

ΙΑΤΡΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Προοπτική έρευνα

	Νόσημα		
Παράγοντας	Ναι	Όχι	Σύνολο
Ναι	a	b	a+b
Όχι	c	d	c+d
Σύνολο	a+c	b+d	n

$I_E = a/(a+b)$ είναι η συχνότητα εμφάνισης του νοσήματος στους εκτεθειμένους

$I_{NE} = c/(c+d)$ η συχνότητα εμφάνισης του νοσήματος στους μη-εκτεθειμένους

Σχετικός Κίνδυνος (Relative Risk) $RR = I_E / I_{NE} = \frac{a}{a+b} \div \frac{c}{c+d} = \frac{a}{c} \times \frac{(c+d)}{(a+b)}$

Εκφράζει πόσο συχνότερη (ή λιγότερο συχνή) είναι η νόσος όταν ο παράγοντας κινδύνου είναι παρών σε σχέση με την αντίστοιχη συχνότητα όταν ο παράγοντας είναι απών

Έρευνες ασθενών - μαρτύρων

- Στις έρευνες ασθενών-μαρτύρων επιλέγονται τα $a+c$ και $b+d$ (από τον τετράπτυχο Πίνακα). Επομένως δεν είναι δυνατόν να υπολογιστεί ο Σχετικός Κίνδυνος και υπολογίζεται μόνο ο Σχετικός Λόγος (ΣΛ).

	Νόσημα		
Παράγοντας	Ναι	Όχι	Σύνολο
Ναι	a	b	a+b
Όχι	c	d	c+d
Σύνολο	a+c	b+d	n

Τι εκφράζουν ο ΣΛ και ο RR??

- Η σημασία των RR και ΣΛ είναι η ίδια: εκφράζουν πόσο μεγαλύτερος είναι ο κίνδυνος να πάθουν τη νόσο οι εκτεθειμένοι στον παράγοντα σε σχέση με τους μη εκτεθειμένους.
- Πιθανές τιμές: 0 μέχρι ∞
- Τιμή 1 σημαίνει απουσία σχέσης
- Τιμή > 1 δείχνει επιβαρυντικό παράγοντα
- Τιμή < 1 δείχνει προστατευτικό παράγοντα

Παράδειγμα προοπτικής έρευνας

Αποτέλεσμα προηγούμενης εγκυμοσύνης	Γυναίκες χωρίς αντισύλληψη για 10 μήνες		Σύνολο
	Εγκυμοσύνη όχι	Εγκυμοσύνη ναι	
Προκλητή έκτρωση	111 (a)	33(b)	144
Άλλο	175 (c)	177 (d)	352
Σύνολο	286	210	496

$$RR = (111/144) \div (175/352) = 1,55$$

Ποιες ομάδες συγκρίνονται και ως προς τι;

Γυναίκες των οποίων η προηγούμενη εγκυμοσύνη κατέληξε σε προκλητή έκτρωση είχαν κατά 55% αυξημένη πιθανότητα να μην μείνουν έγκυες σε διάστημα 10 μηνών σε σχέση με εκείνες των οποίων η προηγούμενη εγκυμοσύνη είχε άλλο αποτέλεσμα

Παράδειγμα έρευνας ασθενών-μαρτύρων

	Καρκίνος ωοθηκών		
Αριθμός παιδιών	Ναι	Όχι	Σύνολο
0	48	30	78
1+	152	150	302
Σύνολο	200	180	380

Σχετικός Λόγος (ΣΛ ή Odds Ratio-OR)= $(48 \times 150) / (30 \times 152) = 1,58$

Συγκρίνει εκείνες που δεν έχουν παιδιά με εκείνες που έχουν ως προς την πιθανότητα να έχουν καρκίνο ωοθηκών

Γυναίκες που δεν είχαν παιδιά είχαν κατά 58% αυξημένο κίνδυνο καρκίνου των ωοθηκών σε σύγκριση με εκείνες που είχαν παιδιά

Παράδειγμα έρευνας ασθενών-μαρτύρων - αναδιάταξη

	Καρκίνος ωοθηκών		
Αριθμός παιδιών	Ναι	Όχι	Σύνολο
1+	152	150	302
0	48	30	78
Σύνολο	200	180	380

Θα μπορούσαμε να υπολογίσουμε αντίστροφα

Σχετικός Λόγος (ΣΛ ή OR)= $(30 \times 152) / (48 \times 150) = 0,63 = 1/1,58$

Συγκρίνει εκείνες που έχουν παιδιά με εκείνες που δεν έχουν ως προς την πιθανότητα να έχουν καρκίνο ωοθηκών

Γυναίκες που είχαν παιδιά είχαν κατά 37% μειωμένο κίνδυνο καρκίνου των ωοθηκών σε σύγκριση με εκείνες που δεν είχαν παιδιά

Όρια αξιοπιστίας (διάστημα εμπιστοσύνης) του σχετικού κινδύνου

- Η κατανομή συχνοτήτων του σχετικού κινδύνου → ασύμμετρη όρια αξιοπιστίας από λογαριθμισμένες τιμές

- RR: προοπτική έρευνα

- 95% O.A. $\ln RR \pm Z_{0,05} * SE(\ln RR)$



οριακή τιμή κανονικής κατανομής για $\alpha=0.05$ (1,96)

- $SE(\ln RR) = \sqrt{\frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}}$

Παράδειγμα προοπτικής έρευνας (συνέχεια)

Αποτέλεσμα προηγούμενης εγκυμοσύνης	Γυναίκες χωρίς αντισύλληψη για 10 μήνες		Σύνολο
	Εγκυμοσύνη όχι	Εγκυμοσύνη ναι	
Προκλητή έκτρωση	111 (a)	33(b)	144
Άλλο	175 (c)	177 (d)	352
Σύνολο	286	210	496

$$RR = (111/144)/(175/352) = 1,55$$

$$\ln(RR) = \ln(1,55) = 0,438$$

$$SE(\ln(RR)) = \sqrt{\frac{1}{111} - \frac{1}{111+33} + \frac{1}{175} - \frac{1}{175+177}} = 0,0707$$

$$95\% \Delta E (\ln RR) = 0,4383 \pm 1,96 * 0,0707 =$$

$\nearrow 0,5768$
 $\searrow 0,2997$

Αντιλογαριθμίζοντας: $e^{0,5768} = 1,78$, $e^{0,2997} = 1,35$,

Το διάστημα εμπιστοσύνης του RR είναι 1,35 έως 1,78

Όρια αξιοπιστίας (διάστημα εμπιστοσύνης) του σχετικού λόγου (για έρευνες ασθενών – μαρτύρων)

- 95% O.A. $\ln\Sigma\Lambda \pm Z_{0,05} * SE(\ln\Sigma\Lambda)$

- $SE(\ln\Sigma\Lambda) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$

όταν a, b, c, d όχι πολύ μικρά

Παράδειγμα έρευνας ασθενών-μαρτύρων

	Καρκίνος ωοθηκών		
Αριθμός παιδιών	Ναι	Όχι	Σύνολο
0	48	30	78
1+	152	150	302
Σύνολο	200	180	380

Σχετικός Λόγος (ΣΛ ή OR) = $(48 \times 150) / (30 \times 152) = 1,58$

$$SE(\ln(OR)) = \sqrt{\frac{1}{48} + \frac{1}{30} + \frac{1}{152} + \frac{1}{150}} = 0,2596$$

$$0,4574 \pm 1,96 * 0,2596 = \begin{cases} 0,9662 \\ -0,0514 \end{cases}$$

$$e^{0,9662} = 2,63$$

$$e^{-0,0514} = 0,95$$

Το διάστημα εμπιστοσύνης του OR είναι 0,95 - 2,63

Κατανομή 118 ασθενών με στεφανιαία νόσο και 92 «υγιών» κατά οικογενειακή κατάσταση

	Έγγαμοι	Άγαμοι	Σύνολο
Ασθενείς	87	31	118
Υγιείς	40	52	92
Σύνολο	127	83	210

$$\chi^2 = (|87*52 - 31*40| - 210/2)^2 * 210 / ((87+40)*(31+52)*(87+31)*(40+52)) = 18.6$$

$$BE=1$$

$$P < 0,001$$

Είναι πιθανό αυτή η σχέση να οφείλεται σε τρίτο παράγοντα?

Σε ποιον πχ?

Διαστρωμάτωση κατά ηλικία

	Μέχρι 29 ετών		
	Έγγαμοι	Άγαμοι	Σύνολο
Ασθενείς	3	5	8
Υγιείς	8	34	42
Σύνολο	11	39	50

$$\chi^2 = 0,47$$

	50+ ετών		
	Έγγαμοι	Άγαμοι	Σύνολο
Ασθενείς	66	14	80
Υγιείς	14	3	17
Σύνολο	80	17	97

$$\chi^2 = 0,11$$

	30-49 ετών		
	Έγγαμοι	Άγαμοι	Σύνολο
Ασθενείς	18	12	30
Υγιείς	18	15	33
Σύνολο	36	27	63

$$\chi^2 = 0,19$$

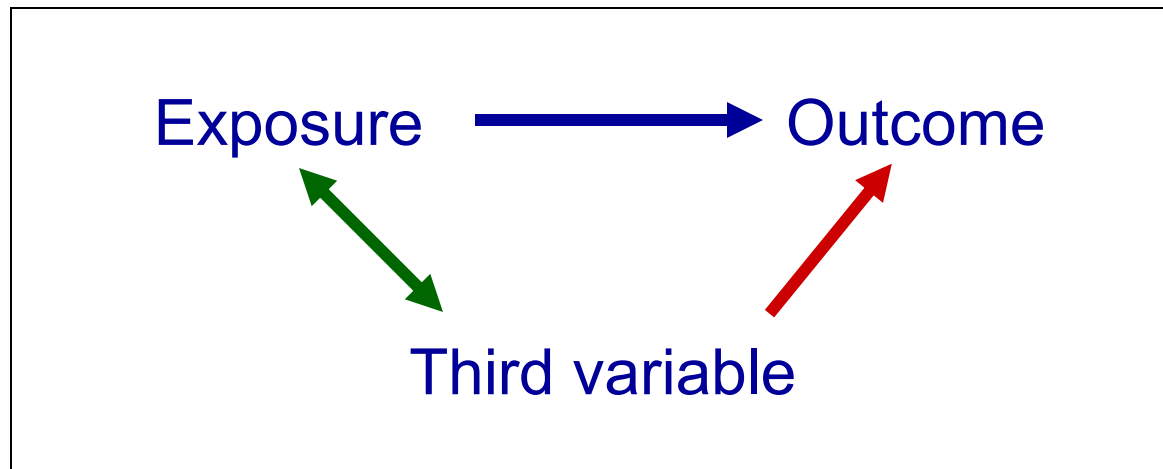
Τι συμπέρασμα μπορεί να βγει;
Τι πρόβλημα υπάρχει;

Συγχυτικοί παράγοντες (Confounding variables): Ορισμός

Συγχυτικός παράγοντας (ΣΠ):

- 1) συσχετιζόμενος με τον εξεταζόμενο παράγοντα κινδύνου (μικρότερη ή μεγαλύτερη συχνότητα του ΣΠ σε εκτεθέντες απ' ότι σε μη εκτεθέντες)
- 2) Ανεξάρτητος προγνωστικός παράγοντας για την έκβαση (δηλαδή να μεταβάλλει την πιθανότητα νόσησης/θανάτου κλπ)

ΣΥΓΧΥΤΙΚΟΙ ΠΑΡΑΓΟΝΤΕΣ: Προϋποθέσεις (II)



Συσχέτιση με τον παράγοντα κινδύνου

Προγνωστικός παράγοντας για την νόσο - ανεξαρτήτως του παράγοντα κινδύνου

ΔΙΑΣΤΡΩΜΑΤΩΣΗ

- Έστω μελετούμενος παράγοντας κινδύνου E με 2 κατηγορίες ($E+$, $E-$)
- Έστω αποτέλεσμα (έκβαση) με 2 κατηγορίες ($N+$, $N-$)
- Έστω πιθανός συγχυτικός παράγοντας με 2 κατηγορίες (Σ_1 , Σ_2)
- Ο τετράπτυχος πίνακας για το σύνολο των δεδομένων κατασκευάζεται χωριστά για τα άτομα που ανήκουν στην κατηγορία Σ_1 και χωριστά για τα άτομα που ανήκουν στην κατηγορία Σ_2 .
- Η διαδικασία αυτή του διαχωρισμού ενός Πίνακα σε περισσότερους, σύμφωνα με τις κατηγορίες μίας μεταβλητής (εδώ του πιθανού συγχυτικού παράγοντα) λέγεται **διαστρωμάτωση**.

ΔΙΑΣΤΡΩΜΑΤΩΣΗ

Μόνον άτομα της κατηγορίας Σ_1

Έκθεση στον παράγοντα

Νόσημα	Ναι	Όχι	Σύνολο
Ναι	a_1	b_1	$a_1 + b_1$
Όχι	c_1	d_1	$c_1 + d_1$
Σύνολο	$a_1 + c_1$	$b_1 + d_1$	$a_1 + c_1 + b_1 + d_1$

Μόνον άτομα της κατηγορίας Σ_2

Έκθεση στον παράγοντα

Νόσημα	Ναι	Όχι	Σύνολο
Ναι	a_2	b_2	$a_2 + b_2$
Όχι	c_2	d_2	$c_2 + d_2$
Σύνολο	$a_2 + c_2$	$b_2 + d_2$	$a_2 + c_2 + b_2 + d_2$

- Υπολογίζουμε Σχετικό Κίνδυνο (ΣΚ) σε καθέναν από τους πίνακες
- «Ελέγχουμε» εάν οι ΣΚ είναι περίπου ίδιοι σε όλες τις κατηγορίες του πιθανού συγχυτικού παράγοντα – Δεν θα πρέπει να διακρίνουμε αλλαγή της κατευθυνσης (π.χ. $\Sigma K > 1$ στην Σ_1 , και $\Sigma K < 1$ στην Σ_2)

Ερώτημα: Παραμένει η σχέση έκθεσης-αποτελέσματος περίπου ίδια σε κάθε Πίνακα της διαστρωμάτωσης;

- Αν δεν είναι ίδια, είναι σημαντικό να περιγραφεί πώς διαφοροποιείται στα διάφορα επίπεδα του παράγοντα (τρίτη μεταβλητή) που χρησιμοποιήθηκε ως κριτήριο της διαστρωμάτωσης (δίνουμε το ΣΚ για κάθε επίπεδο της τρίτης μεταβλητής)
- Αν είναι περίπου ίδια, τότε μπορεί να υπολογιστεί ο συνοπτικός σχετικός κίνδυνος (Mantel- Haenszel) που προκύπτει από όλους τους πίνακες λαμβάνοντας υπόψη τη διαστρωμάτωση.
 - Επίσης μπορούμε να διερευνήσουμε το αν υπάρχει σχέση του αρχικού παράγοντα – αποτελέσματος (αν δηλαδή ο συνοπτικός σχετικός κίνδυνος είναι 1 ή όχι) μέσω του συνοπτικού χ^2 . Το συνοπτικό χ^2 ελέγχει αν ο συνοπτικός ΣΚ διαφέρει σε βαθμό στατιστικά σημαντικό από τη μονάδα, δηλαδή, αν η σχέση μεταξύ των 2 παραγόντων είναι στατιστικά σημαντική **ανεξάρτητα** από την επίδραση του συγχυτικού παράγοντα.

Γενική εξίσωση του συνοπτικού σχετικού κινδύνου και του συνοπτικού χ^2

Επίπεδο (στρώμα) i του συγχυτικού παράγοντα			
	Εκτεθειμένοι	Μη-εκτεθειμένοι	Σύνολο
Ασθενείς	a_i	b_i	n_{1i}
«Υγιείς»	c_i	d_i	n_{0i}
Σύνολο	m_{1i}	m_{0i}	T_i

Όπου $i=1, \dots, I$ και I = αριθμός στρωμάτων στη διαστρωμάτωση (ή κατηγοριών του συγχυτικού παράγοντα)

$$X^2_{M-H} = \frac{\left\{ \sum_{i=1}^I a_i - \sum_{i=1}^I \frac{n_{1i} m_{1i}}{T_i} \right\}^2 - \frac{1}{2}}{\sum_{i=1}^I \frac{n_{1i} n_{0i} m_{1i} m_{0i}}{T_i^2 (T_i - 1)}} \quad \Sigma \Lambda_{M-H} = \frac{\sum_{i=1}^I (a_i d_i / T_i)}{\sum_{i=1}^I (b_i c_i / T_i)}$$

Αξιολογούμε το χ^2_{M-H} στους πίνακες της χ^2 κατανομής με 1 ΒΕ

Κατανομή 2239 γυναικών καρκινοπαθών του μαστού και 1370 μη καρκινοπαθών κατά
 αναμνηστικό προκλητής ή φυσιολογικής εμμηνόπαυσης
και κατά ηλικία

	Μέχρι 49 ετών		
	Προκλ	Φυσιολ	Σύνολο
Καρκιν	140	103	243
Μη-καρκ	210	104	314
Σύνολο	350	207	557

	50-59 ετών		
	Προκλ	Φυσιολ	Σύνολο
Καρκιν	176	567	743
Μη-καρκ	145	345	490
Σύνολο	321	912	1233

$$\Sigma\Lambda_1 = (140 \times 104) / (103 \times 210) = 0,67$$

$$\Sigma\Lambda_2 = (176 \times 345) / (567 \times 145) = 0,74$$

	60-69 ετών		
	Προκλ	Φυσιολ	Σύνολο
Καρκιν	111	680	791
Μη-καρκ	93	278	371
Σύνολο	204	958	1162

	70-79 ετών		
	Προκλ	Φυσιολ	Σύνολο
Καρκιν	42	420	462
Μη-καρκ	25	170	195
Σύνολο	67	590	657

$$\Sigma\Lambda_3 = (111 \times 278) / (680 \times 93) = 0,49$$

$$\Sigma\Lambda_4 = (42 \times 170) / (420 \times 25) = 0,68$$

Οι ΣΛ φαίνεται να μην διαφοροποιούνται συστηματικά

χ^2 και σχετικός λόγος κατά Mantel - Haenszel

$$\Sigma \Lambda_{M-H} = 0,64$$

$$\chi^2_{M-H} = 28,84 \quad BE=1$$

$$10,83 < 28,84$$

$$0,1\% > P$$

Τι συμπεραίνουμε;

Συμπέρασμα - Ερμηνεία

Γυναίκες με προκλητή εμμηνόπαυση έχουν $(1 - 0,64) = 0.36 = 36\%$ μειωμένο κίνδυνο ανάπτυξης καρκίνου του μαστού και μετά την διόρθωση για την επίδραση της ηλικίας (ή ανεξάρτητα από την επίδραση της ηλικίας)

Η σχέση μεταξύ προκλητής εμμηνόπαυσης και καρκίνου του μαστού είναι στατιστικά σημαντική στο επίπεδο του 5% ακόμα και μετά την διόρθωση για την επίδραση της ηλικίας

Συμπέρασμα - Ερμηνεία

$\Sigma\Lambda = 0,50$ πριν την διόρθωση κατά MH

$\Sigma\Lambda = 0,64$ μετά την διόρθωση κατά MH

Η ηλικία «άλλαξε» τον $\Sigma\Lambda$, η ηλικία είναι συγχυτικός παράγοντας στην μελετούμενη σχέση