University of Verona,

School of Exercise and Sport Science,

Laurea magistrale in Scienze motorie preventive ed adattate

(Laurea magistrale in Scienze dello sport e della prestazione fisica)

Metodologia delle misure delle attività sportive

Friday 16/11/2018 8:30÷10

Luca P. Ardigò Ph.D.

A class within an eight-class module

Sensitivity and specificity

Sensitivity and **specificity** are statistical measures of the performance of a binary classification test, also known in statistics as classification function:

- **Sensitivity** (also called the **true positive rate**, the **recall**, or **probability of detection**[1] in some fields) measures the proportion of positives that are correctly identified as such (e.g. the percentage of sick people who are correctly identified as having the condition).
- **Specificity** (also called the **true negative rate**) measures the proportion of negatives that are correctly identified as such (e.g. the percentage of healthy people who are correctly identified as not having the condition).

Precision, sensitivity, and specificity

Terminology and derivations from a confusion matrix

(number of) positive samples (P)

(number of) negative samples (N)

(number of) true positive (TP)

eqv. with hit

(number of) true negative (TN)

eqv. with correct rejection

(number of) false positive (FP)

eqv. with false alarm, Type I error

(number of) false negative (FN)

eqv. with miss, Type II error

---

sensitivity or true positive rate (TPR)

eqv. with hit rate, recall

$$TPR = TP/P = TP/(TP + FN)$$

specificity (SPC) or true negative rate

$$SPC = TN/N = TN/(TN + FP)$$

precision or positive predictive value (PPV)

$$PPV = TP/(TP + FP)$$

Accuracy

accuracy (ACC)

$$ACC = (TP + TN)/(TP + FP + FN + TN)$$

Confusion matrix

		True condition	
	Total population	Condition positive	Condition negative
Predicted condition	Predicted condition positive	**True positive**	**False positive**, Type I error
	Predicted condition negative	**False negative**, Type II error	**True negative**

Table of error types

Table of error types		Null hypothesis (H_0) is	
		True	**False**
Decision About Null Hypothesis (H_0)	**Reject**	Type I error (False Positive)	Correct inference (True Positive)
	Fail to reject	Correct inference (True Negative)	Type II error (False Negative)

Example 1 [edit]

Hypothesis: "Adding water to toothpaste protects against cavities."

Null hypothesis (H_0): "Adding water to toothpaste has no effect on cavities."

This null hypothesis is tested against experimental data with a view to nullifying it with evidence to the contrary.

A type I error occurs when detecting an effect (adding water to toothpaste protects against cavities) that is not present. The null hypothesis is true (i.e., it is true that adding water to toothpaste has no effect on cavities), but this null hypothesis is rejected based on bad experimental data.

Example 2 [edit]

Hypothesis: "Adding fluoride to toothpaste protects against cavities."

Null hypothesis (H_0): "Adding fluoride to toothpaste has no effect on cavities."

This null hypothesis is tested against experimental data with a view to nullifying it with evidence to the contrary.

A type II error occurs when failing to detect an effect (adding fluoride to toothpaste protects against cavities) that is present. The null hypothesis is false (i.e., adding fluoride is actually effective against cavities), but the experimental data is such that the null hypothesis cannot be rejected.

Example 3 [edit]

Hypothesis: "The evidence produced before the court proves that this man is guilty."

Null hypothesis (H_0): "This man is innocent."

A type I error occurs when convicting an innocent person (a miscarriage of justice). A type II error occurs when letting a guilty person go free (an error of impunity).

A positive correct outcome occurs when convicting a guilty person. A negative correct outcome occurs when letting an innocent person go free.

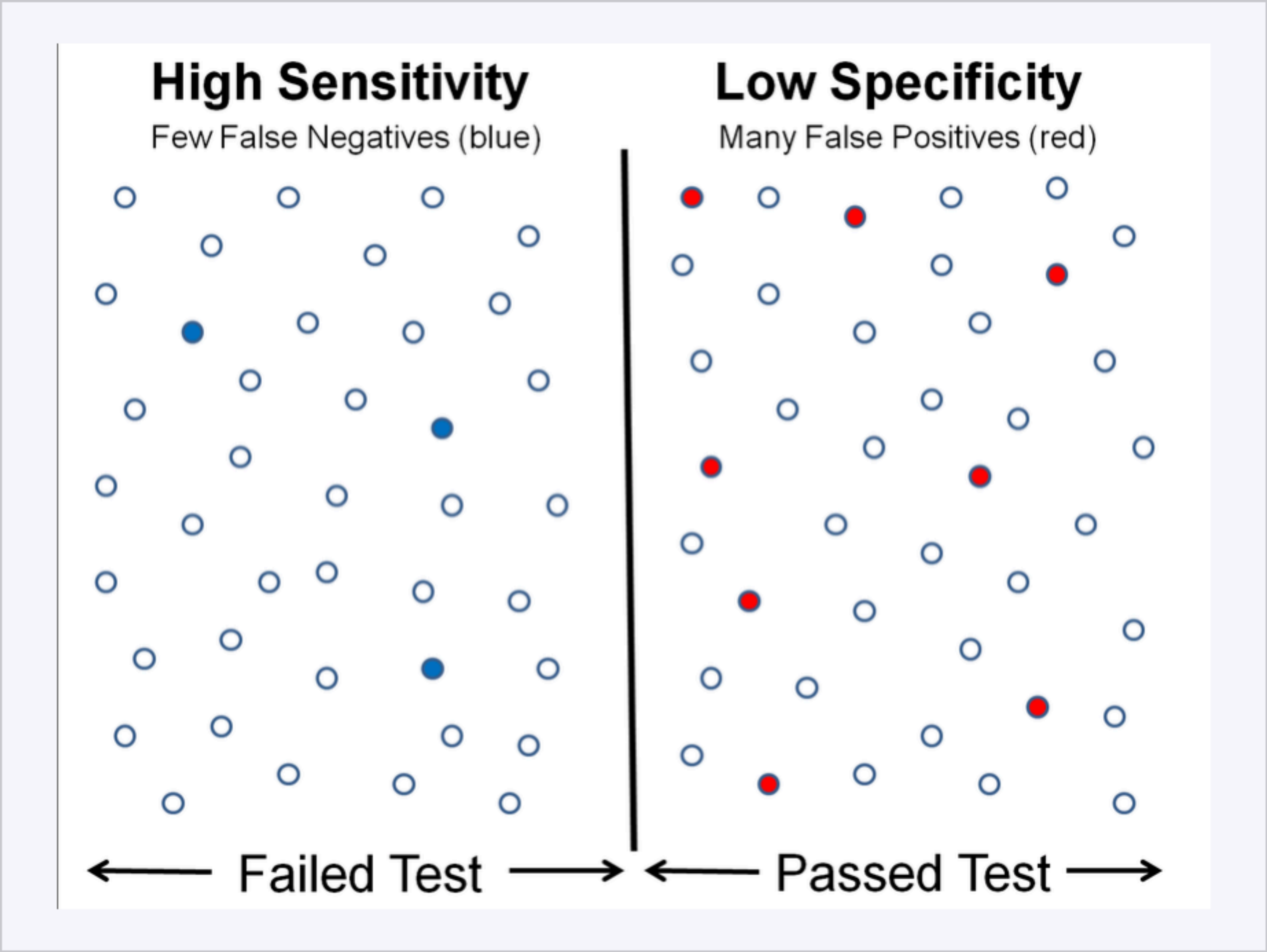
Example 4 [edit]

Hypothesis: "A patient's symptoms improve after treatment A more rapidly than after a placebo treatment."

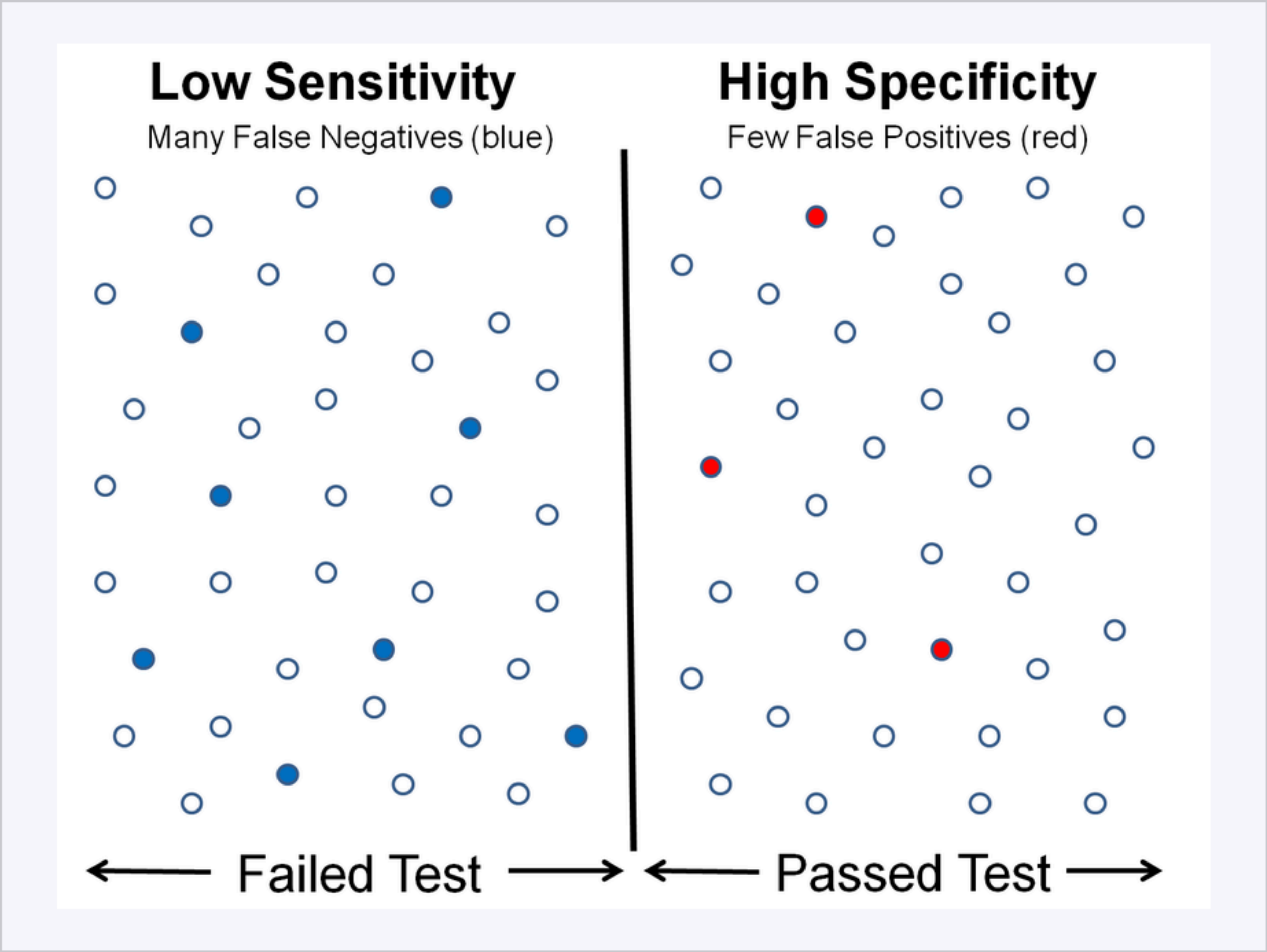
Null hypothesis (H_0): "A patient's symptoms after treatment A are indistinguishable from a placebo."

A Type I error would falsely indicate that treatment A is more effective than the placebo, whereas a Type II error would be a failure to demonstrate that treatment A is more effective than placebo even though it actually is more effective.

Sensitivity and specificity



High sensitivity and low specificity



Low sensitivity and high specificity

Worked example

		Patients with bowel cancer (as confirmed on endoscopy)	
		Condition positive	Condition negative
Fecal occult blood screen test outcome	Test outcome positive	**True positive** (TP) = 20	**False positive** (FP) = 180
	Test outcome negative	**False negative** (FN) = 10	**True negative** (TN) = 1820
		Sensitivity = TP / (TP + FN) = 20 / (20 + 10) ≈ **67%**	**Specificity** = TN / (FP + TN) = 1820 / (180 + 1820) = **91%**

Convergent validity

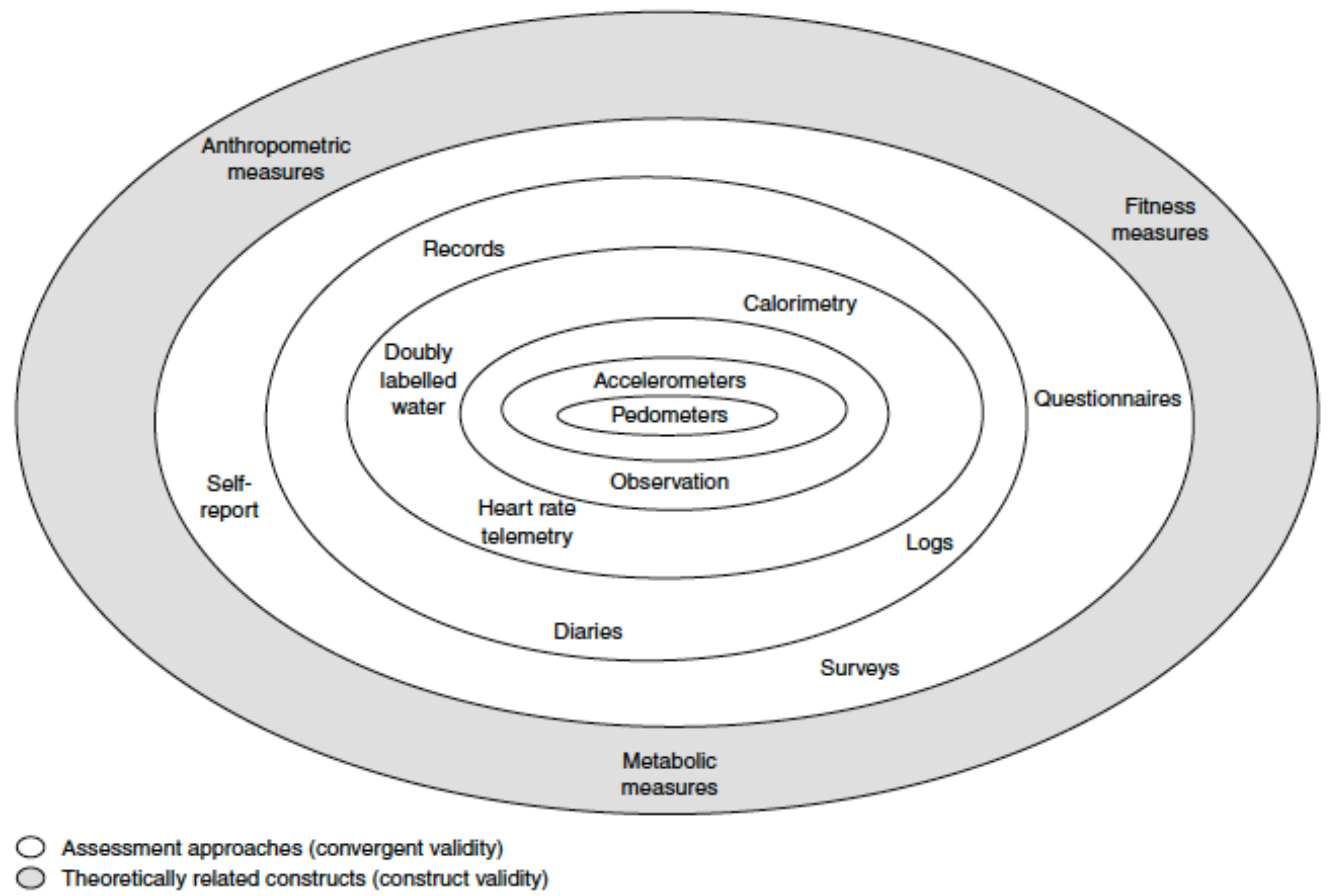


Fig. 1. Concentric circles of anticipated concordance between pedometer and other physical activity measures.

vs. accelerometer;
vs. observation;
vs. HR, V'O2, DLW;
vs. self-report diary

Tudor-Locke et al., 2002

Construct validity

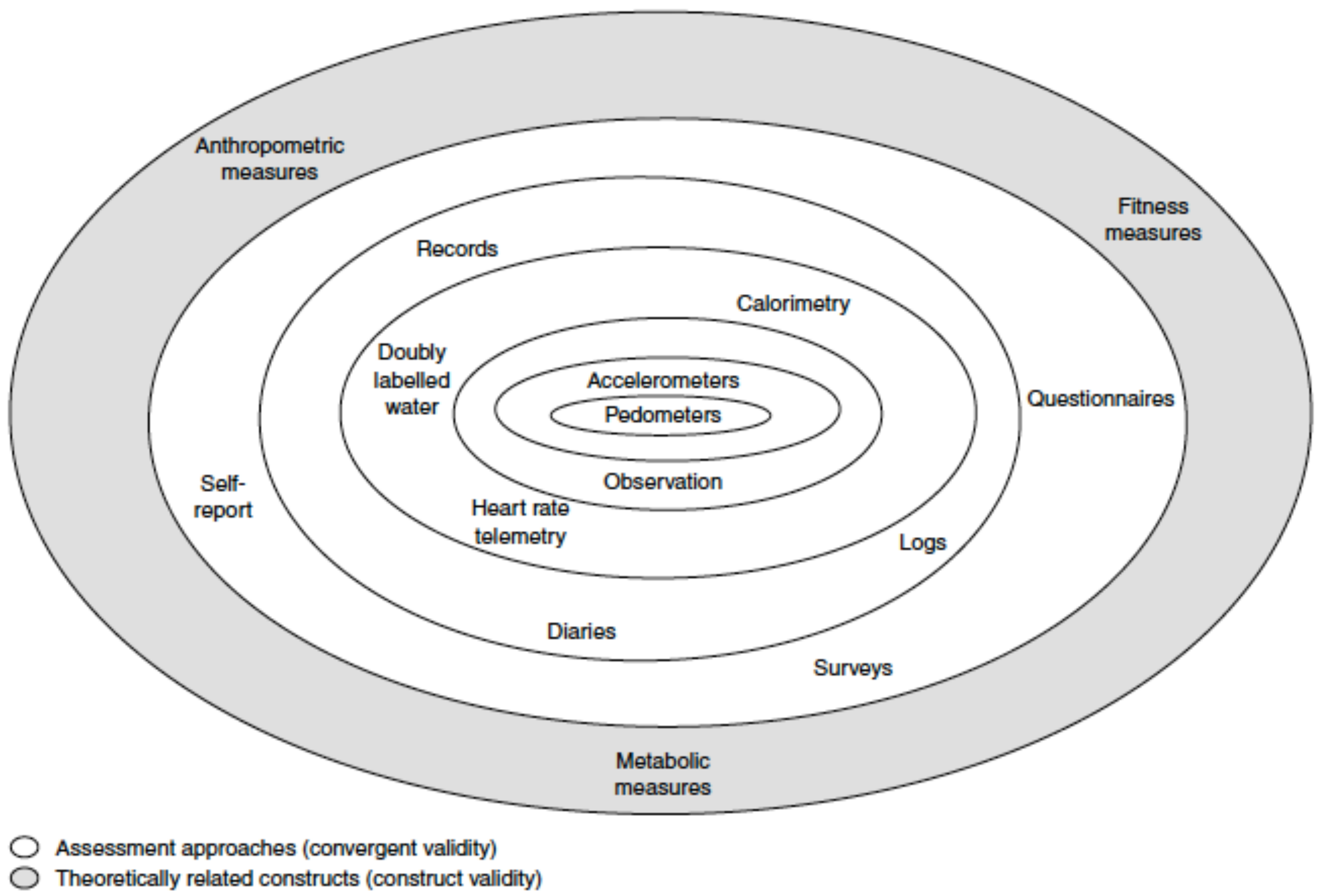


Fig. 1. Concentric circles of anticipated concordance between pedometer and other physical activity measures.

w/age;
w/anthropometry;
w/fitness measures

Tudor-Locke et al., 2004

Pedometer accuracy/validity

measures

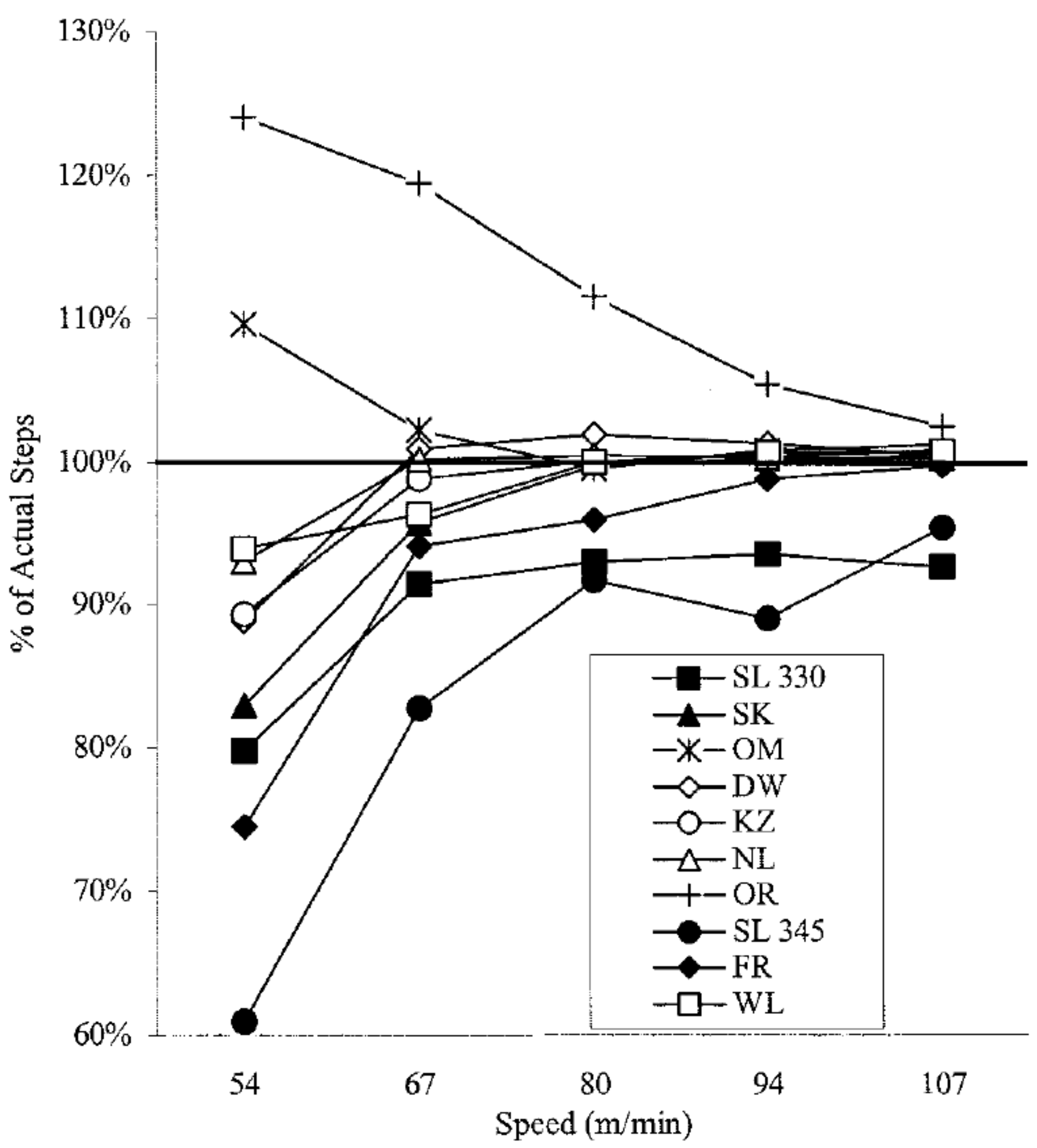


FIGURE 1—Effect of speed on pedometer accuracy (percentage of actual steps) during treadmill walking.

step #

Crouter et al., 2003

Pedometer accuracy/validity

measures

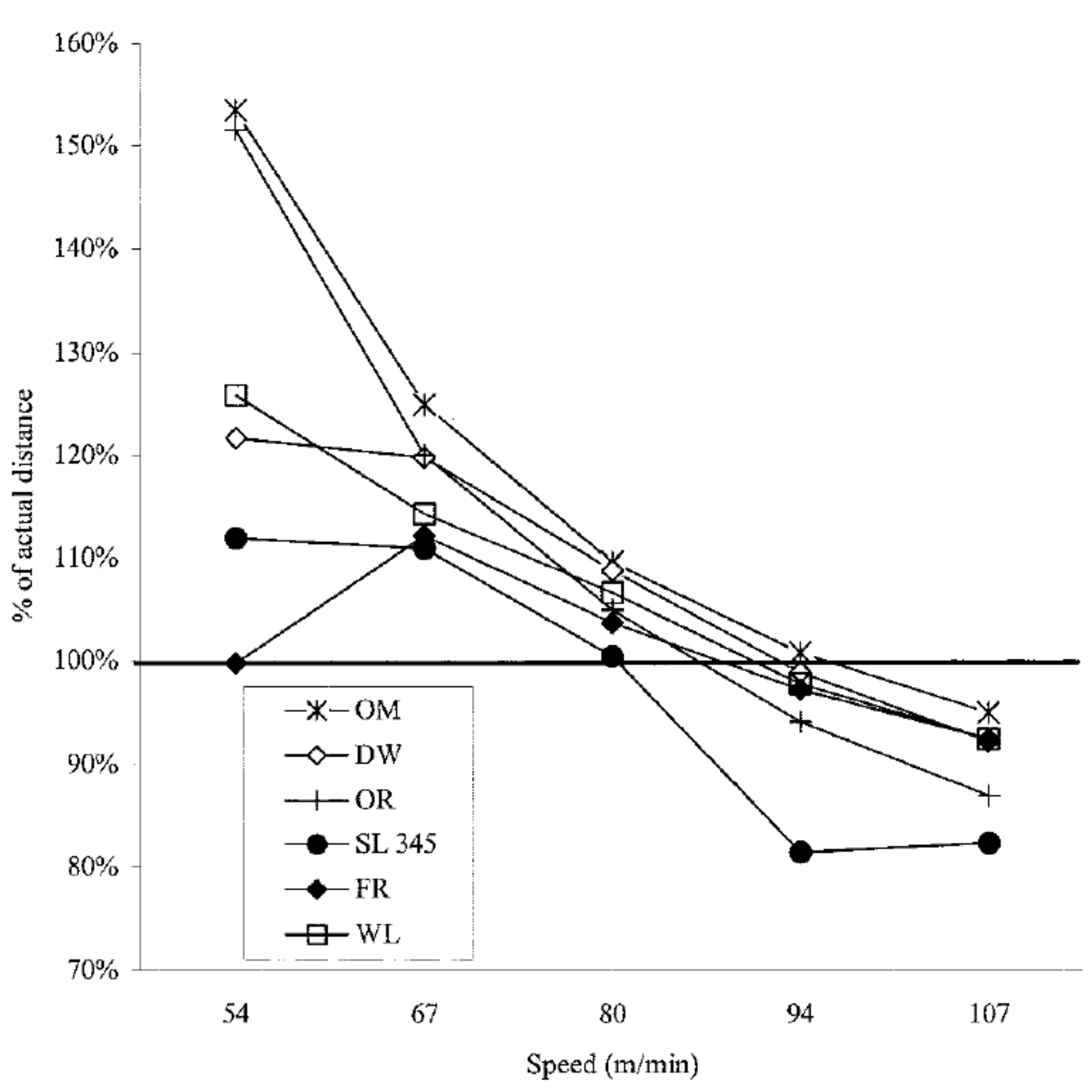


FIGURE 2—Effect of speed on pedometer estimates of percentage of actual distance traveled during treadmill walking.

(estimated) speed

Crouter et al., 2003

Pedometer accuracy/validity

measures

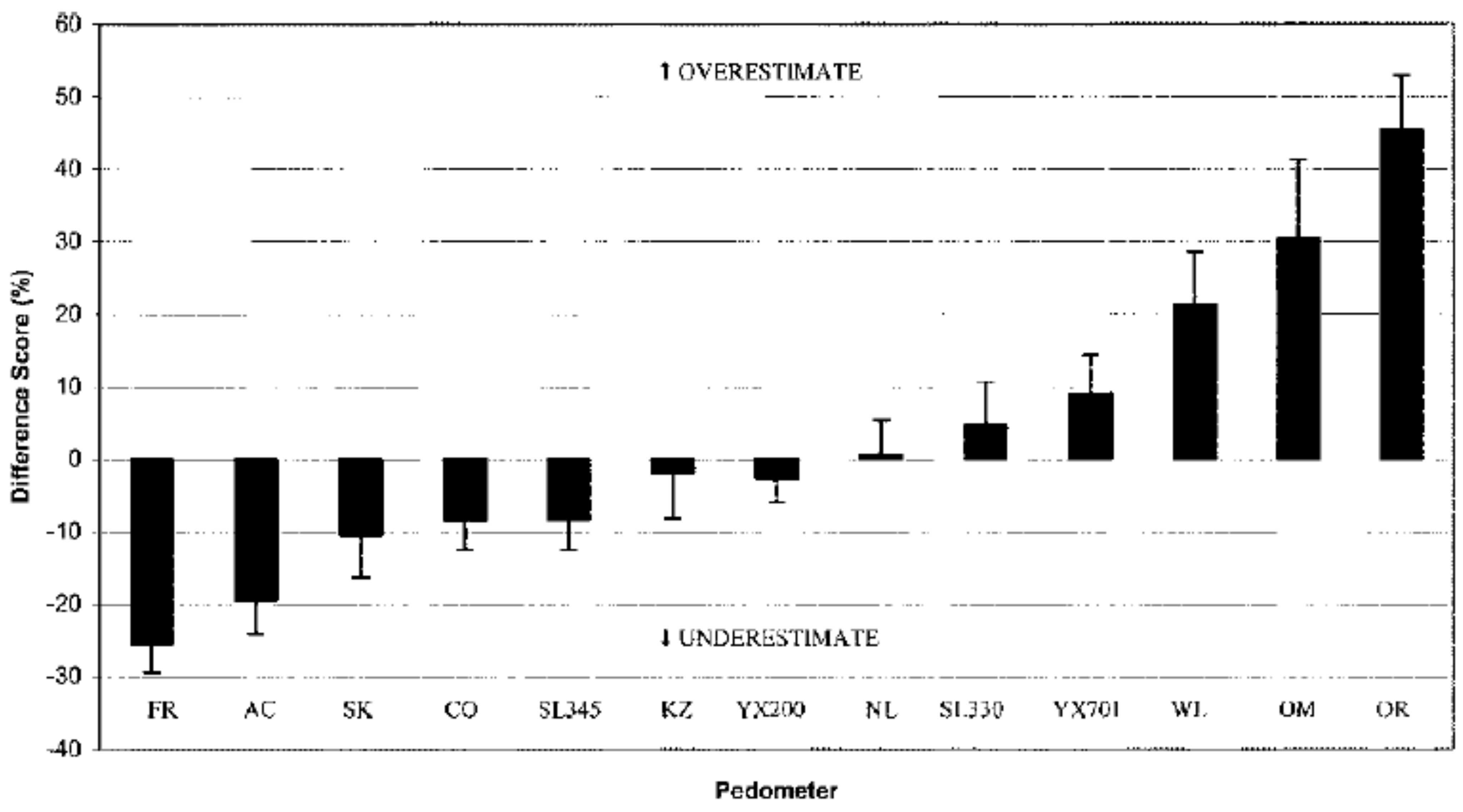


FIGURE 1—Mean difference scores [(comparison − criterion pedometer)/criterion] ± SE as a percentage of the criterion estimated steps over a 24-h period. Positive difference scores represent overestimations, and negative difference scores indicate underestimations of steps compared with the criterion pedometer.

step/day #

Schneider et al., 2004

Pedometer accuracy/validity

measures

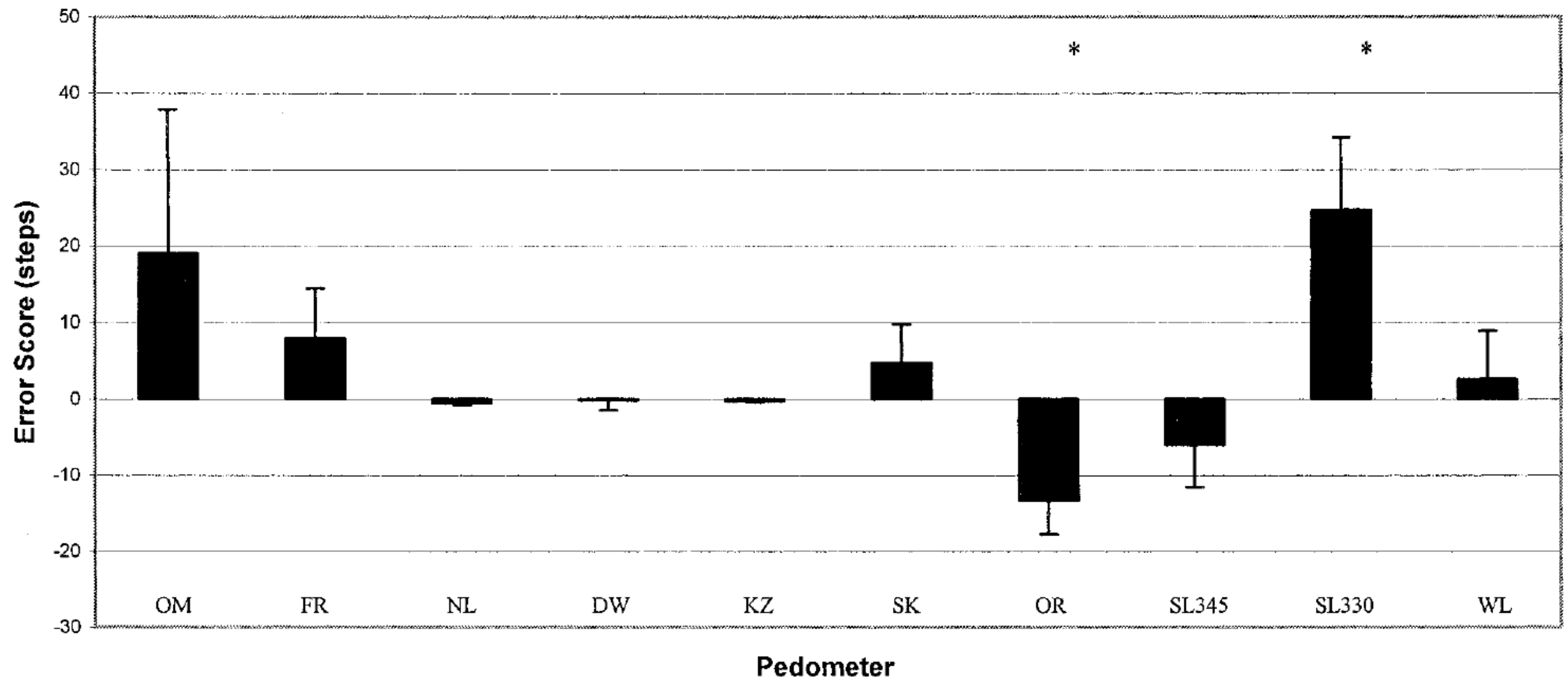


FIGURE 1—Mean error scores (actual − pedometer) ± SE in number of steps during a 400-m track walk at self-selected speeds. * $P < 0.05$.

400-m step #

Schneider et al., 2003

Pedometer accuracy/validity

measures

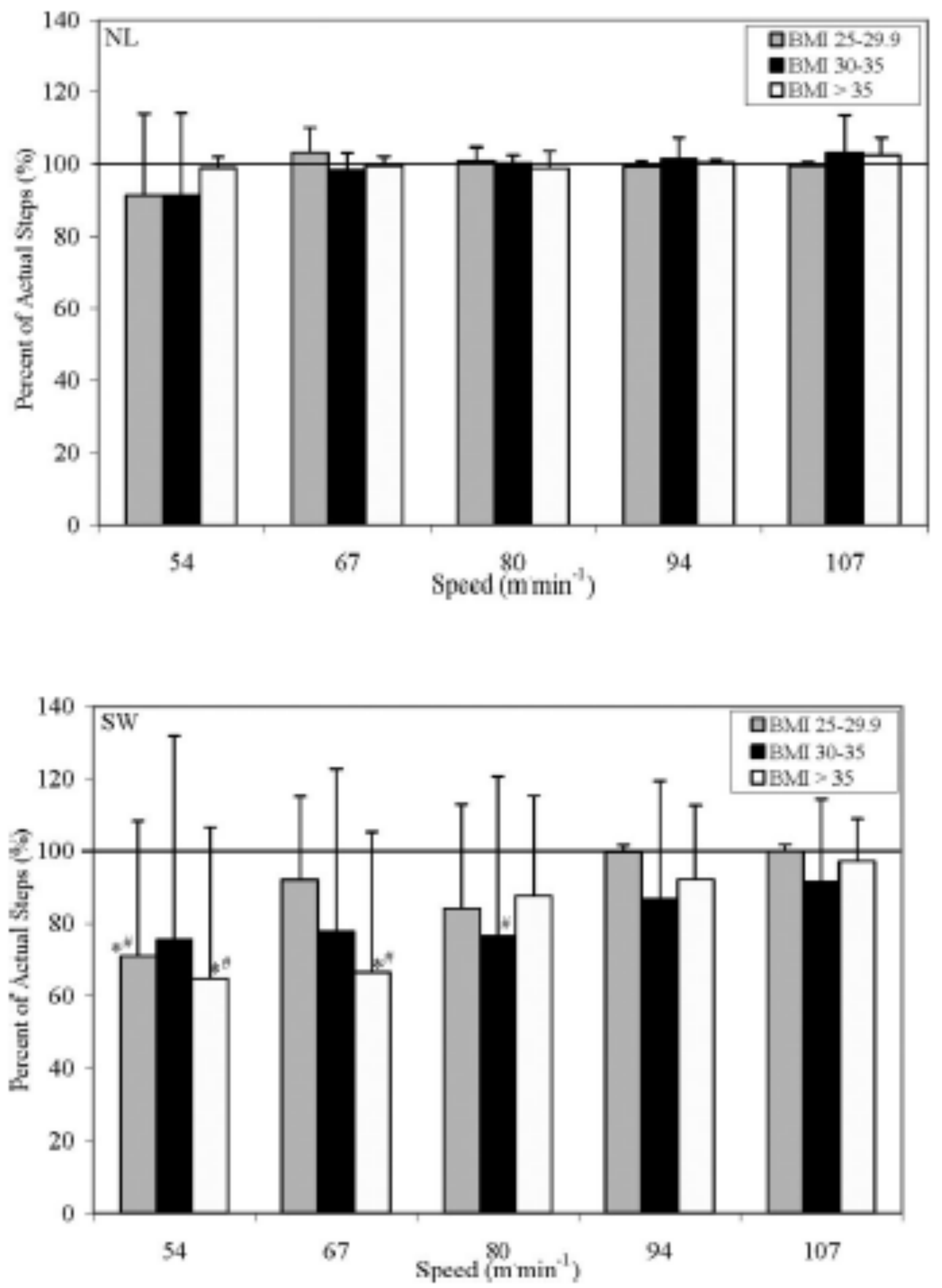


FIGURE 2—Effect of BMI (25–29.9 kg·m^{-2}, 30–35 kg·m^{-2}, and >35 kg·m^{-2}) on the percent of actual steps recorded by the New-Lifestyles NL-2000 (NL) and Yamax Digiwalker SW-200 (SW). Error bars are standard deviation. * Significantly different from actual steps; # significantly different from the NL ($P < 0.05$).

Crouter et al., 2005

(uniaxially accelerometric)

step #

(electromechanical circuit based)

Pedometer

Final pedometry issues

- no discrimination of weight lifting, gradient legged locomotion, cycling, swimming, rowing;
- shoe or ankle accelerometric pedometer -> stride #

Vetrovsky *et al. BMC Public Health* (2018) 18:635
https://doi.org/10.1186/s12889-018-5520-8

BMC Public Health

RESEARCH ARTICLE Open Access

CrossMark

A pedometer-based walking intervention with and without email counseling in general practice: a pilot randomized controlled trial

Tomas Vetrovsky[1*], Jozef Cupka[2], Martin Dudek[3], Blanka Kuthanova[4], Klaudia Vetrovska[5], Vaclav Capek[6] and Vaclav Bunc[1]

2018 study example

Pedometer

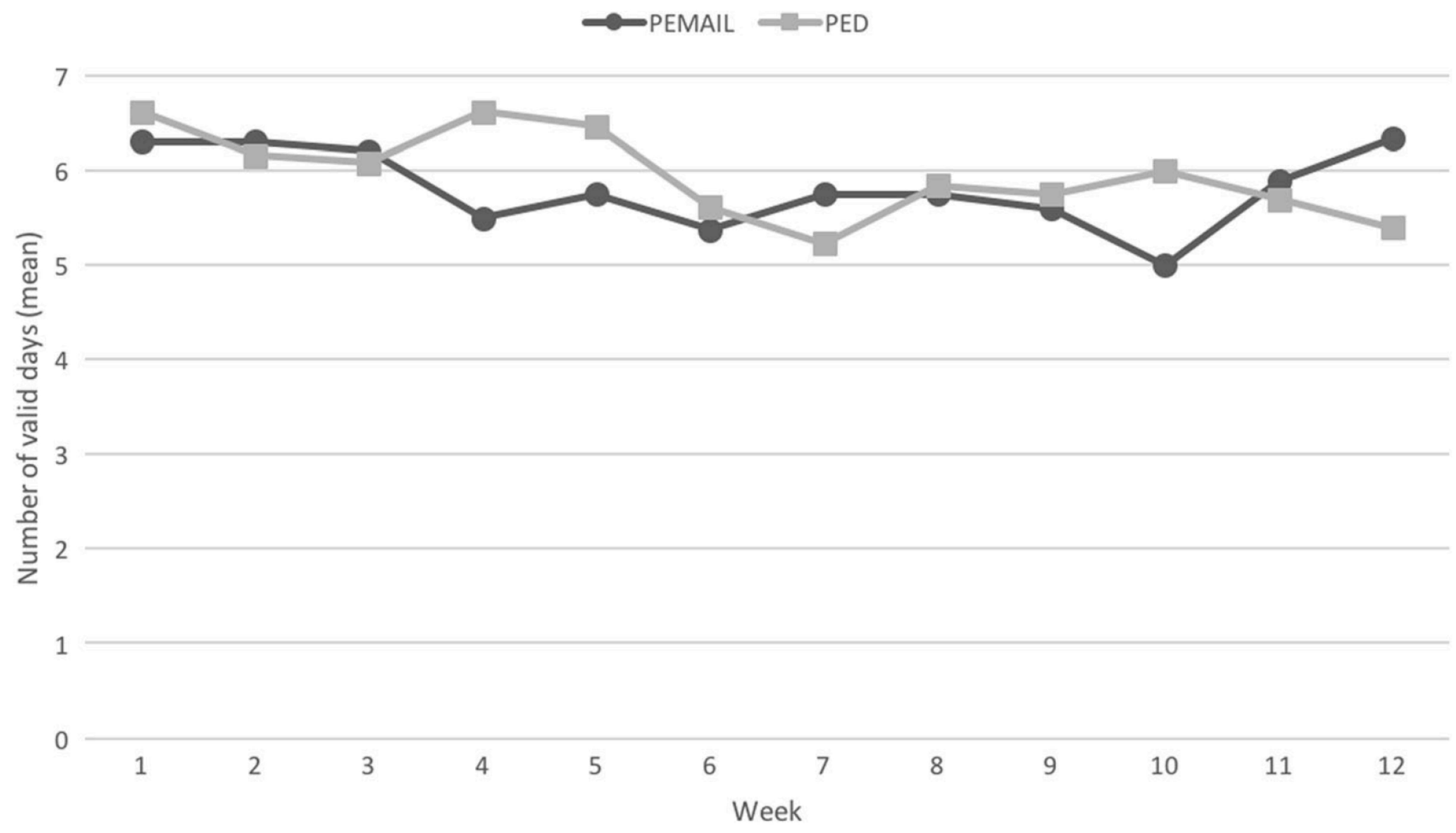


Fig. 2 Adherence to pedometer wear during the intervention period. The effect of time was significant (*P* = 0.008), whereas the effect of group was not

Pedometer

2018 study example

Table 1 Baseline characteristics of study participants, mean (SD)

	Pedometer-plus-email (n = 10)	Pedometer-alone (n = 13)
Age (yr)	44 (10)	39 (9)
BMI (kg/m^2)	33 (7)	33 (8)
Females (%)	30	62
Systolic blood pressure (mm Hg)	133 (9)	130 (18)
Diastolic blood pressure (mm Hg)	89 (10)	83 (15)
Waist circumference (cm)	114 (17)	102 (17)
Hip circumference (cm)	116 (10)	115 (17)
Steps per day	5034 (1431)	5050 (1393)

Pedometer

2018 study example

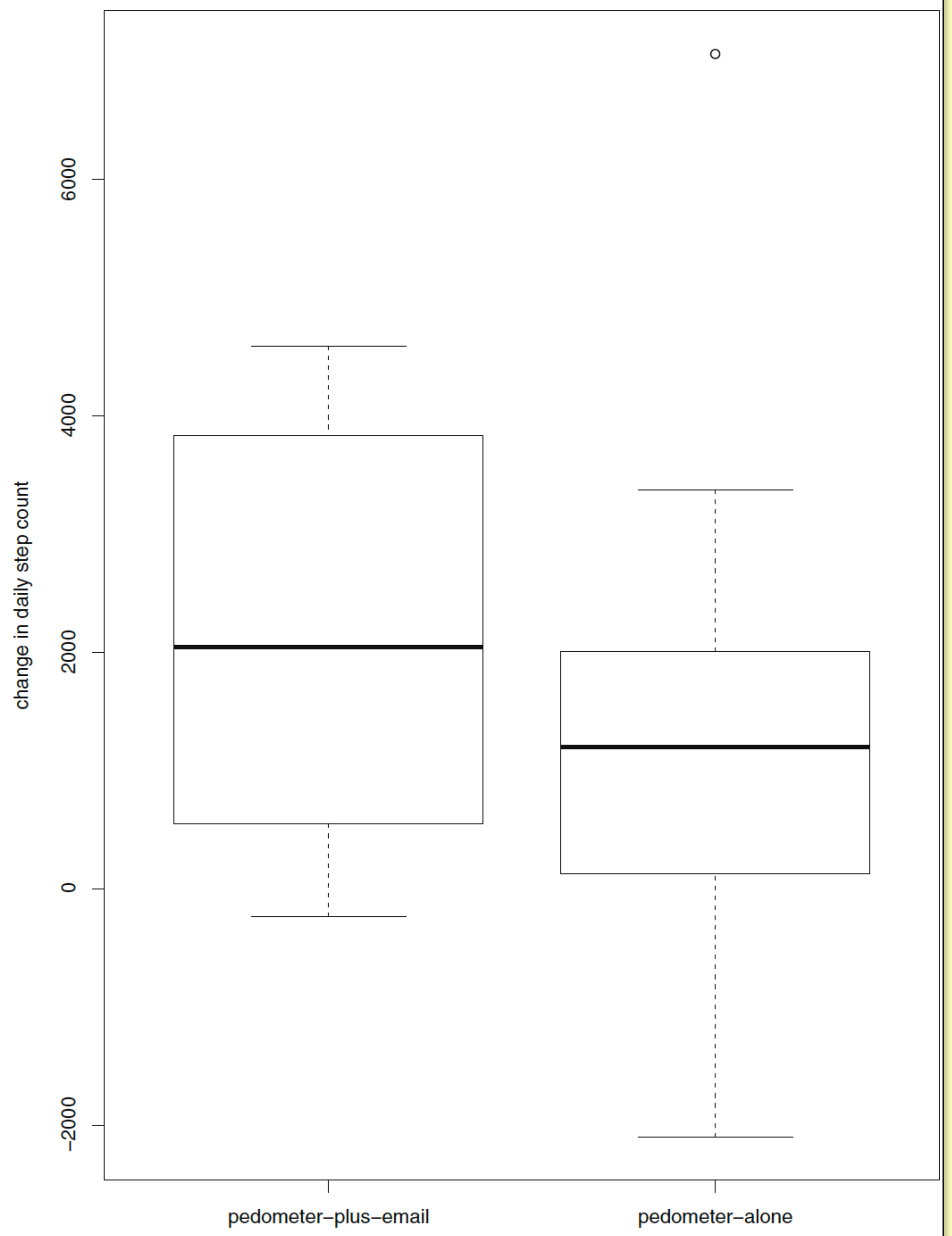


Fig. 3 Changes in the number of steps per day from baseline to post-intervention. The difference between groups was not significant ($p = 0.36$)

Pedometer

2018 study example

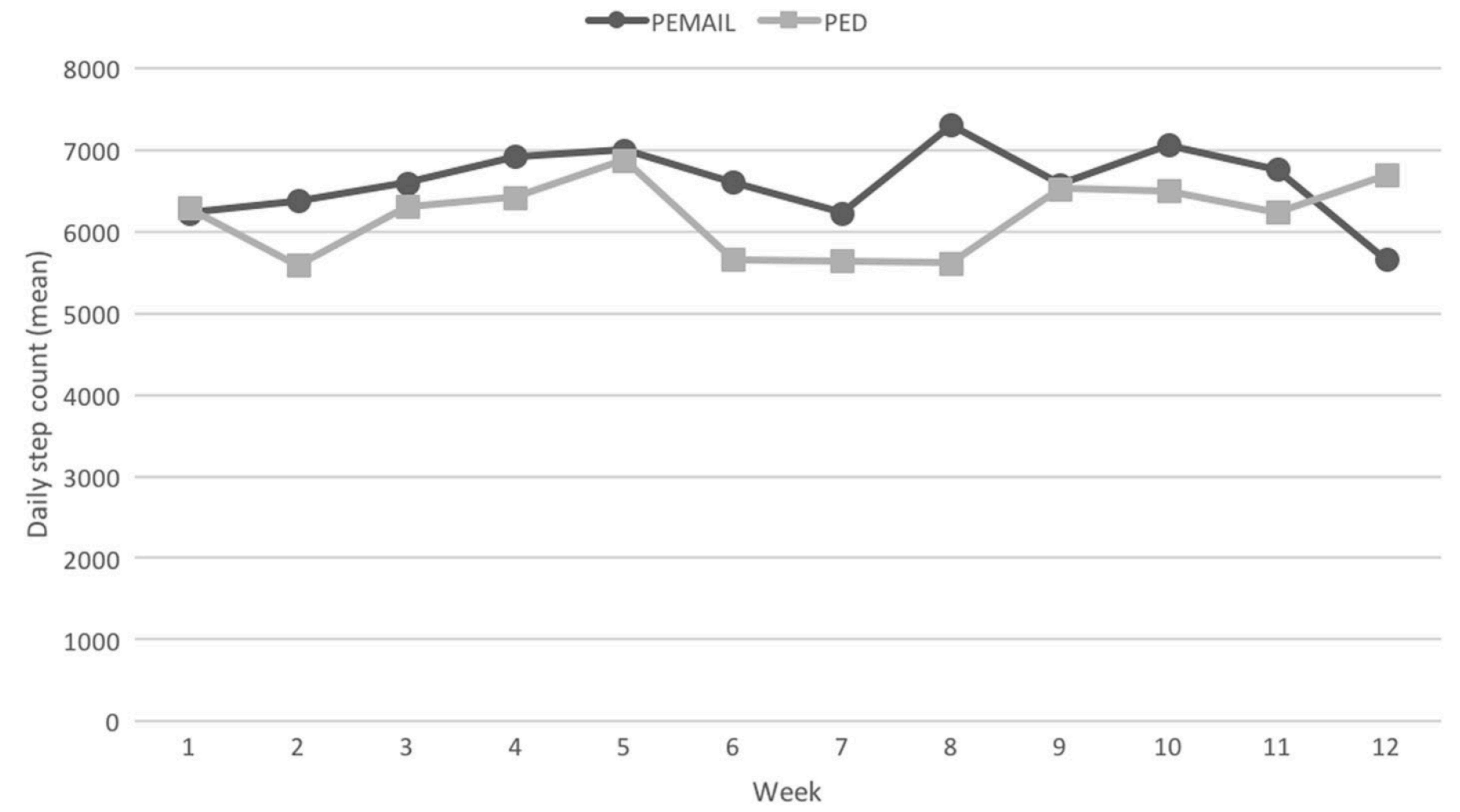


Fig. 4 Mean daily step count during the intervention period. No effect of group or time was observed

Pedometer

2018 study example

Table 2 Baseline (T0) and post-intervention (T12) values of both groups combined, mean (SD)

	T0	T12	Change	*p* value	Cohen's d
Steps per day	5043 (1377)	6719 (2359)	1676 (2066)	.0004	.87
Body mass (kg)	102.8 (21.7)	101.7 (21.6)	−0.7 (1.8)	.044	.05
Systolic blood pressure (mm Hg)	131.5 (14.3)	128.0 (12.4)	−3.5 (9.4)	.045	.26
Diastolic blood pressure (mm Hg)	85.5 (12.9)	83.7 (8.3)	−1.8 (9.7)	.193	.16
Waist circumference (cm)	107.2 (17.7)	105.4 (17.2)	−1.7 (4.0)	.029	.11
Hip circumference (cm)	115.4 (14.5)	114.8 (14.0)	−0.6 (5.0)	.292	.04