, Not Valid

Both Reliable & Valid

SLIDE 5

Test-retest reliability

First survey → 3 weeks → Second survey

Convergent validity
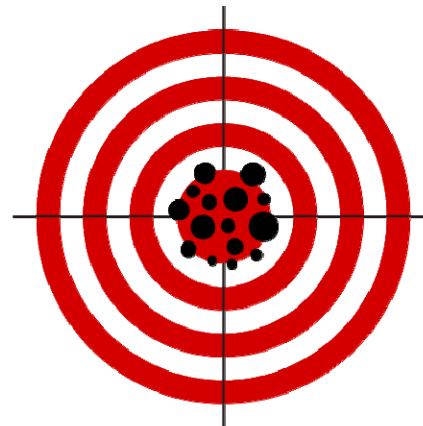
Pregnancy intention question ⟷ Feeling of happiness at the time of Pregnancy
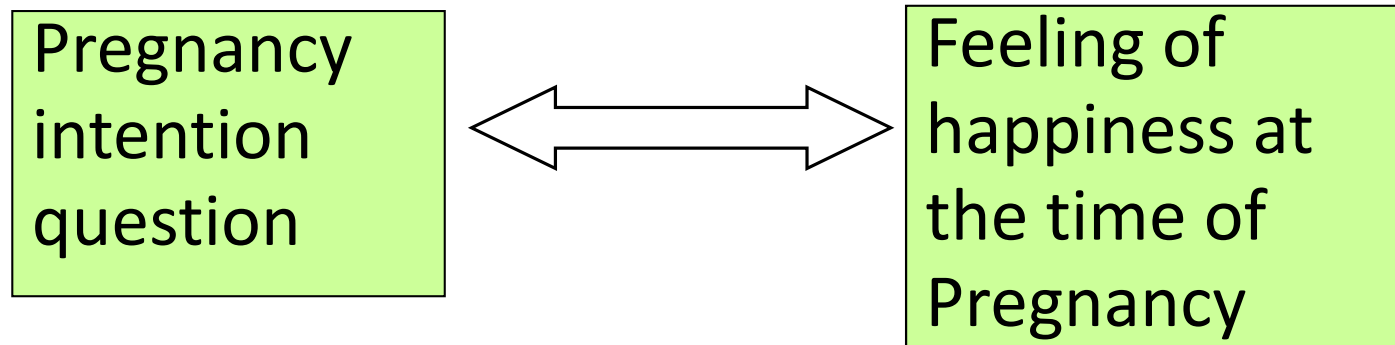
* The same method was applied in the NSFG.

Test-retest reliability

Experience of unintended pregnancy		Second survey	
		Yes	No
First survey	Yes	27	4
	No	3	22

Expected Agreement	Agreement	Kappa	Std. Err.	Z	Prob>Z
87.50%	50.38%	0.7481	0.1335	5.60	0.0000

Kappa

= Measure which quantifies the extent of agreement.

(The extent to which the observed agreement exceeds
that which would be expected by chance alone.)

Interpretation of kappa:

< 0.00	Poor agreement
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-	Almost perfect

Convergent validity

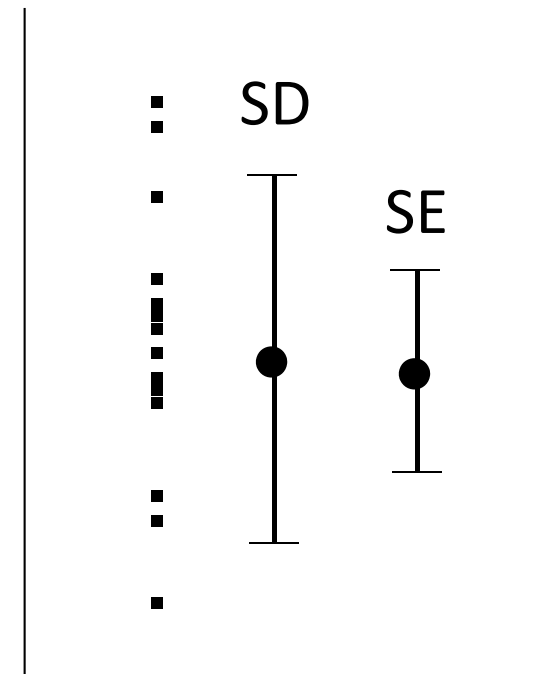|	Intended N=153	Mistimed N=36	Unwanted N=8
Feeling of happiness at the time of pregnancy [Median (min, max)]	10 (1, 10)	8 (3, 10)	5 (4, 8)

Data distribution

SD or SE

Mean (SD)

Standard deviation (SD) Descriptive
Average difference between the
data points and their mean

Standard error (SE)
Inferential
A measure of how variable the
mean will be, if you repeat the
whole study many times.

Cumming G., et al. Error bars in
experimental biology. JCB 2007; 17: 7-11.

⚠️ When both are not applicable...

"Because experimental biologists are usually trying to compare experimental results with controls, it is usually appropriate to show inferential error bars, such as SE or CI, rather than SD.

However, if n is very small (for example $n = 3$), rather than showing error bars and statistics, it is better to simply plot the individual data points."

Cumming G., et al. Error bars in experimental biology. JCB 2007; 177; 7-11.