



# Dodgy statistics

david owens

---

## Hang-gliding accidents

RAIMUND MARGREITER, LOIS-JÖRG LUGGER

Between 1 January 1973 and 31 December 1976 there were 75 reported accidents in which the pilot was injured or (in 7 cases) killed; two men were injured twice. The number increased steeply during these years, being 1, 3, 27, and 44 in the years 1973 to 1976 successively. The age range was 19-61 (mean 317), and three of the pilots were women. Many were experienced hang-gliders, the mean number of previous flights being 94 (range 7-200). None were known to be abusing drugs or alcohol.

reference

Margreiter R, Lugger LJ.  
Hang-gliding accidents  
BMJ 1978; 1: 400-2

---

## Hang-gliding accidents

RAIMUND MARGREITER, LOIS-JÖRG LUGGER

*Time of accidents* - Most accidents happened during May, June, or September (table I), months when hot weather in the Alps creates more turbulent conditions that would partly explain these peaks. Convection currents **may help to account for the fact that 73 % occurred between 11 am and 3 pm**, though this is also the time when most pilots fly.

reference



Margreiter R, Lugger LJ.  
Hang-gliding accidents  
BMJ 1978; 1: 400-2

---

## Hang-gliding accidents

RAIMUND MARGREITER, LOIS-JÖRG LUGGER

We therefore recommend the following precautions that would lessen the risk:

- (7) **Avoidance of flying between 11 am and 3 pm on hot days, when dangerous convection currents are likely.**

reference



Margreiter R, Lugger LJ.  
Hang-gliding accidents  
BMJ 1978; 1: 400-2

This *floating numerator* error often concerns ‘contact cases’ – thinking that people seen as patients and collected as a sample constitute a population. Another well-known example is the collection of a series of in-patient falls, noticing that a great many of them are among the elderly, and concluding that falls are consequently commoner among elderly patients. This may be a true observation but it can’t be derived from these data.

This error can be one due to lazy thinking and action. The necessary denominator data will cost time and effort (money) to gather: for example, if you want to know when is the most dangerous time to hang-glide you need data on people who don't have accidents.

reference



**Margreiter R, Lugger LJ.**  
Hang-gliding accidents  
BMJ 1978; 1: 400-2



THE LANCET, FEBRUARY 8, 1986

---

## Measurement

---

### STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT

J. MARTIN BLAND

DOUGLAS G. ALTMAN

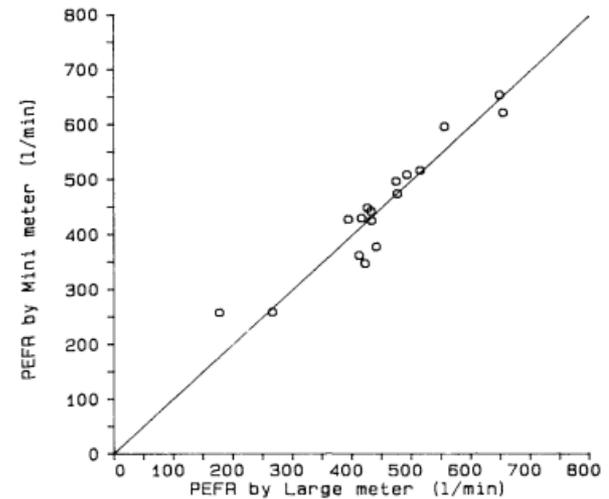


Fig 1—PEFR measured with large Wright peak flow meter and mini Wright peak flow meter, with line of equality.



# correlation and the P-value fallacy

THE LANCET, FEBRUARY 8, 1986

## Measurement

### STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT

J. MARTIN BLAND

DOUGLAS G. ALTMAN

Peak Expiratory Flow Rate

New Mini-meter  
judged against  
Old Large meter

Product-moment correlation represented by 'r' – which can vary from -1 to +1

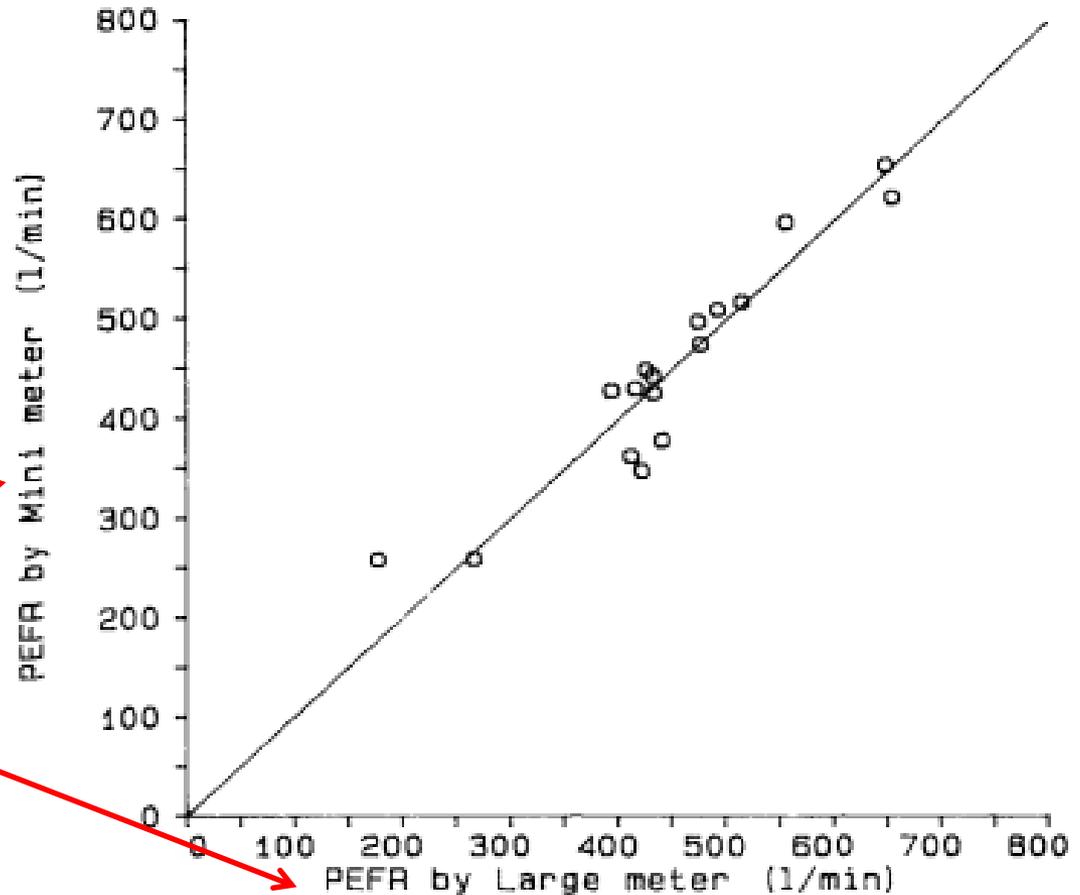


Fig 1—PEFR measured with large Wright peak flow meter and mini Wright peak flow meter, with line of equality.



# correlation and the P-value fallacy

THE LANCET, FEBRUARY 8, 1986

## Measurement

### STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT

J. MARTIN BLAND

DOUGLAS G. ALTMAN

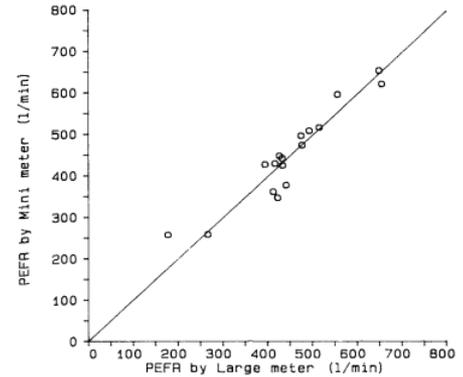


Fig 1—PEFR measured with large Wright peak flow meter and mini Wright peak flow meter, with line of equality.

“r measures the strength of a relation between two variables, not the agreement between them. We will have perfect agreement only if the points in fig 1 lie along the line of equality, but we will have perfect correlation if the points lie along any straight line.”

Imagine reading:  $r = 0.52$ ;  $P = 0.03$

“The test of significance may show that the two methods are related, but it would be amazing if two methods designed to measure the same quantity were not related. The test of significance is irrelevant to the question of agreement.”



# correlation and the P-value fallacy

THE LANCET, FEBRUARY 8, 1986

## Measurement

### STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT

J. MARTIN BLAND

DOUGLAS G. ALTMAN

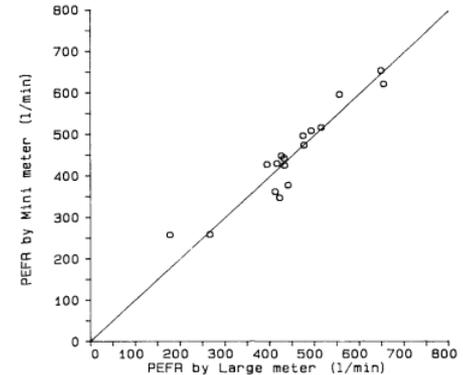


Fig 1—PEFR measured with large Wright peak flow meter and mini Wright peak flow meter, with line of equality.

“r measures the strength of a relation between two variables, not the agreement between them. We will have perfect agreement only if the points in fig 1 lie along the line of equality, but we will have perfect correlation if the points lie along any straight line.”

Imagine reading:  $r = 0.52$ ;  $P = 0.03$

“The test of significance may show that the two methods are related, **but it would be amazing** if two methods designed to measure the same quantity were not related. The test of significance is irrelevant to the question of agreement.”

“When the Okies left Oklahoma [think dust-bowl, Grapes of Wrath ...] and moved to California, they raised the average intelligence levels in both States.”

Imagine bowel cancer diagnosis:

If a population of patients is more accurately staged it will improve the survival of all stages because patients with subtle advanced disease will be up-staged. For example, if liver imaging is routinely included in the staging workup, patients with subclinical liver metastases will be identified, removed from the stage III group and added to the stage IV group thus improving the expected survival of both groups.

Spurious improvements can be deemed due to new kit or treatments, some places can seem to have better or worse outcomes than others...



For an exercise in the metric system, a certain teacher asked her fourth-grade students to calculate their respective heights in centimeters. As they left the classroom for recess, she had the students write their names on the blackboard alongside their heights in two lists as follows:

### Okies

Wailyn 144  
Nollyn 135  
Pailwyn 137  
Courtlyn 124  
Soulyn 140  
Gladwyn 128  
Ashlyn 122  
Ravyn 139  
Ulewyn 142  
Qualyn 138  
Zalyn 147  
Jenwyn 131  
Xelwyn 145

### Calies

Marlyn 134  
Irwyn 146  
Haldwyn 129  
Edwyn 126  
Yawyn 130  
Kamlyn 132  
Louryn 133  
Fairlyn 127  
Oldewyn 136  
Dallwyn 125  
Terryn 141  
Vallewyn 143  
Brandyn 123

### Yaw of Averages

by Paul Niquette

Copyright © 2005 by Paul Niquette. All rights reserved.

"Which group is taller?" asked the teacher, when the class returned.

The students immediately set about to calculate the average heights of the two groups. They determined that the Okies were taller than the Calies by 3.6 centimeters (136.3 vs 132.7).

The teacher smiled and took chalk in hand. The students watched as she crossed out a name from one of the lists and added it to the other list.

"Do the calculations again," she said, clapping the chalk dust from her hands. "You will see that the average heights of both the Okies and the Calies have *increased!*"

Which name did the teacher select?





be very very careful of the odds ratio

Cohort analytic study

		Developed irritable bowel syndrome		
		Yes	No	
Gastro-enteritis episode	Yes	12 <i>a</i>	306 <i>b</i>	318 <i>a+b</i>
	No	2027 <i>c</i>	582,281 <i>d</i>	584,308 <i>a+c</i>
		<i>a+c</i>	<i>b+d</i>	

Risk ratio =  $12 / 318 \div 2027 / 584308 = 10.9$

Odds ratio =  $12 / 306 \div 2027 / 582281 = 11.3$



be very very careful of the odds ratio

Randomised controlled trial

		Stopped breastfeeding by 3 months		
		Yes	No	
Lactation consultant	No	52 <i>a</i>	5 <i>b</i>	57 <i>a+b</i>
	Yes	32 <i>c</i>	19 <i>d</i>	51 <i>a+c</i>
		<i>a+c</i>	<i>b+d</i>	

Risk ratio =  $52 / 57 \div 32 / 51 = 1.5$

Odds ratio =  $52 / 5 \div 32 / 19 = 6.2$



be very very careful of the odds ratio

A review of reviews

		Passive smoking found harmful		
		Yes	No	
Article's authors sponsored by tobacco industry	No	29	2	31
	Yes	10	65	75
		a+c	b+d	

Risk ratio =  $29 / 31 \div 10 / 75 = 7.0$

Odds ratio =  $29 / 2 \div 10 / 65 = 94$



be very very careful of the odds ratio

A review of reviews

		Passive smoking found harmful		
		Yes	No	
Article's authors sponsored by tobacco industry	No	29 <i>a</i>	2 <i>b</i>	31 <i>a+b</i>
	Yes	10 <i>c</i>	65 <i>d</i>	75 <i>a+c</i>
		<i>a+c</i>	<i>b+d</i>	

Risk ratio =  $29 / 31 \div 10 / 75 = 7.0$

Odds ratio =  $29 / 2 \div 10 / 65 = 94$  – this value becoming 88 after logistic regression that took account of article quality, peer review status, article topic and year of publication

be very very careful of the odds ratio



UNIVERSITY OF LEEDS

thebmj

Research ▾

Education ▾

News & Views ▾

Campaigns

Archive

## News

### Links to tobacco industry influences review conclusions

*BMJ* 1998 ; 316 doi: <http://dx.doi.org/10.1136/bmj.316.7144.1553b> (Published 23 May 1998)

Cite this as: *BMJ* 1998;316:1553

Article

Related content

Metrics

Responses

Jacqui Wise

Author affiliations ▾

A review article written by authors with affiliations to the tobacco industry is 88 times more likely to conclude that passive smoking is not harmful than if the review article was written by authors with no connection to the tobacco industry.

Deborah Barnes and Lisa Bero from the University of California searched Medline and Embase and a database of symposium proceedings on passive smoking and identified 106 reviews of the health effects of passive smoking published from 1980 to 1995 (*JAMA* 1998; 279:1566-70). They found that very few reviews had been conducted systematically. Three quarters of the articles failed to disclose the sources of funding for the research.

be very very careful of the odds ratio



UNIVERSITY OF LEEDS

A review article written by authors with affiliations to the tobacco industry is 88 times more likely to conclude that passive smoking is not harmful than if the review article was written by authors with no connection to the tobacco industry.

No! The conclusion (correct in the original paper) is that it is 7 times **as** likely