

Διαστρωματική ανάλυση δεδομένων*

Π. Γαθάνης,
Λ. Σπάρος

Εργαστήριο Κλινικής Επιδημιολογίας,
Τμήμα Νοσηλευτικής, Πανεπιστήμιο
Αθηνών, Αθήνα

Stratified analysis of data

Abstract at the end of the article

1. Εισαγωγή
2. Η έννοια του συγχυτή
3. Η έννοια του τροποποιητή
4. Διαστρωματική ανάλυση δεδομένων
5. Άλλες μέθοδοι εξουδετέρωσης των συγχυτών
6. Συμπέρασμα

Λέξεις ευρετηρίου

Διαστρωματική ανάλυση
Διαστρωμάτωση
Σύγχυση
Συγχυτής
Τροποποίηση του μέτρου αποτελέσματος

Υποβλήθηκε 11.7.2003

Εγκρίθηκε 16.7.2003

1. ΕΙΣΑΓΩΓΗ

Η διαστρωματική ανάλυση (stratified analysis) ή διαστρωμάτωση (stratification) είναι ένας τρόπος μελέτης και ελέγχου της επίδρασης άλλων μεταβλητών, εκτός της έκθεσης και της νόσου, στην ανάλυση δεδομένων. Διαστρωμάτωση σημαίνει ότι τα δεδομένα χωρίζονται σε υποομάδες ή στρώματα (strata). Για παράδειγμα, διαστρωμάτωση κατά φύλο ή ηλικία σημαίνει ότι τα δεδομένα χωρίζονται σε κατηγορίες ανδρών και γυναικών ή κατά ηλικία.¹

Ο πρωταρχικός λόγος για χρησιμοποίηση της διαστρωματικής ανάλυσης είναι να εκτιμηθεί και –αν είναι αναγκαίο– να ελεγχθεί η σύγχυση, η οποία προκύπτει όταν κάποια αιτία, άλλη από την έκθεση που μελετάται, είναι περισσότερο ή λιγότερο επικρατής στην εκτεθειμένη ομάδα από ό,τι στη μη εκτεθειμένη. Όταν τα δεδομένα στρωματοποιούνται κατά επίπεδα του υπό σύγχυση παράγοντα, για παράδειγμα σε άνδρες και γυναίκες, κάθε στρώμα θα είναι ελεύθερο σύγχυσης από τη μεταβλητή διαστρωμάτωσης. Έτσι, αν η συσχέτιση μεταξύ έκθεσης και νόσου αναλυθεί ξεχωριστά, π.χ. σε άνδρες και γυναίκες, καθεμιά από τις ομάδες του φύλου θα δώσει μια εκτίμηση του αποτελέσματος της έκθεσης,

ανεξάρτητη του φύλου. Συχνά, αυτά τα επιμέρους αποτελέσματα κατά στρώματα δεν αναφέρονται ξεχωριστά, αλλά ενοποιημένα (pooled) σε ένα μόνο αποτέλεσμα.^{1,2}

Ο δεύτερος λόγος για διαστρωματική ανάλυση είναι η ανάλυση της τροποποίησης του μέτρου αποτελέσματος (effect measure modification). Τροποποίηση του μέτρου αποτελέσματος σημαίνει ότι το αποτέλεσμα της έκθεσης είναι ισχυρότερο σε μερικά στρώματα σε σχέση με άλλα. Για παράδειγμα, αν η σχετική επίπτωση-πυκνότητα μεταξύ έκθεσης και νόσου είναι 2 στις γυναίκες και 3 στους άνδρες, το φύλο θα τροποποιούσε το αποτέλεσμα της έκθεσης και έτσι θα ήταν ένας τροποποιητής του αποτελέσματος (effect modifier).¹

2. Η ΕΝΝΟΙΑ ΤΟΥ ΣΥΓΧΥΤΗ

Η σύγχυση (confounding) στις επιδημιολογικές μελέτες καθορίζεται ως η απόκλιση ενός μέτρου αποτελέσματος που προκαλείται από μια εξωγενή μεταβλητή και συμβαίνει μόνο μέσα στα πλαίσια μιας συγκεκριμένης μελέτης. Τα δύο κυριότερα μέτρα αποτελέσματος σε μια αιτιολογική μελέτη είναι (α) η διαφορά των επιπτώσεων-πυκνότητας (IDe-IDo) και (β) ο λόγος των επιπτώσεων-πυκνότητας (ή σχετική επίπτωση-πυκνότητα, IDe/IDo). Με IDe συμβολίζεται η επίπτωση-πυκνότητα στην κατηγορία των εκτεθειμένων, ενώ με IDo συμβολίζεται η επίπτωση-πυκνότητα στους μη εκτεθειμένους.³ Η

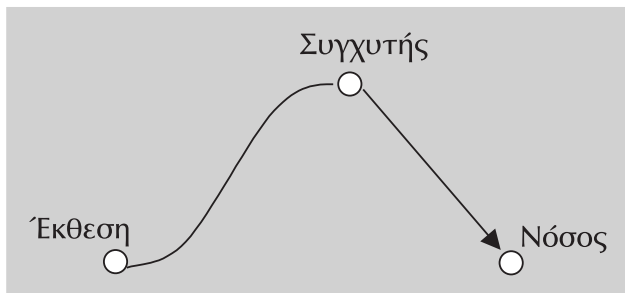
* Μερική χρηματοδότηση: Ειδικός Λογαριασμός Κονδυλίων Έρευνας Πανεπιστημίου Αθηνών

σύγχυση δεν είναι ένα χαρακτηριστικό της κάθε μεταβλητής. Έτσι, μια μεταβλητή που είναι συγχυτής σε μια μελέτη, μπορεί να μην είναι συγχυτής της ίδιας σχέσης σε μια άλλη μελέτη.⁴⁻⁷ Ο συγχυτής (confounder) πρέπει να έχει τα παρακάτω χαρακτηριστικά^{8,9} (εικ. 1):

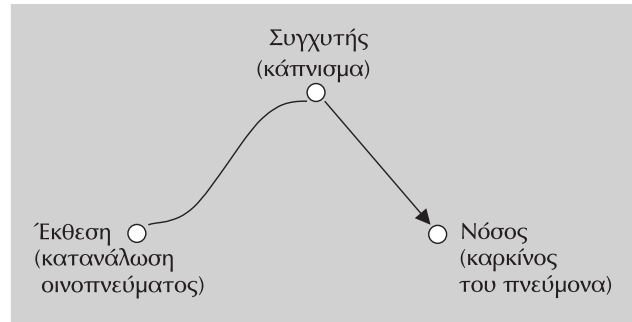
- Ο συγχυτής πρέπει να σχετίζεται με τη νόσο (πρέπει να αποτελεί αιτία της νόσου, αλλά όχι αποτέλεσμα της νόσου).
- Ο συγχυτής πρέπει να σχετίζεται με την έκθεση.
- Ο συγχυτής δεν πρέπει να αποτελεί αποτέλεσμα της έκθεσης. Για παράδειγμα, θεωρήστε μια δίαιτα πλούσια σε κορεσμένα λίπη, η οποία προκαλεί αύξηση των επιπέδων της λιποπρωτεΐνης χαμηλής πυκνότητας (LDL) στο αίμα. Η αυξημένη LDL αυξάνει τις πιθανότητες εμφάνισης της αθηρωματικής νόσου. Έτσι, η αυξημένη LDL σχετίζεται τόσο με τη νόσο (αθηρωματική νόσο) όσο και με την έκθεση (δίαιτα πλούσια σε κορεσμένα λίπη). Ωστόσο, η αυξημένη LDL δεν είναι συγχυτής της σχέσης μεταξύ δίαιτας και αθηρωματικής νόσου, γιατί αποτελεί αποτέλεσμα της έκθεσης, καθώς η πλούσια σε κορεσμένα λίπη δίαιτα προκαλεί αύξηση της LDL.

Ως παράδειγμα σύγχυσης μπορεί να αναφερθεί η μελέτη στην οποία διαπιστώθηκε ότι άτομα που καταναλώνουν μεγάλη ποσότητα οινόπνευματος είχαν υψηλότερο κίνδυνο εμφάνισης καρκίνου του πνεύμονα από εκείνα που δεν καταλάωναν οινόπνευμα. Η αιτία γι' αυτή τη σχέση ήταν ότι τα άτομα της πρώτης κατηγορίας ήταν σε μεγάλη αναλογία βαρείς καπνιστές. Η σύγχυση από το κάπνισμα εξηγούσε την παρατηρηθείσα σχέση μεταξύ κατανάλωσης οινόπνευματος και καρκίνου του πνεύμονα (εικ. 2).¹ Το κάπνισμα είναι συγχυτής της σχέσης μεταξύ οινόπνευματος και καρκίνου του πνεύμονα, καθώς έχει τα παρακάτω χαρακτηριστικά:

- Το κάπνισμα σχετίζεται με τη νόσο, καθώς αποτελεί μια αιτία του καρκίνου του πνεύμονα.
- Το κάπνισμα σχετίζεται με την έκθεση, καθώς τα άτομα που καταναλώνουν μεγάλη ποσότητα οινόπνευ-



Εικόνα 1. Σχηματική απεικόνιση των χαρακτηριστικών ενός συγχυτή.¹



Εικόνα 2. Παράδειγμα συγχυτικής επίδρασης: Η σχέση μεταξύ κατανάλωσης οινόπνευματος και καρκίνου του πνεύμονα.¹

ματος είναι συνήθως και τα άτομα εκείνα που καπνίζουν περισσότερο.

- Το κάπνισμα δεν αποτελεί αποτέλεσμα της έκθεσης, δηλαδή της κατανάλωσης οινόπνευματος.

Μια καλά σχεδιασμένη μελέτη απαιτεί την εξουδετέρωση των συγχυτών, η οποία μπορεί να γίνει με διάφορους τρόπους, όπως η εξομοίωση, η διαστρωμάτωση, η πολυπαραμετρική ανάλυση, η προτύπωση, η περιοριστική μέθοδος και το πείραμα.^{1,4} Η εξουδετέρωση όλων των δυνατών, γνωστών και άγνωστων, συγχυτών είναι απαραίτητη προϋπόθεση για την ανάδειξη μιας σχέσης σε αιτιακή. Η μη εξουδετέρωση όλων των συγχυτών εισάγει συστηματικό σφάλμα στην έρευνα και καθιστά τα αποτελέσματα μη έγκυρα.³ Το γεγονός αυτό έχει ιδιαίτερη σημασία στην περίπτωση της φαρμακοεπιδημιολογίας, καθώς η ύπαρξη συγχυτών επηρεάζει την εγκυρότητα (validity) των αποτελεσμάτων μιας μελέτης και μπορεί να οδηγήσει σε λανθασμένα συμπεράσματα.

3. Η ΕΝΝΟΙΑ ΤΟΥ ΤΡΟΠΟΠΟΙΗΤΗ

Σε αντίθεση με τη σύγχυση, η τροποποίηση καθορίζεται ως η ύπαρξη διαφορετικών τιμών ενός μέτρου αποτελέσματος στα διαφορετικά επίπεδα μιας άλλης μεταβλητής.^{4,10} Εναλλακτικά, θα μπορούσε να λεχθεί ότι τροποποίηση συμβαίνει όταν το αποτέλεσμα της έκθεσης είναι ισχυρότερο σε μερικές κατηγορίες σε σχέση με άλλες.¹ Για παράδειγμα, αν η σχετική επίπτωση-πυκνότητα μεταξύ έκθεσης και νόσου είναι 2 στις γυναίκες και 3 στους άνδρες, τότε το φύλο τροποποιεί το αποτέλεσμα της έκθεσης και είναι τροποποιητής του μέτρου αποτελέσματος (effect measure modifier). Η τροποποίηση είναι ένα εγγενές χαρακτηριστικό της σχέσης μεταξύ δύο αιτιών μιας νόσου.⁴ Αυτή η σχέση δεν εξαρτάται από τις ιδιαιτερότητες κάθε μελέτης, καθώς αποτελεί ένα σταθερό γεγονός της φύσης.

Η τροποποίηση του μέτρου αποτελέσματος πρέπει να λαμβάνεται σοβαρά υπόψη κατά το σχεδιασμό, την ανάλυση, αλλά και την ερμηνεία των αποτελεσμάτων των επιδημιολογικών μελετών.¹¹ Τροποποίηση υπάρχει όταν η σχέση μεταξύ δύο μεταβλητών διαφοροποιείται ανάλογα με τα επίπεδα μιας τρίτης μεταβλητής.^{8,11} Για παράδειγμα, η παχυσαρκία αυξάνει τον κίνδυνο για καρκίνο του μαστού στις μετεμμηνοπαυσιακές γυναίκες, αλλά όχι και στις προεμμηνοπαυσιακές. Έτσι, η ύπαρξη ή όχι της εμμηνόπαυσης τροποποιεί την επίδραση της παχυσαρκίας για τον καρκίνο του μαστού. Η έκθεση στον αμίαντο, π.χ., είναι ένας ισχυρός παράγοντας κινδύνου για τους καπνιστές σε σχέση με τους μη καπνιστές.^{8,11} Συνεπώς, το κάπνισμα τροποποιεί την επίδραση του αμιάντου στη συχνότητα του καρκίνου του πνεύμονα, όπως αυτή εκτιμάται από τη διαφορά ή το λόγο συχνότητων του καρκίνου του πνεύμονα. Η ανακάλυψη της ύπαρξης τροποποίησης στα αποτελέσματα μιας μελέτης είναι ένα σημαντικό στοιχείο της ανάλυσης των επιδημιολογικών δεδομένων.

Οι επιδημιολόγοι χρησιμοποιούν συχνά τον όρο τροποποίηση του αποτελέσματος (effect modification) αντί του όρου τροποποίηση του μέτρου αποτελέσματος (effect measure modification). Η προσθήκη της λέξης μέτρο είναι αναγκαία, για να τονιστεί η εξάρτηση του φαινομένου της αλληλεπίδρασης από την επιλογή του μέτρου αποτελέσματος. Δεν έχει νόημα να αναφέρεται γενικά η παρουσία ή μη της τροποποίησης του αποτελέσματος χωρίς να αναφέρεται το είδος (διαφορά ή λόγος) του μέτρου αποτελέσματος.

Η ύπαρξη ή όχι τροποποίησης του μέτρου αποτελέσματος εξαρτάται αποκλειστικά από το μέτρο αποτελέσματος που χρησιμοποιείται. Συνήθως, όταν η χρήση ενός μέτρου αποτελέσματος (π.χ. διαφορά επιπτώσεων-πυκνότητας) υποδηλώνει την ύπαρξη τροποποίησης, τότε η χρήση ενός άλλου μέτρου αποτελέσματος (π.χ. λόγος επιπτώσεων-πυκνότητας) υποδηλώνει την απουσία τροποποίησης και το αντίστροφο.

4. ΔΙΑΣΤΡΩΜΑΤΙΚΗ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

Όπως αναφέρθηκε παραπάνω, ο πρωταρχικός λόγος για χρησιμοποίηση της διαστρωματικής ανάλυσης είναι η εκτίμηση και –αν είναι αναγκαίο– ο έλεγχος της σύγκυσης, η οποία προκύπτει όταν κάποια αιτία, άλλη από την έκθεση που μελετάται, είναι περισσότερο ή λιγότερο επικρατής στην εκτεθειμένη ομάδα από ό,τι στη μη εκτεθειμένη.^{1,2,8,9,12-15} Όταν τα δεδομένα στρωματοποιούνται κατά επίπεδα του υπό σύγκυση παράγοντα, για παράδειγμα σε άνδρες και γυναίκες, κάθε στρώμα θα είναι ελεύθερο σύγκυσης από τη μεταβλητή διάστρωσης. Έτσι, αν η συσχέτιση μεταξύ έκθεσης και ασθένειας αναλυθεί ξεχωριστά, π.χ. σε άνδρες και γυναίκες, καθεμιά από τις ομάδες του φύλου θα δώσει μια εκτίμηση του αποτελέσματος της έκθεσης, ανεξάρτητη του φύλου. Συχνά, αυτά τα επιμέρους αποτελέσματα κατά στρώματα δεν αναφέρονται ξεχωριστά, αλλά ενοποιημένα (pooled) σε ένα μόνο αποτέλεσμα. Ο δεύτερος λόγος για διαστρωματική ανάλυση είναι η ανάλυση του φαινομένου της τροποποίησης που συμβαίνει όταν το αποτέλεσμα της έκθεσης είναι ισχυρότερο σε μερικές κατηγορίες σε σχέση με άλλες.

Για την καλύτερη κατανόηση των παραπάνω, θεωρήστε τα δεδομένα του πίνακα 1, ο οποίος παρουσιάζει το υλικό από μια επιδημιολογική μελέτη με σχεδιασμό κοορτών. Στην επιδημιολογία, η κοορτή (cohort) χρησιμοποιείται για να περιγράψει τον κλειστό πληθυσμό (closed population).^{1,3,8} Πρόκειται για ένα «κλειστό» σύνολο ατόμων, όπου η ιδιότητα του μέλους καθορίζεται από ένα συμβάν (event) σε μια συγκεκριμένη τοποχρονική περιοχή. Ο χαρακτηρισμός «κλειστός» σημαίνει ότι «απαγορεύεται» η έξοδος των μελών από τον πληθυσμό αυτόν. Η ιδιότητα του μέλους ενός κλειστού πληθυσμού δεν χάνεται ούτε με το θάνατο του μέλους. Παράδειγμα κλειστού πληθυσμού είναι τα άτομα που πέθαναν σε μια συγκεκριμένη περιοχή σε ένα συγκεκριμένο χρονικό διάστημα. Ανοικτός (open) ή δυναμικός (dynamic) πληθυσμός είναι ο πληθυσμός μιας πόλης ή μιας χώρας, οι

Πίνακας 1. Αριθμός ανθρωπο-ετών και αριθμός περιπτώσεων νοσήματος σύμφωνα με τη χρήση οινόπνευματος και καπνού.¹

Οινόπνευμα	Καπνός	Ανθρωπο-έτη	Περιπτώσεις	Επίπτωση-πυκνότητα
Όχι	Όχι	75.000	75	0,001
Όχι	Ναι	25.000	50	0,002
Ναι	Όχι	25.000	50	0,002
Ναι	Ναι	75.000	300	0,004

νοσπλεύόμενοι σε ένα νοσοκομείο κ.ά.^{1,3} Τα μέλη του ανοικτού πληθυσμού εναλλάσσονται στο χρόνο. Ο ανοικτός πληθυσμός, σε αντίθεση με τον κλειστό, δεν γηράσκει. Η ιδιότητα του μέλους του ανοικτού πληθυσμού προσδιορίζεται από μια κατάσταση (state) και διαρκεί όσο διαρκεί η κατάσταση αυτή. Ένα άτομο είναι μέλος του δυναμικού πληθυσμού της Αθήνας όσο ζει στην Αθήνα και για το χρονικό διάστημα που ζει σε αυτή. Χάνει όμως την ιδιότητα του μέλους εφόσον απομακρυνθεί από αυτήν, κάτι που δεν ισχύει για τον κλειστό πληθυσμό.^{1,16}

Ο σκοπός της μελέτης ήταν να εκτιμηθεί η σχέση μεταξύ κατανάλωσης οινόπνευματος και νόσου (καρκίνου του πνεύμονα).¹ Υπήρχε, παρόλα αυτά, η υποψία ότι η χρήση καπνού ίσως να σχετιζόταν με τη χρήση οινόπνευματος και επίσης ότι ο ίδιος ο καπνός μπορούσε να είναι μια επιπλέον αιτία της νόσου. Έτσι, ο καπνός θεωρήθηκε ένας πιθανός συγχυτής στη μελέτη. Γι' αυτόν το λόγο, οι πληροφορίες που συλλέχθηκαν δεν αφορούσαν μόνο στη χρήση οινόπνευματος, αλλά επίσης και καπνού. Αν οι πληροφορίες για τη χρήση καπνού αγνοηθούν, το υλικό θα περιγραφεί όπως στον πίνακα 2.

Ο λόγος των επιπτώσεων-πυκνότητας, που υπολογίστηκε ως το πηλίκο της επίπτωσης-πυκνότητας ανάμεσα σε αυτούς που κατανάλωναν οινόπνευμα προς αυτούς που δεν κατανάλωναν, συνεπάγεται ότι υπάρχει μια ισχυρή συσχέτιση μεταξύ του οινόπνευματος και της νόσου. Αν, παρόλα αυτά, το κάπνισμα είναι περισσότερο συχνό ανάμεσα σε αυτούς που κατανάλωναν οινόπνευμα παρά σε εκείνους που δεν κατανάλωναν και αν, επιπρόσθετα, το κάπνισμα είναι μια αιτία της νόσου, τότε τα παρατηρούμενα αποτελέσματα θα μπορούσαν να προκύψουν

Πίνακας 2. Επίπτωση-πυκνότητα και λόγος επιπτώσεων-πυκνότητας σύμφωνα με τη χρήση οινόπνευματος.¹

Οινόπνευμα	Επίπτωση-πυκνότητα	Λόγος επιπτώσεων-πυκνότητας
Ναι	0,00350	0,00350/0,00125=2,8
Όχι	0,00125	

και χωρίς η χρήση του οινόπνευματος να σχετίζεται αιτιολογικά με τη νόσο. Για να εξεταστεί αυτή η πιθανότητα, τα δεδομένα στρωματοποιούνται σύμφωνα με το κάπνισμα. Σε αυτό το παράδειγμα, επομένως, τα δεδομένα χωρίζονται σε δύο στρώματα, καπνιστές και μη καπνιστές. Τα στρωματοποιημένα δεδομένα περιγράφονται στον πίνακα 3. Ο πίνακας 3 δείχνει ότι και στα δύο στρώματα, π.χ. τόσο για τους καπνιστές όσο και για τους μη καπνιστές, η επίπτωση είναι δύο φορές υψηλότερη για εκείνους που κατανάλωναν οινόπνευμα, σε σχέση με εκείνους που δεν κατανάλωναν.

Ο λόγος των επιπτώσεων-πυκνότητας μειώθηκε από 2,8 σε 2, μέσω του ελέγχου του παράγοντα σύγχυσης «κάπνισμα». Η εκτίμηση του 2,8, που παρατηρήθηκε πριν από την εξουδετέρωση του καπνίσματος, μπορεί να θεωρηθεί ότι αντανακλά ένα μικτό αποτέλεσμα, τόσο της χρήσης οινόπνευματος όσο και αυτής του καπνού.

Μια περαιτέρω ανάλυση του παραδείγματος φανερώνει ότι το κάπνισμα ικανοποιεί τα δύο από τα τρία κριτήρια σύγχυσης που αναφέρθηκαν παραπάνω. Πιο συγκεκριμένα, από τη στήλη που αναφέρεται στον αριθμό των ανθρωπο-ετών, στον πίνακα 1, διαπιστώνεται ότι το κάπνισμα σχετίζεται με την κατανάλωση οινόπνευματος, γιατί ο αριθμός των ανθρωπο-ετών για την κατηγορία έκθεση στο οινόπνευμα και ταυτόχρονη έκθεση στο κάπνισμα (75.000 ανθρωπο-έτη) είναι σημαντικά μεγαλύτερος από τον αριθμό των ανθρωπο-ετών για την κατηγορία έκθεση στο οινόπνευμα αλλά όχι ταυτόχρονη έκθεση στο κάπνισμα (25.000 ανθρωπο-έτη). Παράλληλα, ο πίνακας αποκαλύπτει ότι το κάπνισμα αποτελεί μια αιτία της νόσου –εκτός του οινόπνευματος– καθώς το κάπνισμα μόνο του σχετίζεται με μια αύξηση της επίπτωσης-πυκνότητας, τόσο σε εκείνους που κατανάλωναν οινόπνευμα όσο και σε εκείνους που δεν κατανάλωναν.

Το κάπνισμα είναι ένας συγχυτής της σχέσης μεταξύ κατανάλωσης οινόπνευματος και νόσου. Ωστόσο, δεν είναι ένας τροποποιητής του μέτρου αποτελέσματος, όταν το μέτρο αποτελέσματος που χρησιμοποιείται είναι ο λόγος των επιπτώσεων-πυκνότητας (πίν. 3). Και αυτό

Πίνακας 3. Επίπτωση-πυκνότητα και λόγος επιπτώσεων-πυκνότητας σύμφωνα με τη χρήση οινόπνευματος και καπνού.¹

Κάπνισμα	Οινόπνευμα	Επίπτωση-πυκνότητα (έτος ⁻¹)	Λόγος επιπτώσεων-πυκνότητας
Ναι	Ναι	0,004	0,004/0,002=2
Ναι	Όχι	0,002	
Όχι	Ναι	0,002	0,002/0,001=2
Όχι	Όχι	0,001	

γιατί ο λόγος των επιπτώσεων-πυκνότητας μεταξύ των μη εκτεθειμένων στο κάπνισμα είναι $0,002/0,001=2$, που είναι ίσος με το λόγο των επιπτώσεων-πυκνότητας μεταξύ των εκτεθειμένων στο κάπνισμα ($0,004/0,002=2$). Αν όμως χρησιμοποιηθεί ως μέτρο αποτελέσματος η διαφορά των επιπτώσεων-πυκνότητας, τότε το κάπνισμα είναι ένας τροποποιητής του μέτρου αποτελέσματος (πίν. 4).

Η διαφορά των επιπτώσεων-πυκνότητας μεταξύ των μη εκτεθειμένων στο κάπνισμα είναι $0,002-0,001=0,001$, ενώ η διαφορά των επιπτώσεων-πυκνότητας μεταξύ των εκτεθειμένων στο κάπνισμα είναι $0,004-0,002=0,002$. Επομένως, η χρήση της διαφοράς των επιπτώσεων-πυκνότητας αποδεικνύει ότι το κάπνισμα είναι ένας τροποποιητής του μέτρου αποτελέσματος, ενώ η χρήση του λόγου των επιπτώσεων-πυκνότητας αποδεικνύει ότι το κάπνισμα δεν είναι τροποποιητής του μέτρου αποτελέσματος.

5. ΆΛΛΕΣ ΜΕΘΟΔΟΙ ΕΞΟΥΔΕΤΕΡΩΣΗΣ ΤΩΝ ΣΥΓΧΥΤΩΝ

Μια καλά σχεδιασμένη μελέτη απαιτεί την εξουδετέρωση των συγχυτών, η οποία μπορεί να γίνει με διάφορους τρόπους, όπως η εξομοίωση, η διαστρωμάτωση, η πολυπαραμετρική ανάλυση, η προτύπωση, η περιοριστική μέθοδος και το πείραμα.^{3,5,7,9,12-15,17} Η εξουδετέρωση όλων των δυνατών, γνωστών και άγνωστων, συγχυτών είναι απαραίτητη προϋπόθεση για την ανάδειξη μιας σχέσης σε αιτιακή. Η μη εξουδετέρωση όλων των συγχυτών εισάγει συστηματικό σφάλμα στην έρευνα και καθιστά τα αποτελέσματα μη έγκυρα.³

Στην εξομοίωση (matching) επιλέγεται μια ομάδα ατόμων, έτσι ώστε να είναι παρόμοια με μιαν άλλη ομάδα, ως προς την κατανομή μίας ή περισσότερων μεταβλητών.^{1,8,12,17,18} Υπάρχουν δύο μέθοδοι για να εξασφαλιστεί η ομοιότητα των συγκρινόμενων ομάδων ως προς τους δυνητικούς συγχυτές: η μέθοδος της αναλογικής εξομοίωσης (frequency matching) και η μέθοδος της ατομικής εξομοίωσης (individual matching).^{19,20} Η μέθοδος της αναλογικής εξομοίωσης εφαρμόζεται όταν έχει

ολοκληρωθεί η διαμόρφωση της ομάδας των εκτεθέντων ατόμων και είναι γνωστή η κατανομή τους ως προς τους δυνητικούς συγχυτές. Έτσι, σε δείγμα καπνιστών στο οποίο υπάρχουν 70% άνδρες και 30% γυναίκες, θα πρέπει η συλλογή των μη καπνιστών να γίνει με τέτοιο τρόπο, ώστε η ομάδα τους να περιέχει τις δύο παραπάνω κατηγορίες στις ίδιες περίπου αναλογίες. Αντίθετα, η μέθοδος της ατομικής εξομοίωσης εφαρμόζεται όταν η διαμόρφωση των ομάδων των εκτεθέντων και των μη εκτεθέντων ατόμων γίνεται περίπου συγχρόνως,^{21,22} οπότε για κάθε επιλεγόμενο καπνιστή ορισμένου φύλου επιλέγεται ένας ορισμένος αριθμός μη καπνιστών του ίδιου φύλου (συνήθως 1-4 μη εκτεθειμένα άτομα για κάθε εκτεθειμένο ή αντίστροφα).

Η προτύπωση (standardization) χρησιμοποιείται συνήθως για τη σύγκριση της θνησιμότητας μεταξύ πληθυσμών που έχουν διαφορετική κατανομή κατά ηλικία ή φύλο, ιδιαίτερα στην επαγγελματική επιδημιολογία.^{1,8,9,12,18} Υπάρχουν αρκετές τεχνικές προτύπωσης, αλλά δύο από αυτές χρησιμοποιούνται συχνότερα: η μέθοδος της άμεσης προτύπωσης (direct standardization) και η μέθοδος υπολογισμού του «προτυπωμένου πληθικού θνησιμότητας» (standardized mortality ratio, SMR).²³ Στη μέθοδο της άμεσης προτύπωσης επιλέγεται ένας πληθυσμός που θεωρείται ως «πρότυπος». Η επιλογή μπορεί να γίνει αυθαίρετα, αρκεί να χρησιμοποιείται ο ίδιος «πρότυπος» πληθυσμός για όλες τις συγκρινόμενες ομάδες. Συνεπώς, ως «πρότυπος» πληθυσμός μπορεί να θεωρηθεί ο πληθυσμός της Ευρώπης το 1960 ή ο πληθυσμός της Ελλάδας το 2000. Οποιοσδήποτε και αν είναι ο πρότυπος πληθυσμός, έχει μια δεδομένη σταθερή σύνθεση κατά ηλικία. Κατά την εφαρμογή της άμεσης μεθόδου προτύπωσης, οι ειδικοί κατά ηλικία δείκτες θνησιμότητας του καθένα από τους συγκρινόμενους πληθυσμούς (π.χ. του πληθυσμού Α) εφαρμόζονται (πολλαπλασιάζονται) στις αντίστοιχες ομάδες ηλικιών του πρότυπου πληθυσμού. Έτσι υπολογίζονται οι θάνατοι που θα συνέβαιναν στον πρότυπο πληθυσμό, αν ίσχυαν σε αυτόν οι ειδικοί κατά ηλικίες δείκτες θνησιμότητας του πληθυσμού Α. Η αδρή (γενική) θνησιμότητα του πρότυπου πληθυσμού, που θα υπήρχε αν ίσχυαν σε αυτόν οι κατά ηλικίες δείκτες

Πίνακας 4. Επίπτωση-πυκνότητα και διαφορά επιπτώσεων-πυκνότητας σύμφωνα με τη χρήση οιοπνεύματος και καπνού.¹

Κάπνισμα	Οιοπνευμα	Επίπτωση-πυκνότητα (έτος ⁻¹)	Διαφορά επιπτώσεων-πυκνότητας
Ναι	Ναι	0,004	0,004-0,002=0,002
Ναι	Όχι	0,002	
Όχι	Ναι	0,002	0,002-0,001=0,001
Όχι	Όχι	0,001	

θνησιμότητας του πληθυσμού A, είναι ο «προτυπωμένος κατά ηλικία δείκτης θνησιμότητας του πληθυσμού A» (με την άμεση μέθοδο και με βάση το συγκεκριμένο πληθυσμό που επιλέχθηκε ως πρότυπος). Εφόσον προτυπωθούν στον ίδιο πρότυπο πληθυσμό, οι δείκτες θνησιμότητας των πληθυσμών A, B, Γ κ.ά. μπορούν να συγκριθούν μεταξύ τους, γιατί οι προτυπωμένοι δείκτες βασίζονται σε πληθυσμό με την ίδια σύνθεση ηλικίας.

Η μέθοδος υπολογισμού του «προτυπωμένου πληθυσμίου θνησιμότητας» ονομάζεται και έμμεση προτύπωση και βασίζεται στην εξής αρχή. Οι ειδικοί κατά ηλικίες δείκτες ενός ορισμένου πληθυσμού αποτελούν τη βάση των υπολογισμών, ενώ η κατά ηλικία σύνθεση αυτού του πληθυσμού δεν έχει οποιαδήποτε σημασία. Οι δείκτες αυτοί («πρότυποι δείκτες») εφαρμόζονται στις αντίστοιχες ομάδες ηλικιών του πληθυσμού, στον οποίο γίνεται η προτύπωση. Έτσι υπολογίζεται ο αριθμός των θανάτων που θα συνέβαιναν στον πληθυσμό που εξετάζεται, αν ίσχυαν σε αυτόν οι πρότυποι δείκτες (αναμενόμενοι θάνατοι=E), και στη συνέχεια ο αριθμός αυτός συγκρίνεται με τον αριθμό των θανάτων που συνέβησαν πραγματικά (O). Το πηλίκο (O/E), πολλαπλασιαζόμενο με το 100, αποτελεί το προτυπωμένο πηλίκο θνησιμότητας (SMR). Όταν το SMR είναι >100, σημαίνει ότι η προτυπωμένη θνησιμότητα στον εξεταζόμενο πληθυσμό είναι μεγαλύτερη από τη θνησιμότητα στον πληθυσμό του οποίου οι δείκτες θεωρήθηκαν ως πρότυποι δείκτες και αντίστροφα. Έτσι, αναφέρεται ότι το SMR των μεταλλωρύχων είναι >100, ενώ το SMR των κηπουρών είναι <100, με βάση (=100) τη θνησιμότητα στο γενικό πληθυσμό όλων των επαγγελματιών μαζί. Η έμμεση προτύπωση είναι ιδιαίτερα χρήσιμη, όταν ο αριθμός των θανάτων στον εξεταζόμενο πληθυσμό είναι μικρός και όταν δεν είναι γνωστή η κατά ηλικία κατανομή των θανάτων αυτών.

Στην περιοριστική μέθοδο (restriction), η μελέτη αφορά μόνο στο ένα επίπεδο του συγχυτή. Για παράδειγμα, αν

μελετηθεί το αποτέλεσμα της δράσης ενός φαρμάκου μόνο σε μία ηλικιακή κατηγορία, τότε εξουδετερώνεται η σύγχυση που προκαλείται από τη διαφορετική κατανομή κατά ηλικία. Ωστόσο, με τον τρόπο αυτόν τα αποτελέσματα της μελέτης περιορίζονται σε μια συγκεκριμένη ηλικιακή κατηγορία.^{8,12,18}

Τέλος, το πείραμα (experiment) αποτελεί μια έρευνα όπου η έκθεση ελέγχεται από τον ερευνητή, με σκοπό τη βελτίωση της εγκυρότητας της έρευνας, όπως, π.χ., με τυχαία χορήγηση της έκθεσης στα άτομα του πληθυσμού της μελέτης (τυχαιοποιημένο πείραμα).^{1,8}

6. ΣΥΜΠΕΡΑΣΜΑ

Η ύπαρξη συγχυτών σε μια μελέτη αποτελεί σφάλμα, το οποίο ο ερευνητής πρέπει να εντοπίσει και να εξουδετερώσει, έτσι ώστε τα αποτελέσματα της μελέτης να είναι έγκυρα. Ωστόσο, η τροποποίηση του μέτρου αποτελέσματος δεν αποτελεί σφάλμα, αλλά ένα εύρημα, το οποίο πρέπει να εντοπίζεται και να αναφέρεται. Γι' αυτό, οι ερευνητές πρέπει να εξουδετερώνουν τους συγχυτές και να ανακαλύπτουν την ύπαρξη της τροποποίησης του μέτρου αποτελέσματος.^{8,9} Για το σκοπό αυτό χρησιμοποιείται η διαστρωματική ανάλυση των δεδομένων μιας μελέτης.¹ Ο πρωταρχικός λόγος για χρησιμοποίηση της διαστρωματικής ανάλυσης είναι η εκτίμηση και, αν είναι αναγκαίο, ο έλεγχος της σύγχυσης. Ο δεύτερος λόγος για διαστρωματική ανάλυση είναι η ανάλυση του φαινομένου της τροποποίησης, που συμβαίνει όταν το αποτέλεσμα της έκθεσης είναι ισχυρότερο σε μερικές κατηγορίες σε σχέση με άλλες. Μια καλά σχεδιασμένη μελέτη απαιτεί την εξουδετέρωση των συγχυτών, η οποία μπορεί να γίνει με διάφορους τρόπους, εκτός από τη διαστρωματική ανάλυση, όπως με την εξομοίωση, την πολυπαραμετρική ανάλυση, την προτύπωση, την περιοριστική μέθοδο και το πείραμα.

ABSTRACT

Stratified analysis of data

P. GALANIS, L. SPAROS

Laboratory of Clinical Epidemiology, School of Nursing, University of Athens, Athens, Greece

Archives of Hellenic Medicine 2004, 21(4):378-384

Stratified analysis or stratification is used to study and control for the effect of other than variables exposure and disease, in data analysis. Stratification means that the data are separated into categories or strata. For example stratification by sex or age means that the data are separated into male and female groups or into categories by age. The aims of stratification in an epidemiological study are control of confounding and analysis

of effect measure modification. A simple definition of confounding would be the confusion, or mixing, of effects. This definition implies that the effect of the exposure is mixed with the effect of another variable, leading to a bias. When data are stratified by confounder, for example, into men and women, each stratum would be free of confounding by that variable. Thus, if the correlation between exposure and disease is analyzed separately in men and women, each category of gender would give an estimation of the result of exposure, independent of gender. A well-designed study requires control for confounding, which can be achieved in various ways, apart from stratification, such as adjustment, standardization, restriction and experiment. Effect measure modification should be taken into account during the planning, analysis and interpretation of results of epidemiological studies. Modification exists when the relation between two variables is different depending on the levels of a third variable. For example, obesity increases the risk for breast cancer in postmenopausal but not in premenopausal women. In this way the presence or absence of menopause modifies the effect of obesity on breast cancer. Thus menopausal status is an effect modifier of obesity on breast cancer. The modifier is the characteristic of individuals which modifies the effect of an exposure on the incidence of a disease, or in other words, the characteristic that modifies the relation between exposure and disease.

Key words: Confounder, Confounding, Effect measure modification, Stratification, Stratified analysis

Βιβλιογραφία

1. AHLBOM A, NORELL S. *Εισαγωγή στη σύγχρονη επιδημιολογία*. Ιατρικές Εκδόσεις Λίτσας, Αθήνα, 1992:137–143
2. GAIL MH. Stratification. In: Armitage P, Colton T (eds) *Encyclopedia of biostatistics*. John Wiley and Sons, New York, 1998:43–46
3. ΣΠΑΡΟΣ Λ. *Μεταεπιδημιολογία*. Εκδόσεις ΒΗΤΑ, Αθήνα, 2001:37–41, 45–50
4. ROTHMAN KJ. Causes. *Am J Epidemiol* 1976, 104:587–592
5. STRIKE PW. *Statistical methods in laboratory medicine*. Butterworth-Heinemann, Oxford, 1991:499–502
6. CLAYTON D, HILLS M. *Statistical models in epidemiology*. Oxford University Press, Oxford, 1993:261–270
7. ΚΑΗΝ ΗΑ, ΣΕΜΠΟΣ ΣΤ. *Statistical methods in epidemiology*. Oxford University Press, Oxford, 1989:105–108
8. ROTHMAN KJ. *Epidemiology: An introduction*. Oxford University Press, New York, 2002:1–3, 101–111, 158–166
9. ROTHMAN KJ, GREENLAND S. *Modern epidemiology*. 2nd ed. Lippincott Williams & Wilkins, Philadelphia, 1998:253–256, 260, 261
10. MANTEL N, BROWN C, BYAR DP. Tests for homogeneity of effect in an epidemiologic investigation. *Am J Epidemiol* 1977, 106:125–129
11. BROWNSON RC, PETITTI DB. *Applied epidemiology*. Oxford University Press, Oxford, 1998:49–50
12. ΤΡΙΧΟΠΟΥΛΟΣ Δ. *Επιδημιολογία*. Εκδόσεις Παρισιάνος, Αθήνα, 1982:91–94, 194
13. FREEMAN J, GOLDMANN DA, MCGOWAN JE. Confounding and the analysis of multiple variables in hospital epidemiology. *Infect Control* 1987, 8:465–473
14. FREEMAN J, GOLDMANN DA, MCGOWAN JE. Methodologic issues in hospital epidemiology. IV. Risk ratios, confounding, effect modification, and the analysis of multiple variables. *Rev Infect Dis* 1988, 10:1118–1141
15. CONSONNI D, BERTAZZI PA, ZOCCHETTI C. Why and how to control for age in occupational epidemiology. *Occup Environ Med* 1997, 54:772–776
16. MIETTINEN OS. *Theoretical epidemiology. Principles of occurrence research in medicine*. John Wiley and Sons, New York, 1985:5–10, 262–265
17. RIGBY AS, ROBINSON MB. Statistical methods in epidemiology. IV. Confounding and the matched odds ratio. *Disabil Rehabil* 2000, 22:259–265
18. STROM BL. *Pharmacoepidemiology*. 2nd ed. John Wiley and Sons, New York, 1994:619–624, 695
19. WORCESTER J. Matched samples in epidemiologic studies. *Biometrics* 1964, 20:840–848
20. MIETTINEN OS. The matched pairs design in the case of all-or-none responses. *Biometrics* 1968, 24:339–352
21. DALING JR, EMANUEL I. Induced abortion and subsequent outcome of pregnancy in a series of American women. *N Engl J Med* 1977, 297:1241–1245
22. SILVERMAN WA, FERTIG JW, BERGER AP. The influence of the thermal environment upon the survival of newly-born premature infants. *Pediatrics* 1958, 22:876–886
23. DECOUFLE P, THOMAS TL, PICKLE LW. Comparison of the proportionate mortality ratio and standardized mortality ratio risk measures. *Am J Epidemiol* 1980, 111:263–269

Corresponding author:

L.D. Sparos, 123 Papadiamadopoulou street, GR-115 27 Athens, Greece
e-mail: pgalanis@nurs.uoa.gr