

---

# Υπολογισμός μεγέθους δείγματος

---

Απόστολος Βανταράκης  
Καθηγητής Υγιεινής  
avanta@upatras.gr

# Πόσο μεγάλο πρέπει να είναι το δείγμα?

---

Μια επιδημία Σαλμονέλλας συνέβη σε εθνικό επίπεδο με πολλά περιστατικά

Σχεδιάζεις μια case-control μελέτη για να ταυτοποιήσεις αν η κατανάλωση του τροφίμου X συνδέεται με τη μόλυνση

Πόσα περιστατικά και controls θα χρησιμοποιήσεις?

# Πόση ισχύ έχεις?

---

Μια επιδημία 14 περιστατικών HAV έχει συμβεί σε ένα σχολείο X

Υποψιάζεσαι έκθεση από τρόφιμο Y που σχετίζεται με την ασθένεια και σχεδιάζεις να κάνεις μελέτη κοορτών

Πόση ισχύ έχεις για να ανιχνεύσεις ένα σχετικό κίνδυνο  $(RR) = 1.5$  ;

# Αντικείμενο μελέτης: μέγεθος δείγματος

---

- Γιατί υπολογίζουμε το μέγεθος δείγματος
- Αρχές υπολογισμού μεγέθους δείγματος
- Βήματα για να υπολογιστεί το μέγεθος δείγματος

# Γιατί να υπολογίσουμε το μέγεθος δείγματος?

---

- Πόσοι άνθρωποι θα απαντήσουν τις ερωτήσεις σου;;
- Χωρίς νόημα αν η μελέτη είναι πολύ μικρή;
- Σπατάλη πόρων αν είναι πολύ μεγάλο.

# Γιατί να υπολογίσουμε το μέγεθος δείγματος?

---

- ❑ Υπολογίζουμε τον αριθμό που χρειάζεται για να μετρήσουμε τον παράγοντα υπό μελέτη
- ❑ Ισορροπία μεταξύ μεγέθους και πόρων...
- ❑ Το μέγεθος δείγματος καθορίζεται από διάφορους παράγοντες :
  - ✓ Επίπεδο σημαντικότητας (“alpha”)
  - ✓ Ισχύ (“1-beta”)
  - ✓ Αναμενόμενος επιπολασμός του παράγοντα που μελετάμε

# Έλεγχος σημαντικότητας: Μηδενική και εναλλακτική υπόθεση

---

## Μηδενική υπόθεση ( $H_0$ )

Δεν υπάρχει διαφορά. Οποιαδήποτε διαφορά είναι λόγω τύχης

## Εναλλακτική υπόθεση ( $H_1$ )

Υπάρχει αληθινή διαφορά

# Παραδείγματα μηδενικών υποθέσεων

---

## Case-control μελέτη

$$H_0: OR=1$$

“οι πιθανότητες έκθεσης στα περιστατικά είναι οι ίδιες με αυτές στους ελέγχους”

## Μελέτη κοορτών

$$H_0: RR=1$$

“Ο δείκτης προσβολής ανάμεσα στους εκτεθειμένους είναι η ίδια με το δείκτη προσβολής ανάμεσα στους μη εκτεθειμένους”



# Σημαντικότητα και ισχύ

		Αλήθεια	
		$H_0$ αλήθεια	$H_0$ ψέμα Διαφορά
Απόφαση	Δεν μπορεί να απορριφθεί $H_0$	Καμμία διαφορά Σωστή απόφαση	Type II λάθος = $\beta$
	Απόρριψη $H_0$	Type I λάθος = $\alpha$ σημαντικότητα	Σωστή απόφαση ισχύ = $1-\beta$

# Επίπεδο σημαντικότητας (p-value)

---

$\alpha$  or type I error: Πιθανότητα να βρεθεί διαφορά (RR $\neq$ 1, απόρριψη  $H_0$ ), όταν δεν υπάρχει διαφορά

p-value χρησιμοποιείται για να απορριφθεί  $H_0$  (επίπεδο σημαντικότητας);

Μια υπόθεση είναι πάντα «αποδεκτή» συνήθως στο 5%

# Type II λάθος και ισχύ

---

$\beta$  είναι το λάθος τύπου II

Πιθανότητα να μην βρεθεί διαφορά όταν η διαφορά υπάρχει

**Ισχύ** είναι  $(1-\beta)$  και συνήθως τίθεται

Πιθανότητα εύρεσης διαφοράς που πραγματικά υπάρχει (=ευαισθησία)

# Ισχύ

---

Οι ευκολότεροι τρόποι για να αυξήσουμε την ισχύ είναι να:

- ❑ Αυξήσουμε το μέγεθος δείγματος
- ❑ Αυξήσουμε την επιθυμητή διαφορά (ή το μέγεθος της επίπτωσης) απαιτείται
  - ✓ NB! αυξάνοντας την επιθυμητή διαφορά σε RR/OR εννοεί την μετακίνηση από το 1!
- ❑ Αυξήσουμε το επίπεδο σημαντικότητας που επιθυμείται ( $\alpha$  error) σε π.χ. 10%
  - → Επέκταση των επιπέδων εμπιστοσύνης

# Η επίδραση του μεγέθους του δείγματος

---

Θεωρούμε 3 μελέτες κοορτών που ελέγχουν την έκθεση στα οστρακοειδή με  $N=10, 100, 1000$

Σε όλες τις 3 μελέτες, το 60% των εκτεθειμένων ασθενεί σε σύγκριση με το 40% που δεν εκτίθεται ( $RR = 1.5$ )

# Πίνακας Α

---

		Ασθενεί		
		Ναι	Σύνολο	AR
Τρώνε οστρακοειδή	Ναι	3	5	3/5
	Όχι	2	5	2/5
	Σύνολο	5	10	5/10

**RR=1.5, 95% CI: 0.4-5.4, p=0.53**

# Πίνακας Β

---

		Ασθενεί		AR
		Ναι	Σύνολο	
Τρώνε οστρακοειδή	Ναι	30	50	30/50
	Όχι	20	50	20/50
	Σύνολο	50	100	50/100

**RR=1.5, 95% CI: 1.0-2.3, p=0.046**

# Πίνακας C

---

		Ασθενεί		AR
		Ναι	Όχι	
Τρώνε οστρακοειδή	Ναι	300	500	300/500
	Όχι	200	500	200/500
	Σύνολο	500	1000	500/1000

**RR=1.5, 95% CI: 1.3-1.7,  $p<0.001$**



# Μέγεθος δείγματος και ισχύ

---

Στο Πίνακα Α, με  $n=10$  δείγματα, δεν υπάρχει σημαντική σχέση με τα οστρακοειδή, αλλά υπήρχε σχέση με το μέγεθος δείγματος.

Αυτό δείχνει τη σχέση μεταξύ ισχύος και μέγεθος δείγματος

# Βήματα στον υπολογισμό του δείγματος για την περιγραφική μελέτη

---

- ❑ Ταυτοποίηση του κύριου είδους της μελέτης
- ❑ Καθορισμός τύπου υπολογισμού  
(%, μέσος, αναλογία,...)
- ❑ Δείχνουμε την αναμενόμενη συχνότητα του παράγοντα ενδιαφέροντος
  - ✓ Βιβλιογραφία, γνώμη ειδικού
- ❑ Αποφασίζουμε την επιθυμητή ακρίβεια του υπολογισμού

# Μέγεθος δείγματος κοορτής: θέματα

---

- ❑ Αναλογία κινδύνου
- ❑ Αναμενόμενη συχνότητα της ασθένειας στον μη εκτεθειμένο πληθυσμό
- ❑ Αναλογία του μη εκτεθειμένου προς τον εκτεθειμένο
- ❑ Επιθυμητό επίπεδο σημαντικότητας ( $\alpha$ )
- ❑ Ισχύς μελέτης ( $1-\beta$ )
  - ✓ Να ανιχνεύσουμε σε στατιστικά σημαντικό επίπεδο μια χαρακτηριστική αναλογία κινδύνου

## Episheet © 2002, 2004

### Spreadsheets for the Analysis of Epidemiologic Data.

Written by Ken Rothman. These spreadsheets may be distributed freely. No warranty is implied about the accuracy of the results. Version of November 10, 2004. Send comments to Ken@Rothman.name

**Program not working? Check your Excel security settings to be certain that the macros load (Tools/Macros).**

#### Major References:

1. Edwards JH: The recognition and estimation of cyclic trends. *Ann Hum Gen* 1961;25:83-87.
2. Rothman KJ: *Epidemiology. An Introduction.* Oxford University Press, New York, 2002.
3. Rothman KJ: Estimation of confidence limits for the cumulative probability of survival in life-table analysis. *J Chron Dis* 1978;31:557-560.
4. Rothman KJ and JD Boice Jr: *Epidemiologic Analysis with a Programmable Calculator.* U.S. Department of Health, Education and Welfare, Public Health Service, National Institutes of Health, NIH Publication No. 79-1649, 1979.
5. Rothman KJ, Greenland S: *Modern Epidemiology*, 2nd Ed., Lippincott, Williams & Wilkins, Philadelphia, 1998

[Rate Data](#)

[Case-control Power](#)

[Matched Case-control data](#)

[Risk Data](#)

[Study Size](#)

[O. R. Exact](#)

[Case-control Data](#)

[Quickcalc](#)

[P-Value Functions](#)

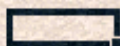
[Standardization](#)

[Seasonal Analysis](#)

[Meta-Analysis \(contributed by T. Andersson & A. Ahlbom\)](#)

[Cohort Power](#)

[Life Table](#)



Menu / Rate Data / Risk Data / Case-control Data / Standardization / Cohort Power

Draw | AutoShapes | [Drawing Tools: Line, Arrow, Rectangle, Circle, Text, Eraser, Undo, Redo, Copy, Paste, Fill, Stroke, Selection] | [Navigation: Home, Back, Forward, Stop]

# Κοορτή:

## Ερι-sheet υπολογισμός ισχύος

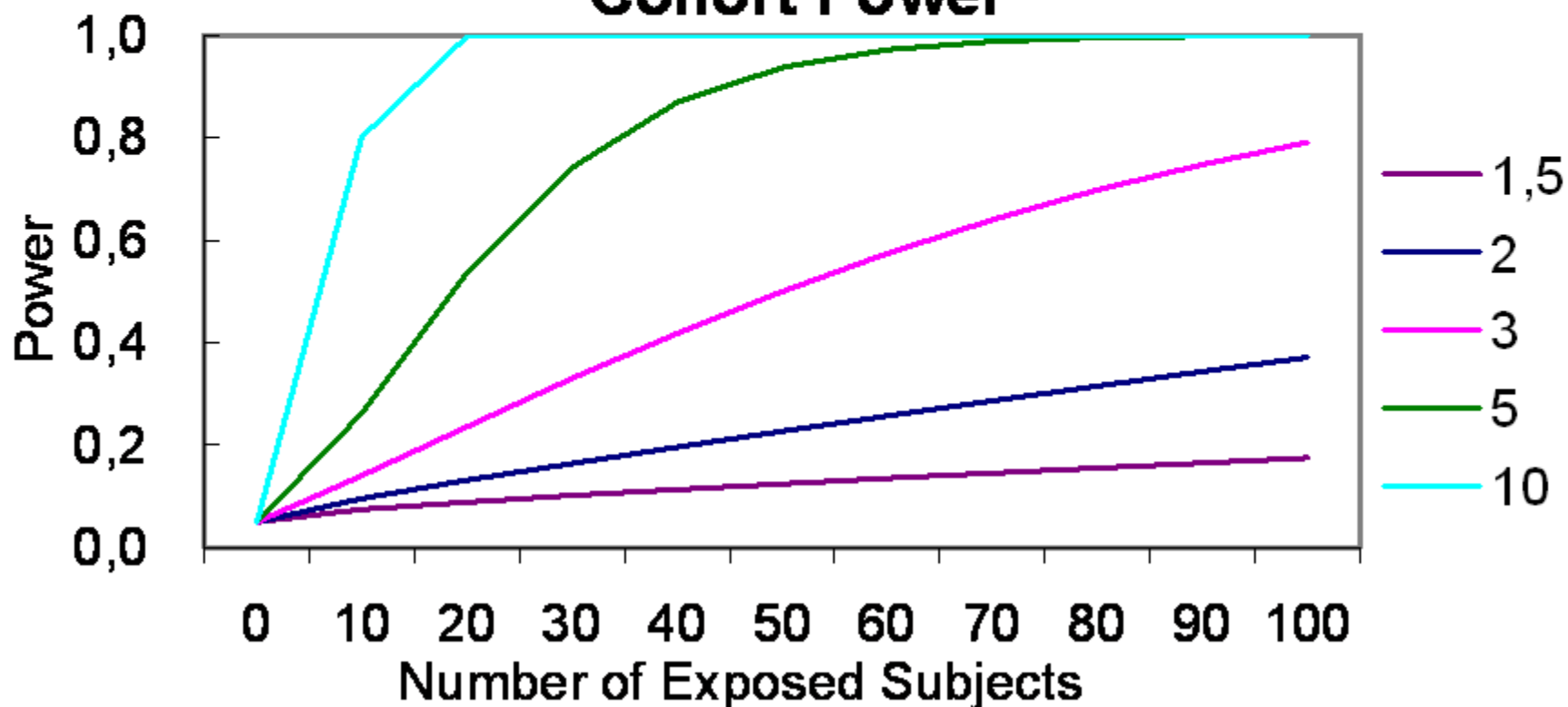
---

Κίνδυνος α error	5%
Πληθυσμός που εκτίθεται	100
Αναμ.συχνότητα ασθένειας σε μη εκτεθ	5%
Αναλογία μη εκτεθειμένου/εκτεθειμένου	1:1
RR να ανιχνευθεί	$\geq 1.5$

Z-alpha:	1,96
Unexp:Exp Ratio:	1
Risk in Unexposed:	0,05
Max. No. Exposed:	100

RR Effect Levels:		4
1,5	5	
2	7	
3	10	

### Cohort Power



# Case-control & μέγεθος δείγματος: θέματα

---

- ❑ Αριθμός περιστατικών
- ❑ Αριθμός ελέγχων/περιστατικό
- ❑ Λόγος αναλογιών αξίζει ανίχνευση
- ❑ % των εκτεθειμένων ατόμων στον πληθυσμό στόχο
- ❑ Επιθυμητό επίπεδο σημαντικότητας ( $\alpha$ )
- ❑ Ισχύς μελέτης ( $1-\beta$ )
  - ✓ Να ανιχνεύσουμε σε ένα στατιστικά σημαντικό επίπεδο a particular odds ratio

# Case-control:

## Υπολογισμός ισχύος

---

- ❑ Κίνδυνος  $\alpha$  error 5%
- ❑ Αριθμός περιστατικών 200
- ❑ Αναλογία controls που εκτέθηκαν 5%
- ❑ OR να βρεθεί  $\geq 1.5$
- ❑ Αριθμός controls:case 1:1



Z-alpha: 1,96

Control/Case Ratio: 1

Exposure Prevalence: 0,05

Maximum No. Cases: 200

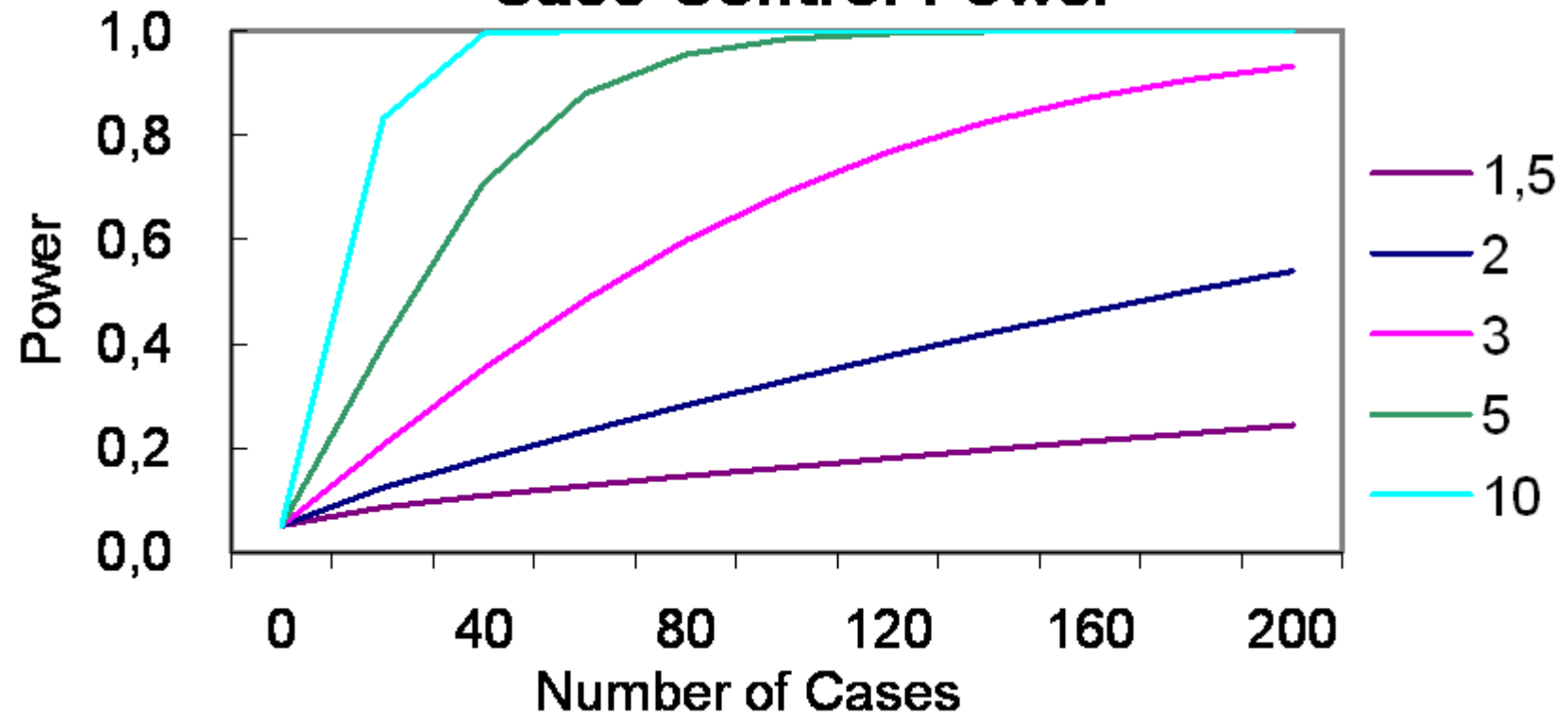
OR Effect Levels: 4

1,5 5

2 7

3 10

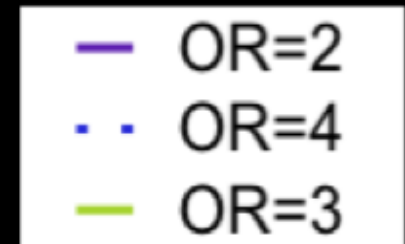
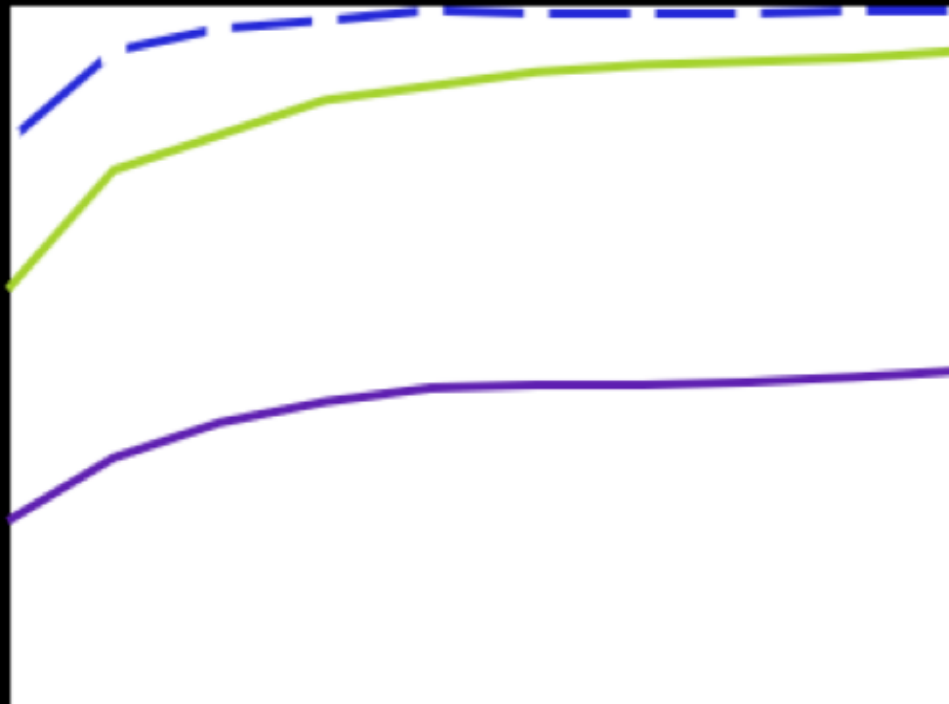
### Case-Control Power



# Στατιστική ισχύ μιας Case-Control μελέτης

για διαφορετικές control-to-case αναλογίες και λόγο πιθανοτήτων odds ratios (50  
περιστατικά)

---



# Συμπεράσματα

---

- ❑ Μην ξεχνάτε να υπολογίζετε το μέγεθος δείγματος /ισχύ
- ❑ Χρησιμοποιούμε όλες τις πηγές της διαθέσιμης πληροφορίας για να ενημερώσουμε τους υπολογισμούς
- ❑ Προσπαθούμε πολλά σενάρια
- ❑ Να υπολογίζουμε για τις μη απαντήσεις

# Συμπεράσματα

---

If in doubt...

**Call a statistician !!!!**