



Εθνικό Μετσόβιο Πολυτεχνείο

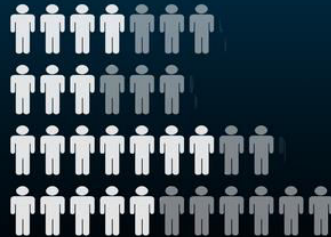
*Σχολή Εφαρμοσμένων Μαθηματικών
και Φυσικών Επιστημών*

*Μεταπτυχιακό Εφαρμοσμένες Μαθηματικές
Επιστήμες, Τομέας Μαθηματικών*

ΜΕΘΟΔΟΙ ΑΝΑΛΥΣΗΣ ΕΠΙΒΙΩΣΗΣ, ΚΛΙΝΙΚΕΣ ΔΟΚΙΜΕΣ ΚΑΙ
ΔΕΙΓΜΑΤΟΛΗΠΤΙΚΕΣ ΤΕΧΝΙΚΕΣ
ΣΤΗΝ ΒΙΟΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΤΗΝ ΕΠΙΔΗΜΙΟΛΟΓΙΑ



Analysis



Επιβλέπων Καθηγητής:
Κουκουβίνος Χρήστος
Καθηγητής Ε.Μ.Π

Διπλωματική Εργασία
της
Σωτηροπούλου Μαρίας

Αθήνα, Ιούλιος 2015

Ευχαριστίες

Η ολοκλήρωση της μεταπτυχιακής αυτής εργασίας θα ήταν αδύνατη χωρίς την πολύτιμη υποστήριξη του καθηγητή μου , Καθηγητή του ΕΜΠ, Κου Χρήστου Κουκουβίνου. Του οφείλω ένα μεγάλο ευχαριστώ γιατί μετά από μία πολύ δύσκολη περίοδο για μένα , μου έδωσε την ευκαιρία να ολοκληρώσω το μεταπτυχιακό μου και μου χάρισε κουράγιο και ενθουσιασμό να συνεχίσω. Επιπλέον θα ήθελα να εκφράσω την βαθιά εκτίμηση μου καθώς ως εκπαιδευτικός αποτέλεσε την αφορμή να αγαπήσω την στατιστική και υπήρξε σημαντικός καθοδηγητής μου σε όλη την διάρκεια της φοίτησης στο μεταπτυχιακό.

Ένα εξίσου μεγάλο ευχαριστώ θα ήθελα να πω και στον σημαντικό συμπαραστάτη μου σε όλη αυτή την διαδικασία ,τον διδάκτορα του ΕΜΠ, Μάνο Ανδρουλακη και να του εκφράσω την βαθιά μου ευγνωμοσύνη για την προσφορά του απαραίτητου υλικού που χρειαζόμουν για την εκπόνηση της εργασίας και ακόμα μεγαλύτερη ευγνωμοσύνη για την συνεχή συμπαράσταση μέχρι την ολοκλήρωση της και τον πρόθυμο τρόπο με τον οποίο αντιμετώπιζε όλες μου τις απορίες. Εύχομαι μία επιτυχημένη πορεία σε οτιδήποτε αποφασίσει να ακολουθήσει από εδώ και στο εξής.

Περίληψη

Βιοστατιστική ονομάζεται ο τομέας της Στατιστικής που ασχολείται με την εφαρμογή των στατιστικών μεθόδων στη Βιολογία, και κυρίως, στην Ιατρική. Πιο συγκεκριμένα συμπεριλαμβάνει κάθε ποσοτική προσέγγιση που είναι ικανή να δώσει απαντήσεις σε ερευνητικά ερωτήματα ή να ελέγξει την ορθότητα επιστημονικών θεωριών, υποθέσεων και σεναρίων. Οι πιο γνωστές εφαρμογές στην Βιοστατιστική είναι οι Μπεϋζιανές μέθοδοι, η Ανάλυση επιβίωσης, η Επιδημιολογία και οι Κλινικές δοκιμές, η εκτενής αναφορά των οποίων γίνεται στην εργασία.

Οι Μπεϋζιανές μέθοδοι είναι οι μέθοδοι προσέγγισης της πιθανότητας βασισμένες σε διαθέσιμες ενδείξεις και λογικά επιχειρήματα. Είναι δηλαδή μέθοδοι ανάπτυξης υποκειμενικής πιθανότητας.

Η ανάλυση επιβίωσης ασχολείται με τις μελέτες στις οποίες η μεταβλητή καθορίζεται από το χρονικό διάστημα που περνάει πριν συμβεί κάποιο γεγονός. Το γεγονός αυτό μπορεί να είναι θάνατος ως συμβάν, ή ο θάνατος εξαιτίας κάποιας συγκεκριμένης ασθένειας.

Οι κλινικές δοκιμές είναι πειράματα κάτω από ελεγχόμενες συνθήκες με σκοπό να συγκριθούν η αποτελεσματικότητα και η ασφάλεια των ασθενών όταν υπάρχει ιατρική παρέμβαση. Οι ασθενείς που συμμετέχουν σε αυτά υποφέρουν από κάποια ή κάποιες συγκεκριμένες ασθένειες, και η συμμετοχή τους στις κλινικές δοκιμές γίνεται με σκοπό να βελτιωθεί η κατάσταση της υγείας τους.

Τέλος, η επιδημιολογία είναι ο κλάδος της βιοστατιστικής που ασχολείται με την κατανομή μιας συγκεκριμένης ασθένειας στον πληθυσμό ή των παραγόντων που επηρεάζουν την κατανομή αυτή. Στο πεδίο της ανήκει και η μελέτη χρόνιων ασθενειών καθώς επίσης και μεταδοτικών ασθενειών που οδηγούν στην αύξηση της επιδημίας των κλασσικών τύπων ασθενειών.

Abstract

Biostatistics (or biometry) is the application of statistics to a wide range of topics in biology. The science of biostatistics encompasses the design of biological experiments, especially in medicine. More specifically it includes any quantitative approach that is able to provide answers to research questions or to verify the accuracy of scientific theories, assumptions and scenarios. The best known applications in Biostatistics are Bayesian inferences, Survival analysis, Epidemiology and clinical trials which will be thoroughly and extensively analyzed in this paper.

Bayesian inferences are probability approach methods based on available evidence and logical arguments. Thus, Bayesian inferences are closely related to subjective probability.

Survival analysis deals with studies in which the variable is defined by the time interval that passes before the occurrence of an event. This can be a death incident, or death due to a particular disease.

Clinical trials are experiments under controlled conditions in order to compare the efficacy and safety of patients under medical intervention. Patients who participate in them suffer from any or some specific diseases, and their participation in clinical trials is to improve the state of their health.

Finally, Epidemiology is the science that studies the patterns, causes, and effects of health and disease conditions in defined populations. Major areas of epidemiological studies include the study of chronic diseases and communicable diseases that lead to the increase of the epidemic of classical types of diseases.

Πίνακας περιεχομένων

Μπεϋζιανές μέθοδοι (Bayesian methods)	5
1.1 Εισαγωγή : Υποκειμενική και αντικειμενική πιθανότητα	5
Υπόβαθρο.....	10
Προγενέστερες και μεταγενέστερες κατανομές (Prior and posterior distributions)	11
Μεταγενέστερες κατανομές (Posterior distributions)	12
Προγενέστερες Κατανομές (Prior distributions)	18
1.2 Το Μπεϋζιανό γραμμικό μοντέλο (Bayesian Linear Model)	20
Γραμμικό μοντέλο με γνωστή διακύμανση	21
Συζυγείς και συσχετισμένες αναλύσεις των δειγμάτων που ακολουθούν κανονική κατανομή με άγνωστη διακύμανση	23
Ανάλυση Επιβίωσης (Survival Analysis).....	29
2.1 Εισαγωγή	29
2.2 Life-Table	31
2.3 Μελέτες υπό παρακολούθηση	34
2.4 Δειγματικά σφάλματα στο life-table.....	40
2.5 Ο εκτιμητής Kaplan-Meier	42
2.6 Το Logrank test.....	43
2.7 Παραμετρικές μέθοδοι.....	51
2.8 Αναλογικά μοντέλα κινδύνου και μοντέλα παλινδρόμησης.....	54
Αναλογικό μοντέλου κινδύνου του Cox(Cox's proportional-hazards model)	56
Κλινικές δοκιμές (clinical trials).....	58
3.1 Εισαγωγή	58
3.2 Κλινικές δοκιμές Phase I και Phase II	59
3.3 Σχεδιάζοντας μία Phase III δοκιμή	61
Το πρωτόκολλο	62
Ο καθορισμός των ασθενών.....	63
Ο καθορισμός των θεραπειών.....	64
Βασικές μεταβλητές και μεταβλητές απόκρισης(baseline and response variable). 64	
Μέγεθος δοκιμής.....	67
Τυχαιοποίηση	67
Ισορροπία.....	68
Κατανομή εξαρτώμενη από τα δεδομένα	69
Τυχαιοποιημένη συναίνεση	70
3.5 Η αξιολόγηση της απόκρισης.....	71

3.6 Ο έλεγχος των δεδομένων	77
Η επιτροπή ελέγχου των δεδομένων	78
Διαδοχική Ανάλυση	78
Ομαδικά σχέδια αλληλουχίας	83
Τυχαία περικοπή (stochastic curtailment)	85
3.7 Ερμηνεία των αποτελεσμάτων της δοκιμής	86
Αριθμός που χρειάζεται να αναλυθεί (number needed to treat-NNT)	86
3.8 Ειδικοί σχεδιασμοί	88
Διασταυρωμένη δοκιμή (cross-over trial)	88
Ο απλός διασταυρωμένος σχεδιασμός (the simple cross-over design)	88
Θεραπεία × περίοδος αλληλεπίδρασης και διασταύρωσης.....	94
Άλλες μεταβλητές απόκρισης.....	97
Δοκιμές n-of-1.....	97
Ισοδύναμες δοκιμές.....	98
Τυχαιοποιημένα συμπλέγματα (cluster randomization).....	100
Στατιστικές μέθοδοι στην επιδημιολογία (statistical methods in epidemiology).....	103
4.1 Εισαγωγή	103
4.2 Ο σχεδιασμός των ερευνών	104
Συστηματικό δείγμα (systematic sampling)	106
Στρωματοποιημένο δείγμα (stratified sampling)	107
Πολυεπίπεδο δείγμα (multistage sampling)	111
Ομαδικό δείγμα (cluster sampling)	113
Επίδραση σχεδιασμού (design effect).....	113
Σημαδεμένα-Ξανασυλληφθέντα δείγματα (mark-recapture sampling).....	115
Απόδοση (imputation)	116
4.3 Δείκτες και προτυποποιήσεις	117
Προτυποποίηση	119
Άμεση μέθοδος.....	121
Έμμεση μέθοδος.....	124
4.4 Έρευνες για την διερεύνηση συνδέσμων	126
4.5 Σχετικός κίνδυνος(Relative Risk)	132
4.6 Αποδοτέος κίνδυνος(Attributable Risk).....	141
4.7 Μέθοδος κατά ηλικία(Subject-years method).....	145
4.8 Ανάλυση ηλικίας-περιόδου-ομάδας(Age-Period-Cohort analysis)	150
4.9 Το μέτρο Κάπα της συμφωνίας(Kappa measure of agreement)	154
Έμμεση μέθοδος.....	158

ΚΕΦΑΛΑΙΟ 1

Μπεϋζιανές μέθοδοι (Bayesian methods)

1.1 Εισαγωγή : Υποκειμενική και αντικειμενική πιθανότητα

Η προσέγγιση μας για την ερμηνεία της πιθανότητας, και την εφαρμογή της στην Στατιστική συμπερασματολογία βασίζεται μέχρι σήμερα στις μεθόδους *frequentist*. Αυτό σημαίνει, ότι θεωρούμε την πιθανότητα ενός τυχαίου γεγονότος ως το ποσοστό των περιπτώσεων στις οποίες το γεγονός αυτό συμβαίνει, υπό συγκεκριμένες συνθήκες. Με παρόμοιο τρόπο μία P-value τιμή καθορίζεται ως η αναλογία των δοκιμών κατά την οποία κάποια παρατηρούμενα αποτελέσματα θα μπορούσαν να είχαν παρατηρηθεί στην ψευδή υπόθεση. Και ένα διάστημα εμπιστοσύνης χαρακτηρίζεται από την πιθανότητα της ένταξης της πραγματικής τιμής της παραμέτρου στα επαναλαμβανόμενα δείγματα.

Η Μπεϋζιανή θεωρία μας επιτρέπει να προσδιορίσουμε προγενέστερες πιθανότητες (*prior probabilities*) βασισμένες σε υποθέσεις πριν τη συλλογή δεδομένων και μεταγενέστερες πιθανότητες (*posterior probabilities*) μετά την συλλογή των δεδομένων. Στην ιατρική διάγνωση, για παράδειγμα, θα μπορούσαμε να μιλήσουμε για τις πιθανότητες των δεδομένων (συμπτώματα κτλ.) σε βέβαιες υποθέσεις (διαγνώσεις) και να αποδώσουμε τις πιθανότητες στην διάγνωση που εμφανίστηκε σύμφωνα με το ιστορικό παρόμοιου τύπου ασθενών.

Θα ήταν ελκυστικό αν κάποιος μπορούσε να προσδιορίσει τις πιθανότητες σε υποθέσεις όπως η ακόλουθη : « Η χρήση αντιτετανικού εμβολίου σε περιπτώσεις κλινικού τετάνου μειώνει την θνησιμότητα από την ασθένεια περισσότερο από 20%». Μια τέτοια προσέγγιση γίνεται πιθανή μόνο αν ερμηνεύσουμε την πιθανότητα μιας υπόθεσης σαν ένα μέτρο του δικού μας βαθμού εμπιστοσύνης στην αλήθεια της. Μία μηδενική πιθανότητα θα μπορούσε να αντιστοιχεί στην απόλυτη δυσπιστία, μία τιμή που αντιπροσωπεύει την απόλυτη βεβαιότητα.

Είναι φανερό ότι κάτι τέτοιο είναι περισσότερο ‘φυσική’ ερμηνεία της πιθανότητας παρά προσέγγιση της με βάση την συχνότητα. Και οι μη στατιστικοί χρήστες των στατιστικών μεθόδων συχνά παρερμηνεύουν τα συμπεράσματα των τεστ σημαντικότητας ή των διαστημάτων εμπιστοσύνης βασισμένοι σε τέτοιες υποκειμενικές ερμηνείες. Αυτό σημαίνει ότι ένα μη σημαντικό αποτέλεσμα ίσως λόγω παρερμηνείας

δείχνει ότι η ψευδής υπόθεση έχει χαμηλή πιθανότητα , και ισχυρίζεται ότι ίσως μια παράμετρος έχει 95 % πιθανότητα να βρίσκεται εντός του διαστήματος εμπιστοσύνης. Η συμφωνία αυτή δεν θα έπρεπε να χρησιμοποιείται για να αιτιολογήσει μία λανθασμένη ερμηνεία αλλά για να παροτρύνει την προσπάθεια ανάπτυξης λογικών προσεγγίσεων λαμβάνοντας υπόψη το βαθμό αξιοπιστίας.

Το βασικό πρόβλημα έγκειται στο πως θα καθοριστούν οι προγενέστερες πιθανότητες σε καταστάσεις όπου οι ερμηνεία της συχνότητας δεν βγάζει νόημα, αλλά χρειάζονται οι τιμές των άκρων ,μεταξύ δηλαδή της απόλυτης δυσπιστίας και της απόλυτης βεβαιότητας. Μία προσέγγιση είναι να αναρωτηθεί κάποιος τι αναλογία είναι διατεθειμένος να δεχτεί αν στοιχηματίσει στην αλήθεια ή το ψέμα μιας συγκεκριμένης κατάστασης. Αν η αποδεκτή αναλογία κρίθηκε να είναι 4 έναντι 1 , τότε η αντίστοιχη πιθανότητα είναι 1/5 ή 0,2. Παρόλα αυτά, η πρόβλεψη των υποθέσεων κρύβει ένα ρίσκο που πολλές φορές δεν αντιλαμβανόμαστε , και για το λόγο αυτό η παραπάνω λογική φαίνεται να είναι ακατάλληλη για έναν μεγάλο αριθμό εκτιμήσεων των πιθανοτήτων. Είναι συνεπώς πιο πρακτικό να χρησιμοποιήσουμε μερικές πιο ευέλικτες προσεγγίσεις για να συλλάβουμε τα βασικά χαρακτηριστικά μιας προγενέστερης εκτίμησης για την αξιοπιστία των διαφορετικών υποθέσεων.

Οι περισσότερες εφαρμογές της στατιστικής εσωκλείουν στα μοντέλα συμπεράσματα για τις παραμέτρους. Συχνά επίσης συμβαίνει να προτείνουν μια οικογένεια κατανομών για τις παραμέτρους , η ποικιλομορφία των οποίων επιτρέπει επαρκή ελαστικότητα στο να καλυφθούν οι ανάγκες των περισσότερων καταστάσεων. Στο ένα άκρο είναι κατανομές με πολύ μεγάλη διασπορά, για να εκπροσωπούν καταστάσεις όπου ο χρήστης έχει μικρή προγενέστερη γνώση, και στο άλλο άκρο είναι κατανομές με πολύ μικρή διασπορά όπου ο χρήστης είναι πεπεισμένος ότι η παράμετρος βρίσκεται εντός ενός μικρού διαστήματος.

Το πρώτο άκρο στο οποίο αναφερθήκαμε , οδηγώντας σε μία prior με μεγάλη διασπορά, παρουσιάζει ιδιαίτερο ενδιαφέρον γιατί υπάρχουν πολλές καταστάσεις στις οποίες ο ερευνητής έχει πολύ λίγες βάσεις δεδομένων για να σχηματίσει μία εικασία , ιδιαίτερα όταν η επιστημονική έρευνα γίνεται για πρώτη φορά. Στις περιπτώσεις αυτές φαίνεται πολύ δελεαστικό να προταθεί μια προγενέστερη κατανομή, η οποία θα μπορούσε να δώσει ίσες πιθανότητες για όλες τις τιμές των παραμέτρων. Η προσέγγιση αυτή όμως είναι ασαφής , διότι μία ομοιόμορφη κατανομή πιθανότητας σε όλη την έκταση των τιμών μιας παραμέτρου θα μπορούσε να οδηγήσει σε μία μη ομοιόμορφη κατανομή πάνω σε μία τροποποιημένη κλίμακα μιας μέτρησης. Για παράδειγμα , για μια παράμετρο θ η οποία συμβολίζει την αναλογία των επιτυχιών σε ένα πείραμα , μια

ομοιόμορφη κατανομή του θ μεταξύ του μηδενός και του ενός δεν θα οδηγούσε σε μία ομοιόμορφη κατανομή $\text{logit}^{(1)}$ του θ μεταξύ του $-\infty$ και του $+\infty$.

Αυτό το πρόβλημα ήταν ένα από τα βασικότερα αντικείμενα της Μπεϋζιανής μεθόδου που αναπτύχθηκε κατά την διάρκεια του 19^{ου} αιώνα.

Ένας πρακτικός τρόπος που δεν εμφανίζει καμία ιδιαίτερη δυσκολία είναι να χρησιμοποιήσουμε την οικογένεια των *conjugate priors* (συζυγών προγενέστερων κατανομών), κατάλληλων για την υπό εξέταση κατάσταση και να διαλέξουμε το ακραίο μέλος αυτής της οικογένειας να αντιπροσωπεύει την άγνοια. Το μέλος αυτό καλείται *non-informative* ή *vague prior* (μη πληροφοριακή ή ασαφής προγενέστερη κατανομή).

Μια περεταίρω υπόθεση είναι ότι ο ακριβής τύπος της προγενέστερης κατανομής είναι σημαντικός μόνο για μικρές βάσεις δεδομένων. Όταν η βάση δεδομένων είναι μεγάλη, η συνάρτηση πιθανοφάνειας συγκεντρώνεται γύρω από την τιμή της μέγιστης πιθανοφάνειας, και το μόνο χαρακτηριστικό της *prior*, το οποίο έχει και μεγάλη επιρροή στην Μπεϋζιανή θεωρία, είναι η συμπεριφορά της γύρω από αυτήν την περιοχή.

Οποιαδήποτε κατανομή *prior* θα παρουσιάζει ομαλότητα στην περιοχή αυτή εκτός κι αν είναι πολύ συγκεντρωμένη σε αυτή περιοχή ή κάπου αλλού. Μια τέτοια *prior* θα οδηγούσε σε μία κατανομή παραπλήσια με την πιθανοφάνεια, και συνεπώς σχεδόν ανεξάρτητη της *prior*. Με άλλα λόγια, όπως ήταν αναμενόμενο, για να καθοριστεί απόλυτα η μεταγενέστερη κατανομή χρειάζονται μεγάλα σύνολα δεδομένων, εκτός κι να ο χρήστης έχει πολύ ισχυρές ενδείξεις από προηγούμενες έρευνες.

Οι Μπεϋζιανές μεθοδολογίες βασισμένες σε κατάλληλες επιλογές μη πληροφοριακών *prior* (Lindley, 1965) συχνά ανταπεξέρχονται με ακρίβεια έναντι των πιο παραδοσιακών μεθόδων, αρκεί να γίνουν οι κατάλληλες αλλαγές στην διατύπωση. Παρόλα αυτά υπάρχουν σημεία στα οποία εμφανίζεται σύγκρουση λόγω διαφορετικών οπτικών.

⁽¹⁾logit function : είναι το αντίστροφο της συνάρτησης logistic. Όταν δηλαδή η παράμετρος της συνάρτησης αντιπροσωπεύει μία πιθανότητα p , τότε η συνάρτηση logit δίνει τον λογάριθμο του όρου $\frac{p}{1-p}$.

Τόσο οι Μπεϋζιανές όσο και οι μη Μπεϋζιανές μέθοδοι κατέχουν το δικό τους σημαντικό κομμάτι στην στατιστική μεθοδολογία . Αν σκοπός του αναλυτή είναι να εκφράσει τον τρόπο με τον οποίο τροποποιείται ένα σύνολο από αρχικές εικασίες λόγω κάποιων ισχυρών ενδείξεων που προέρχονται από τα δεδομένα, τότε η Μπεϋζιανή θεωρία είναι ξεκάθαρα κατάλληλη. Η επίσημη χρήση αυτού του τύπου στατιστικών μεθόδων αποφεύγεται στις εργασιακές πρακτικές των περισσότερων επιστημόνων, αλλά η ανεπίσημη σύνθεση προγενέστερων πεποιθήσεων και η εκτίμηση ενδείξεων από τα δεδομένα είναι σίγουρα κάτι κοινότυπο. Οποιαδήποτε λοιπόν χρήση «ευαίσθητης» στατιστικής πληροφορίας πρέπει να λαμβάνει υπόψη σε μεγάλο βαθμό τις προγενέστερες γνώσεις και τις προηγούμενες εκτιμήσεις σχετικά με την αληθοφάνεια των διαφόρων υποθέσεων. Σε ένα πείραμα στο οποίο το άτομο καλείται να «μαντέψει την κάρτα» με σκοπό να ερευνηθεί η υπεραισθητήρια όραση ,για παράδειγμα ένα υπερβολικό σκορ πρόβλεψης που θα ήταν αρκετά σημαντικό σε επίπεδο 1% θα μπορούσε να αντιμετωπιστεί από τους περισσότερους ανθρώπους με αρκετό σκεπτικισμό. Πολλοί προτιμούν να σκεφτούν ότι η υπερβολή εμφανίστηκε στην τύχη του ατόμου παρά σε κάποιου είδους τηλεπάθεια ή διορατικότητα. Από την άλλη μεριά σε μία κλινική δοκιμή για την σύγκριση ενός ενεργού φαρμάκου με ένα placebo, ένα παρόμοια σημαντικό αποτέλεσμα θα ήταν ευρέως αποδεκτό ως στοιχείο για τον βαθμό επίδρασης του φαρμάκου γιατί τέτοια πορίσματα προκύπτουν συχνά. Η ερώτηση είναι λοιπόν αν θα μπορούσαν στην περίπτωση αυτή να ληφθούν υπόψη οι προγενέστερες πεποιθήσεις και ακόμα καλύτερα αν αυτό θα μπορούσε να γίνει επίσημα διαμέσου μιας Μπεϋζιανής ανάλυσης,ή ανεπίσημα χρησιμοποιώντας μεθόδους frequentist.

Η επίσημη προσέγγιση χρησιμοποιείται κυρίως όταν χρειάζεται να παρθούν αποφάσεις , για παράδειγμα σχετικά με το εάν μια φαρμακευτική εταιρεία θα μπορούσε να εξελιχθεί και να επεκταθεί με την δημιουργία ενός καινούργιου προϊόντος. Εδώ, το στοιχείο, υποκειμενικά και αντικειμενικά , είναι η απόλυτη αποτελεσματικότητα μιας νέας παραγωγής , που χρειάζεται να εκτιμηθεί μαζί με το οικονομικό και άλλα κόστη ώστε να πάρει η εταιρεία εναλλακτικές πορείες δράσεις αν κριθεί απαραίτητο.

Μια άλλη διαφωνία που έχει προκύψει τις τελευταίες δεκαετίες σε σχέση με τις Μπεϋζιανές μεθόδους και η οποία είναι αποτέλεσμα έρευνας σε νέα μοντέλα που αφορούν τις κατασκευές πολύπλοκων συνόλων δεδομένων αναπτύσσεται παρακάτω. Γενικά η Μπεϋζιανή θεωρία οδήγησε σε μία απλοποίηση των υπολογιστικών διαδικασιών κατά τις οποίες οι υπολογισμοί απαιτούν την συνάρτηση πιθανοφάνειας βασισμένη στα παρατηρούμενα δεδομένα, ενώ οι μέθοδοι συχνότητας (frequentist methods) χρησιμοποιούν την περιοχή της ουράς της πιθανότητας και απαιτούν τα

αποτελέσματα να είναι ενσωματωμένα πάνω σε σύνολα δεδομένων και όχι σε πραγματικά παρατηρούμενα δεδομένα. Παρόλα αυτά οι Μπεϋζιανοί υπολογισμοί για πολύπλοκα προβλήματα εσωκλείουν αξιοθαύμαστες υπολογιστικές διεξόδους , που πια είναι διαθέσιμες σε υπολογιστικά πακέτα (Goldestein, 1998) καθώς επίσης και σε εξειδικευμένα πακέτα όπως BUGS (Thomas, 1992).

Στο πεδίο της ιατρικής όπου οι βάσεις δεδομένων εμφανίζονται περισσότερο ξεκάθαρες, ο ερευνητής ίσως δεν έχει δυνατές προγενέστερες πεποιθήσεις να ενσωματώσει μέσα στην ανάλυση , κι έτσι δίνει έμφαση στα στοιχεία που μπορεί να αντλήσει μέσα από τα δεδομένα. Οι στατιστικοί τότε έχουν δύο εναλλακτικές : είτε να χρησιμοποιήσουν μεθόδους συχνότητας , είτε να διατηρήσουν το Μπεϋζιανό πλαίσιο υπολογίζοντας πιθανοφάνειες.

Το τελευταίο μπορεί να παρουσιαστεί απευθείας καθώς , ο ερευνητής ή οι άλλοι εργαζόμενοι καθίστανται ικανοί να ενσωματώσουν οποιαδήποτε προγενέστερη γνώση επιθυμούν ίσως να χρησιμοποιήσουν. Για τέτοιες περιπτώσεις ίσως είναι χρήσιμο να εφαρμόσουμε *ανάλυση ευαισθησίας (sensitivity analysis)* στην οποία μπορούν να διερευνηθούν επιδράσεις διαφορετικών προγενέστερων υποθέσεων.

Στη συνέχεια θα αναφερθούμε σε μεθόδους και σε κάποιες από τις βελτιώσεις που επέτρεψαν στις Μπεϋζιανές μεθόδους να επεκταθούν στην ανάλυση περισσότερο πολύπλοκων καταστάσεων.

Υπόβαθρο

Οποιαδήποτε κατανομή *prior* θα παρουσιάζει ομαλότητα στην περιοχή αυτή εκτός κι αν είναι πολύ συγκεντρωμένη σε αυτή περιοχή ή κάπου αλλού. Μια τέτοια *prior* θα οδηγούσε σε μία κατανομή παραπλήσια με την πιθανοφάνεια., και συνεπώς σχεδόν ανεξάρτητη της *prior*. Με άλλα λόγια , όπως ήταν αναμενόμενο, για να καθοριστεί απόλυτα η μεταγενέστερη κατανομή χρειάζονται μεγάλα σύνολα δεδομένων, εκτός κι να ο χρήστης έχει πολύ ισχυρές ενδείξεις από προηγούμενες έρευνες.

Οι Μπεϋζιανές μεθοδολογίες βασισμένες σε κατάλληλες επιλογές μη πληροφοριακών *prior* (Lindley ,1965) συχνά ανταπεξέρχονται με ακρίβεια έναντι των πιο παραδοσιακών μεθόδων , αρκεί να γίνουν οι κατάλληλες αλλαγές στην διατύπωση. Παρόλα αυτά υπάρχουν σημεία στα οποία εμφανίζεται σύγκρουση λόγω διαφορετικών οπτικών.

Ο ανταγωνισμός μεταξύ των δυο σχολών είναι κατά κύριο λόγο κάτι που αφορά το παρελθόν και προς το παρόν τουλάχιστον, πολλοί στατιστικοί δέχονται ότι και οι δύο έχουν κάτι να προσφέρουν. Κι αυτό θα συνεχίσει να υφίσταται μέχρι την περίπτωση που θα αποκτήσουμε κάποιες πιο εμπειρικές παρά φιλοσοφικές διαφορές.

Αυτό το κεφάλαιο δίνει μια σύντομη και αναγκαία αλλά περιγραφή για κάποιες καινοτομίες οι οποίες έχουν μετασηματίσει την ευελιξία των Μπεϋζιανών αναλύσεων. Περισσότερο αναλυτικές θεωρίες μπορούμε να συναντήσουμε σε πολλά εξαιρετικά κείμενα. Παραδείγματα με έμφαση σε πρακτικά προβλήματα είναι του Gelman(1995), των Gilks (1996) και των Carlin και Louis (2000), ενώ οι Bernardo και Smith (1994) και ο O' Hagan (1994) δίνουν περισσότερο έμφαση σε θεωρητικά θέματα. Εφαρμογές των Μπεϋζιανών μεθόδων σε ιατρικά παραδείγματα μπορούμε να συναντήσουμε για παράδειγμα στο βιβλίο του Breslow (1990) και πιο συγκεκριμένα στο ειδικό θέμα "Statistical Methods in Medical Research" (1996). Επιπλέον, μια αξιοσημείωτη αναφορά μπορούμε να βρούμε και στο βιβλίο του Lindley (2000).

Προγενέστερες και μεταγενέστερες κατανομές (Prior and posterior distributions)

Όπως έχουμε ήδη παρατηρήσει οι Μπεϋζιανές μέθοδοι καθοδηγούνται από την prior και την posterior κατανομή.

Στο κεφάλαιο αυτό είναι χρήσιμο να έχουμε συγκεκριμένους συμβολισμούς για τις παραπάνω κατανομές, ακόμα και αν θα ήταν πιο βολικό να παρεκκλίνουμε λιγάκι από αυτούς όταν αναφερόμαστε σε συγκεκριμένα παραδείγματα. Θα χρησιμοποιήσουμε το θ για να δηλώσουμε μια παράμετρο που, γενικά, θα είναι ένας δείκτης κ- διαστάσεων με μεμονωμένα στοιχεία θ_i , και ένα x που θα δηλώνει τα δεδομένα. Για τις περισσότερες από αυτές τις περιπτώσεις αρκεί να σκεφτούμε τις παραμέτρους και τα δεδομένα που είχαν συνεχείς κατανομές. Οι τροποποιήσεις για τις διακριτές κατανομές εμπλέκουν εναλλασσόμενους ακέραιους που προστίθενται. Η πυκνότητα της prior κατανομής είναι $p(\theta)$ και της posterior $p(\theta/x)$. Ο κατάλληλος συμβολισμός είναι να χρησιμοποιούμε το $p(\theta)$ για την πυκνότητα των δεδομένων, υπό την υπόθεση ότι οι παράμετροι έχουν γνωστές τιμές, οι οποίες προκύπτουν από την πιθανοφάνεια που υπολογίζεται από τα δεδομένα. Σε κάποιες περιπτώσεις θα αναφερθούμε και στην κοινή κατανομή(joint distribution) των παραμέτρων και των δεδομένων, που θα γράφεται ως $p(x,\theta)$.

Μεταγενέστερες κατανομές (Posterior distributions)

Οι Posterior κατανομές είναι το ποσοτικό κλειδί στα Μπεϋζιανά συμπεράσματα. Συνδυάζει τις σχετικές με τα δεδομένα πληροφορίες και με την prior κατανομή, και όταν τα δεδομένα x συλλεχθούν, ό,τι σχετίζεται με το θ θα πρέπει να βασίζεται στο $p(x, \theta)$. Έτσι για παράδειγμα, ο μέσος του θ θα υπολογίζεται από τον τύπο

$$\int_{-\infty}^{+\infty} \theta p(\theta/x) d\theta$$

Για να εκτιμήσουμε το παραπάνω ολοκλήρωμα χρειαζόμαστε και την posterior πυκνότητα, η οποία δίνεται από τον τύπο

$$p(\theta/x) = Cp(x/\theta)p(\theta)$$

όπου το $C = C(x)$ επιλέχθηκε για να εξασφαλίσουμε ότι η posterior είναι πραγματική κατανομή (συντά αναφέρεται ως proper κατανομή) – το οποίο σημαίνει ότι το ολοκλήρωμα πάνω στις πιθανές τιμές του θ ισούται 1. Έτσι,

$$C^{-1} = C(x)^{-1} = \int p(x/\theta)p(\theta)d\theta \quad (1.1)$$

Προσπαθώντας να καθορίσουμε αυτήν την τιμή προκύπτει το σημαντικότερο εμπόδιο στην πρακτική χρήση των Μπεϋζιανών μεθόδων. Το ολοκλήρωμα δεν μπορεί να εκτιμηθεί αναλυτικά εκτός από κάποιες περιπτώσεις, όπως όταν χρησιμοποιούνται οι συζυγείς κατανομές. Οι αριθμητικές μέθοδοι επικεντρώνονται στην εκτίμηση του C όπως έχει προταθεί από τους Naylor και Smith(1982) και μπορεί να αποδειχθούν κατάλληλες αν δεν υπάρχουν τόσες πολλές παράμετροι, αλλά αν το k είναι μεγαλύτερο από περίπου 5, αυτή η προσέγγιση συνήθως δεν δουλεύει. Αναλυτικές προσεγγίσεις βασισμένες στη μέθοδο Laplace για τα ολοκληρώματα έχουν προταθεί από τους Tierney και Kadane (1986). Συνεπώς, οι πρόσφατες μελέτες έχουν επικεντρωθεί σε μεθόδους που επιτρέπουν να περιγράψουμε την $p(x, \theta)$ χωρίς να χρειάζεται να γνωρίσουμε την τιμή του C .

Πράγματι, το πρόβλημα δεν περιορίζεται απλά στην εκτίμηση του C . Ένας χρήσιμος τύπος της posterior κατανομής μπορεί να απαιτεί τον υπολογισμό των posterior στιγμών, συγκεκριμένα των μέσων, των διακυμάνσεων και των συνδιακυμάνσεων. Επίσης, οι τύποι των μεμονωμένων παραμέτρων επικεντρώνονται στην marginal κατανομή, δηλαδή, την κατανομή ενός μόνο στοιχείου του θ . Αν το $\theta = (\theta_i, \theta_{-i})$ όπου θ_{-i} είναι το θ με το θ_i να παραλείπεται, τότε η posterior κατανομή του θ_i είναι:

$$p(\theta_i / x) = \int p(\theta_i, \theta_{-i} / x) d\theta_{-i}$$

Η ακριβής αναλυτική εκτίμηση είναι του παραπάνω ολοκληρώματος πιθανόν να είναι αδύνατη.

Αν και η posterior κατανομή είναι το κλειδί για τα Μπεϋζιανά συμπεράσματα, σε γενικές γραμμές είναι δύσκολα υπολογίσιμη ποσότητα, με αποτέλεσμα να χρειάζονται άλλοι περισσότερο σύντομοι και εύχρηστοι τύποι. Επιπλέον μπορούμε να αντλήσουμε πολλές πληροφορίες από τις γραφικές παραστάσεις της κατανομής κάθε στοιχείου του θ . Σε κάποιες περιπτώσεις είναι χρήσιμο να έχουμε μερικούς από τους τύπους των γραφικών παραστάσεων των κοινών κατανομών για τα ζεύγη των στοιχείων, όπως το διάγραμμα της καμπύλης της διμεταβλητής πυκνότητας. Επιπλέον συχνά είναι απαραίτητοι αριθμητικοί τύποι κάθε παραμέτρου όπως οι posterior μέσοι τους. Αν η posterior έχει μια απλή κλίμακα και είναι επαρκώς συμμετρική ως προς αυτή ώστε δικαιολογημένα να είναι κοντά στην κανονική κατανομή, τότε ίσως είναι βοηθητικό να γνωρίζουμε τις posterior διακυμάνσεις και συνδιακυμάνσεις του θ . Οι ποσότητες της posterior, όπως η διάμεσος και τα τεταρτημόρια και ίσως το 5^ο και 95^ο εκατοστημόριο είναι επίσης σημαντικά.

Με τον καιρό το ενδιαφέρον ίσως συγκεντρώνεται στην συνάρτηση των στοιχείων του x . Ένα παράδειγμα είναι η αναλογία θ_1 / θ_2 . Σε μια τοποθέτηση frequentist όμως κάτι τέτοιο θα ήταν λιγάκι προβληματικό, με συμπεράσματα να έχουν βγει προσεγγιστικά, όπως για παράδειγμα η μέθοδος delta. Η Μπεϋζιανή προσέγγιση όμως, είναι θεωρητικά, περισσότερο ικανοποιητική. Οι παράμετροι στο θ μπορεί να τροποποιηθούν έτσι ώστε ο ένας από αυτούς να είναι η αναλογία, ή οποιαδήποτε άλλη παράμετρος μας ενδιαφέρει, έτσι ώστε να μπορούμε να βγάλουμε συμπεράσματα χρησιμοποιώντας την κατάλληλη posterior κατανομή. Οι εκτιμήσεις διαστημάτων

παίζουν ένα σημαντικό ρόλο στην Μπεϋζιανή ανάλυση: μια $100\alpha\%$ εκτίμηση του διαστήματος, για παράδειγμα του θ_1 είναι οποιοδήποτε ζευγάρι $(\theta_{1L}, \theta_{1U})$ όπου

$$\alpha = \int_{\theta_{1L}}^{\theta_{1U}} p(\theta_1, \theta_{-1} / x), d\theta_1 d\theta_{-1} \quad (1.2)$$

Συνεπώς, υπάρχει μια πιθανότητα α το θ_1 να βρίσκεται μεταξύ των θ_{1L} και θ_{1U} - ένας πολύ πιο φυσικός ορισμός από αυτόν για το διάστημα εμπιστοσύνης frequentist.

Η εξίσωση (1.2) δεν καθορίζει ένα μοναδικό διάστημα και συνεπώς στην Μπεϋζιανή ανάλυση συναντάμε διαφορετικούς τύπους για το διάστημα. Μια διαισθητική επιλογή που εμφανίζεται για την εκτιμήτρια του διαστήματος είναι η μεγαλύτερη posterior πυκνότητα (highest posterior density-HPD) ή το πιο αξιόπιστο διάστημα (most credible interval). Αυτό περιλαμβάνει ένα σύνολο τιμών του θ_1 που έχουν πιθανότητα α κάτω από την posterior κατανομή και τέτοια ώστε η posterior πυκνότητα να είναι μεγαλύτερη σε οποιοδήποτε σημείο εντός του συνόλου σε σύγκριση με οποιοδήποτε σημείο εκτός αυτού. Αυτό οδηγεί στην αποκαλυπτική ιδιότητα ότι το ολικό μέγεθος του συνόλου είναι πιο μικρό από οποιοδήποτε άλλου συνόλου που έχει την ίδια posterior πιθανότητα. Συνεπώς, υπάρχουν δυσκολίες με τέτοιου τύπου διαστήματα. Δηλαδή:

1. Δεν είναι αμετάβλητα κάτω από οποιονδήποτε μετασχηματισμό. Έτσι, για παράδειγμα, το HPD για το $\log \theta_1$ ίσως δεν είναι ξεκάθαρα συσχετισμένο με αυτό του θ_1 .
2. Υπάρχει δυσκολία στον υπολογισμό, και
3. αν η posterior δεν είναι κατά μοναδικό τρόπο τροποποιημένη το σύνολο ίσως δεν είναι ένα διάστημα του τύπου $(\theta_{1L}, \theta_{1U})$.

Μπροστά στις δυσκολίες αυτές, είναι κοινώς αποδεκτό να κλίνουμε προς την χρήση των διαστημάτων ίσης ουράς (equal tail interval), τα οποία καθορίζονται από την απαίτηση ότι η πιθανότητα που βρίσκεται κάτω από το θ_{1L} είναι ίδια με αυτήν που βρίσκεται πάνω από το θ_{1U} . Για τις posterior κατανομές που είναι μονότροπες και προσεγγιστικά συμμετρικές, τα διαστήματα HPD και ίσης ουράς είναι πιθανό να είναι παρόμοια. Αν η posterior είναι εμφανώς αλλοιωμένη τότε τα δύο είδη του διαστήματος μπορεί να είναι αρκετά διαφορετικά.

Είναι πιθανό να υπολογίζουμε αναλυτικά τους τύπους του $p(x, \theta)$ που αναφέραμε παραπάνω, αλλά στις περισσότερες εφαρμογές δεν θα γίνεται κάτι τέτοιο. Μερικές από τις πιο πρόσφατες βελτιωμένες τεχνικές για τον καθορισμό της posterior κατανομής επιτυγχάνεται με το να επιτρέψουμε σε έναν αναλυτή να πάρει δείγμα από την posterior κατανομή. Αυτό σημαίνει ότι ο αναλυτής μπορεί να διαλέξει να παράγει οποιοδήποτε πλήθος κατανομής μιας μεταβλητής π.χ. $\theta_1, \theta_2, \dots, \theta_N$, καθεμιά από τις οποίες προσεγγίζεται από μια κατανομή με πυκνότητα $p(x, \theta)$. Σ' αυτές τις περιπτώσεις οι τύποι βρίσκονται γενικά από υπολογιστικές κατάλληλες δειγματικές αναλογίες.

Για παράδειγμα, μια εκτίμηση του posterior μέσου θα είναι,

$$\frac{1}{N} \sum_{l=1}^N \theta_l$$

Ενώ αυτός ο τρόπος της συντομογραφίας μιας κατανομής ίσως φαίνεται κάτι ανακριβές, η προσέγγιση είναι πολύ ευέλικτη. Για παράδειγμα, η πιθανότητα ότι ένα στοιχείο, ας πούμε, θ_1 , είναι μικρότερο από το c εκτιμάται από την αναλογία των τιμών θ_{l_i} , οι οποίες είναι μικρότερες του c . Σε κάποιες περιπτώσεις είναι χρήσιμο επομένως αντί για την προαναφερθείσα συντομογραφία να γράφουμε,

$$\frac{1}{N} \sum_{l=1}^N I(\theta_{l_i} < c)$$

όπου το $I(\theta < c)$ ισούται με 1 αν $\theta < c$ και με 0 διαφορετικά. Περισσότερο πολύπλοκες συντομογραφίες μπορούν να βρεθούν επίσης εύκολα. Για παράδειγμα μια εκτίμηση του $P(\theta_1 < c_1 / \theta_2 < c_2)$ υπολογίζεται σαν μια αναλογία εκείνων των μεταβλητών που έχουν $\theta_1 < c_1$ δεδομένου ότι $\theta_2 < c_2$. Συντομογραφίες που σχετίζονται με τις posterior ιδιότητες των συναρτήσεων των στοιχείων του θ , όπως η αναλογία $\frac{\theta_1}{\theta_2}$, μπορούν επίσης να

γραφούν με τον προφανή τρόπο. Θεωρητικά, η Στατιστική μπορεί να ελέγξει την ακρίβεια των συντομογραφιών με την κατάλληλη επιλογή του N , αν και για πολύπλοκα μοντέλα το υπολογιστικό φορτίο που επιβάλλεται με την επιλογή μεγάλων τιμών μπορεί να είναι σημαντικό.

Γραφικά οι κατανομές των μεμονωμένων στοιχείων του θ μπορεί να είναι απλά ιστογράμματα, αν και η εκτίμηση της πυκνότητας μέσω Kernels θα παράγει πιο ομαλές εκτιμήσεις (βλέπε για παράδειγμα Silverman, 1986). Συνεπώς, μπορεί να είναι περισσότερο κατάλληλες ειδικές τεχνικές σε βέβαιες περιστάσεις (Gelfand και Smith, 1990). Για γραφικές αναπαραστάσεις δυο διαστάσεων, υπάρχουν διαθέσιμοι σε πολλά στατιστικά πακέτα αλγόριθμοι για εκτίμηση διαγραμμάτων με καμπύλες που βασίζονται στα δεδομένα.

Οι τιμές που προσεγγίζονται από τα $p(x, \theta)$, και συμβολίζονται με $\theta_1, \theta_2, \dots, \theta_N$ είναι απαραίτητα κατανεμημένες με τον ίδιο τρόπο αλλά όχι υποχρεωτικά και ανεξάρτητες: το αν είναι ή όχι είναι ανεξάρτητες βασίζεται στον λογάριθμο που χρησιμοποιήθηκε. Πολλές από τις κοινές συντομεύσεις είναι ανεπηρέαστες από την εξαρτημένη κατασκευή.

Για παράδειγμα, οι προσδοκίες των

$$\frac{1}{N} \sum_{l=1}^N \theta_l \text{ και } \frac{1}{N} \sum_{l=1}^N I(\theta_l < c)$$

είναι αντίστοιχα, ο posterior μέσος και η posterior πιθανότητα όπου $\theta_1 < c_1$ παρά την παρουσία της συσχέτισης μεταξύ των $\theta_1, \theta_2, \dots, \theta_N$.

Ο φυσικός εκτιμητής της posterior διακύμανσης του θ_1 για παράδειγμα είναι

$$\frac{1}{N-1} \sum_{l=1}^N (\theta_{1l} - \bar{\theta}_1)^2 = \frac{1}{N-1} \left[\sum_{l=1}^N \theta_{1l}^2 - N\bar{\theta}_1^2 \right]$$

Θεωρώντας ότι οι προσομοιωτές είναι ανεξάρτητοι η προσδοκία του εκτιμητή αυτού είναι η posterior διακύμανση να είναι σ^2 . Γενικά η προσδοκία είναι $\sigma^2(1-\bar{p})$ όπου το \bar{p} είναι ο μέσος των $\frac{1}{2} N(N-1)$ συσχετίσεων μεταξύ των $\theta_{11}, \theta_{12}, \dots, \theta_{1N}$. Ακόμα κι όταν οι προσομοιωτές από την posterior είναι εξαρτημένοι, η μέση συσχέτιση είναι συχνά επαρκώς μικρή για τον συσχετιζόμενο παράγοντα $1-\bar{p}$ ώστε να αγνοηθεί. Αν και δεν είναι πάντα αυτό που μας αφορά. Αν ο αναλυτής επιθυμεί να ξέρει πόσο καλός είναι ο μέσος του δείγματος του θ_1 ως εκτιμητής του posterior μέσου, τότε το σύνθημα μέτρο

αυτού, δηλαδή το σταθερό σφάλμα του μέσου, μπορεί να είναι αισθητά επηρεασμένο από την συσχέτιση μεταξύ διαδοχικών τιμών. Η διακύμανση του

$$\frac{1}{N} \sum_{i=1}^N \theta_i$$

είναι

$$\frac{\sigma^2}{N} [1 + (N-1) \bar{\rho}]$$

και ο διορθωτικός παράγοντας $1 + (N-1) \bar{\rho}$ μπορεί να αποκλίνει ουσιαστικά από το 1, όπου σ' αυτήν την περίπτωση χρειάζεται να εφαρμοστεί η συσχέτιση.

Προγενέστερες Κατανομές (Prior distributions)

Μια Μπεϋζιανή ανάλυση ξεκινά από την prior κατανομή, $p(\theta)$, η οποία είναι μια αριθμητική περιγραφή των πεποιθήσεων που ο αναλυτής έχει για τις παραμέτρους του μοντέλου $p(x/\theta)$ πριν οποιοδήποτε δεδομένο x συλλεχθεί. Κατά συνέπεια, μια prior κατανομή πρέπει να καθορίζεται προτού διεκπεραιωθεί οποιαδήποτε Μπεϋζιανή ανάλυση.

Πριν συζητήσουμε τους διάφορους τύπους της prior κατανομής που χρειαζόμαστε, είναι χρήσιμο να κάνουμε μερικά γενικά σχόλια.

1) Αφού ισχύει η ισότητα $p(x, \theta) = Cp(x/\theta)p(\theta)$, τότε οποιαδήποτε τιμή του θ η οποία έχει μηδενική prior πιθανότητα θα έχει απαραίτητα μηδενική και την posterior πιθανότητα. Επομένως είναι απαραίτητο όταν καθορίσουμε την prior κατανομή να μην είναι τόσο ρυθμιστική σχετικά με αδύνατες τιμές παραμέτρων. Πρακτικά, μηδενικές τιμές για την prior κατανομή εμφανίζονται συχνά εξαιτίας του επιτρεπόμενου εύρους που επιβάλλεται σε βέβαιες παραμέτρους. Αν είναι απαραίτητα τέτοια εύρη, είναι σημαντικό να είμαστε προσεκτικοί σχετικά με το ότι δεν είναι τόσο αυστηρά καθορισμένα.

2) Εννοείται ότι η prior κατανομή πρέπει να συγκεντρώνει τις πεποιθήσεις του αναλυτή σχετικά με το θ προτού συλλεχθούν τα δεδομένα. Φυσικά, κάποιος αξιολογώντας την ανάλυση ίσως να είχαν διαφορετική άποψη, η οποία θα εκφραζόταν με διαφορετική prior. Είναι αμφιλεγόμενο επίσης το αν θα ήταν σημαντικό για τα συμπεράσματα των αναλύσεων να αναφερθεί με ποιον τρόπο κάθε αναγνώστης θα μπορούσε να καταλάβει τον τρόπο με τον οποίο κάθε prior επηρέασε την συλλογή των δεδομένων. Ίσως αν κάτι τέτοιο είναι επιθυμητό, τότε υπάρχουν πρακτικοί περιορισμοί που εμποδίζουν την επέκταση αυτού του θέματος. Παρόλα αυτά, είναι σημαντικό το γεγονός ότι η επίδραση που έχει η prior κατανομή στην ανάλυση είναι ολοφάνερη. Ευαίσθητες αναλύσεις στις οποίες συγκρίνονται τα αποτελέσματα των posterior κατανομών με διάφορες prior κατανομές, φαίνεται να είναι απαραίτητο βοήθημα για οποιαδήποτε Μπεϋζιανή ανάλυση.

3) Η posterior κατανομή θα είναι πρακτικά πολύ πιο ανεξάρτητη πάνω στην prior κατανομή, εκτός κι αν το δείγμα είναι μικρό ή η prior κατανομή αρκετά συμπυκνωμένη σε σχέση με κάποιες τιμές. Επομένως, το αντίστροφο είναι ότι αν το σύνολο των δεδομένων είναι μικρό, τότε ο αναλυτής ίσως χρειαστεί να κάνει ξεκάθαρο ότι τα posterior αποτελέσματα σε μεγάλο βαθμό αντανακλούν τις προγενέστερες υποθέσεις, πάρα οποιαδήποτε πληροφορία επιτεύχθηκε από τα δεδομένα.

Η prior κατανομή μπορεί από μόνη της να βασίζεται σε παραμέτρους, η , για τις περιπτώσεις που αυτές είναι προκαθορισμένες, και τότε είναι πιο συνηθισμένο να γράφεται $p(\theta/n)$. Για να διαχωρίσουμε αυτές τις παραμέτρους από εκείνες που περιέχονται ξεκάθαρα στο προκαθορισμένο μοντέλο, οι η συχνά αναφέρονται ως υπερπαραμέτροι (hyperparameters). Οι τιμές των υπερπαραμέτρων μπορούν να επιλεγούν από τον αναλυτή έτσι ώστε το $p(\theta/n)$ να περιγράφει επαρκώς τις prior πεποιθήσεις. Η κατανομή των δεδομένων θα εξαρτάται από το η γιατί

$$p(x/\eta) = \int p(x/\theta)p(\theta/\eta)d\theta$$

Μια προσέγγιση για την επιλογή του η είναι να βρεθούν τιμές για τις οποίες η $p(x/\eta)$ ταιριάζει καλύτερα με τα δεδομένα. Αυτή είναι μία εμπειρική προσέγγιση και υπάρχουν δυο κριτικές γι' αυτήν, μια πρακτική και μια φιλοσοφική. Το πρακτικό πρόβλημα είναι ότι υπάρχει αβεβαιότητα στην εκτίμηση του η αλλά χρησιμοποιείται σαν να είναι ακριβής, έτσι με τον τρόπο αυτό υπερεκτιμάται η ακρίβεια οποιασδήποτε επακόλουθης ανάλυσης. Τυπικά, οι εκτιμήτριες διαστημάτων θα είναι πολύ περιορισμένες.

Το φιλοσοφικό πρόβλημα είναι αν η prior κατανομή πρέπει να προσδιοριστεί πριν τα δεδομένα, ακόμη κι αν οι τύποι των δεδομένων είναι γνωστοί και αντικατοπτρίζουν τις prior πεποιθήσεις που αφορούν το θ .

Μια ολοκληρωμένη Μπεϋζιανή προσέγγιση στο πρόβλημα της αβεβαιότητας των υπερπαραμέτρων είναι να τοποθετηθεί η κατανομή πάνω σ' αυτά – hyperprior distribution. Με αυτό ασχολήθηκε, για παράδειγμα, ο Gelman (1995). Στο κεφάλαιο αυτό θα αποφύγουμε το επίπεδο της πολυπλοκότητας και θα υποθέσουμε, όπου χρειάζεται, ότι οι υπερπαραμέτροι δίνουν εκ των προτέρων τιμές που καθορίζουν την προγενέστερη αβεβαιότητα.

Πρακτικά, οι prior κατανομές παίρνουν έναν εκ των δυο τύπων, είτε τον ενημερωμένο είτε τον μη-ενημερωμένο (informative ή non-informative form τον οποίο μερικές φορές θα συναντήσουμε να αναφέρεται ως vague). Στην προηγούμενη περίπτωση ο αναλυτής κάνει μια προσεκτική προσπάθεια να περιγράψει αριθμητικά τις prior πεποιθήσεις. Από την άλλη πλευρά, αν χρησιμοποιηθεί μια non-informative prior, τότε θα δεν υπάρξει καμία τέτοια πρόθεση. Μπορεί να υπάρχουν ποικίλοι τρόποι να χρησιμοποιήσουμε μια vague prior. Ίσως να μην υπάρχει καμία prior πληροφορία. Εναλλακτικά μπορεί να υπάρχει η επιθυμία να χρησιμοποιήσουμε μία αντικειμενική prior, υπό την έννοια ότι τα αντίστοιχα posterior συμπεράσματα ίσως πρέπει να βασιστούν εξ' ολοκλήρου στα δεδομένα.

1.2 Το Μπεϋζιανό γραμμικό μοντέλο (Bayesian Linear Model)

Η διακύμανση που συναντάμε σε πολλές μεταβλητές στις ιατρικές έρευνες μπορεί να μοντελοποιηθεί από μια κανονική κατανομή. Αν το συμπέρασμα ακολουθεί μια κανονική κατανομή με μέσο μ και διακύμανση σ^2 , και αν υποθέσουμε ότι το μ ακολουθεί μια κανονική prior κατανομή με μέσο μ_0 και διακύμανση σ_0^2 , τότε η ανάλυση θα δείξει ότι η posterior κατανομή για το μ είναι επίσης κανονική. Ο posterior μέσος είναι ένας σταθμισμένος μέσος του δειγματικού μέσου και του prior μέσου.

Γραμμικό μοντέλο με γνωστή διακύμανση

Υπάρχει μια σημαντική συμφωνία που επιτεύχθηκε από την επέκταση του απλού κανονικού μοντέλου με τέτοιο τρόπο ώστε ο μέσος να μπορεί να ποικίλει ανάλογα με τις συνδιακυμάνσεις, όπως στην πολλαπλή παλινδρόμηση και στην ανάλυση της διακύμανσης. Μια Μπεϋζιανή ανάλυση, που είναι σχεδόν ανάλογη με αυτό για την απλή περίπτωση, πιθανόν να μπορεί να χρησιμοποιηθεί για την επεκταμένη περίπτωση.

Το σύνηθες γραμμικό μοντέλο $y = X\beta + \varepsilon$ μπορεί να γραφεί με βάση το $n \times 1$ διάνυσμα του συμπεράσματος y ως εξής

$$y / \beta \sim N(X\beta, \sigma^2 I_n)$$

που σημαίνει ότι, δίνοντας μια τιμή για το β , ο δείκτης στο αποτέλεσμα θα έχει μέσο $X\beta$, και τα εξαγόμενα συμπεράσματα θα είναι ασυσχέτιστα, και καθένα από αυτά θα έχει σταθερή διακύμανση σ^2 . Η υπόθεση που ακολουθεί δηλαδή ότι η διασπορά του y με γνωστό το β έχει αυτόν τον συγκεκριμένο τύπο, γίνεται γιατί ο τύπος αυτός χρησιμοποιείται ευρέως και επεξηγεί επαρκώς την Μπεϋζιανή προσέγγιση του γραμμικού μοντέλου. Παρόλα αυτά, αν δεν είναι κατάλληλος, μπορεί εύκολα να βελτιωθεί ή να τροποποιηθεί.

Ο μέσος τώρα καθορίζεται από έναν δείκτη β και για την ποσότητα αυτή απαιτείται μια πολυμεταβλητή prior κατανομή. Μια φυσική επιλογή, η οποία τείνει στην πιθανοφάνεια, είναι μια πολυμεταβλητή κανονική κατανομή με μέσο β_0 και σύμπλεγμα διασποράς D . Η posterior κατανομή του β είναι μια πολυμεταβλητή κανονική με διασπορά M^{-1} και μέσο

$$M^{-1}(\sigma^{-2}X^T y + D^{-1}\beta_0) \quad (1.3)$$

όπου
$$M = \sigma^{-2}X^T X + D^{-1} \quad (1.4)$$

Παρατηρούμε ότι αυτοί οι τύποι είναι πολυμεταβλητές αναλογίες παραλλαγής της μονομεταβλητής, όπου ο posterior μέσος προέρχεται από ένα άθροισμα σταθμισμένο με ακρίβεια. Ο όρος $\sigma^{-2}X^T X$ στο (1.4), όντας το αντίστροφο του $s^2(X^T X)^{-1}$, μπορεί να

θεωρηθεί ότι είναι η ακριβής πολυμεταβλητή της παραμέτρου β , και ο δεύτερος όρος η ακριβής prior. Η αναλογία του δειγματικού μέσου είναι $\hat{\beta}$, καθορισμένη από τον τύπο

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

Και από αυτόν συνεπάγεται ότι $X^T y = (X^T X)\hat{\beta}$. Αντικαθιστώντας την τελευταία σχέση στον τύπο (1.3) αποδεικνύεται ότι ο posterior μέσος έχει ξεκάθαρα τον τύπο του σταθμισμένου αθροίσματος ανάλογο με την περίπτωση της μονομεταβλητής.

Η κατάσταση για την οποία δεν υπάρχει καμία προηγούμενη πληροφορία διαθέσιμη θα μπορούσε να μοντελοποιηθεί χρησιμοποιώντας μια εσφαλμένη prior με $D^{-1} = 0$. Κατά συνέπεια, από τους τύπους (1.3) και (1.4), ο posterior μέσος και η διακύμανση υπολογίζονται αντίστοιχα από τις ποσότητες

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad \text{και} \quad (y - X\hat{\beta})^T (y - X\hat{\beta}) = y^T y - \hat{\beta}^T X^T y$$

τις οποίες έχουμε συναντήσει ξανά στις αναλύσεις frequentist. Φυσικά, καθώς το β θεωρείται τώρα μία τυχαία μεταβλητή, θα υπάρξουν λεπτές διαφορές στον τρόπο που θα εκφραστεί η αβεβαιότητα κατά την εκτίμηση των συντελεστών παλινδρόμησης.

Για τον λόγο αυτό η απλούστερη παραλλαγή μονομεταβλητής β , από μόνη της, δεν είναι χρήσιμη επειδή στην Μπεϋζιανή ανάλυση υποθέτουμε ότι η σ^2 είναι γνωστή και δεν έχει θεωρηθεί σαν μια τυχαία ποσότητα. Το πώς θα συμπεριλάβουμε αυτήν την παράμετρο θα συζητηθεί, αλλά, όπως θα δούμε, οι ρεαλιστικές προσεγγίσεις αυξάνουν και την πολυπλοκότητα.

Συζυγείς και συσχετισμένες αναλύσεις των δειγμάτων που ακολουθούν κανονική κατανομή με άγνωστη διακύμανση

Όταν για κάποιο πρόβλημα αναζητούνται συζυγείς πληροφοριακές priors δημιουργούνται μερικές δυσκολίες, και αυτό είναι κάτι με το οποίο θα ασχοληθούμε παρακάτω. Παρόλα αυτά, μπορούμε να πάρουμε μια ιδέα για τον τύπο της κατανομής που χρειαζόμαστε κι από την υπόθεση όπου θεωρούμε μια ακατάλληλη μονομεταβλητή prior. Η μονομεταβλητή prior για (μ, σ^{-2}) είναι $p(\mu, \sigma^{-2}) = \frac{1}{\sigma^2}$ και από αυτή συνεπάγεται ότι η posterior είναι

$$p(\mu, \sigma^2 / y) = \sigma^{-n-2} \exp \left\{ -\frac{1}{2\sigma^2} [(n-1)s^2 + n(\bar{y} - \mu)^2] \right\} \quad (1.5)$$

όπου y είναι ο μέσος όρος ενός δείγματος μεγέθους n που έχει διακύμανση σ^2 . Από τον τύπο (1.5) μπορούν να προκύψουν ποικίλα συμπεράσματα. Η posterior κατανομή μόνο για το μ (δηλαδή η marginal posterior distribution) προκύπτει από το ενσωματωμένο $p(\mu, \sigma^2 / y)$ με εκτίμηση του σ^2 και δίνεται από τον τύπο

$$p(\mu / y) = \left[1 + \frac{n(\mu - \bar{y})^2}{(n-1)s^2} \right]^{-n/2}$$

Αυτός είναι ένας τύπος t κατανομής ο οποίος εσωκλείει μια παράμετρο τοποθεσίας και μία κλίμακας, καθώς επίσης και τους συνηθισμένους βαθμούς ελευθερίας. Λεπτομέρειες μπορούμε να βρούμε στον Πίνακα Α του Gelman (1995), αλλά μια απλούστερη περιγραφή είναι ότι η οριακή posterior distribution του μέσου είναι η ίδια όπως κι αυτή του $\bar{y} + Ts/Tn$, όπου το T έχει μια σταθερή t κατανομή με $n-1$ βαθμούς ελευθερίας.

Η οριακή posterior (marginal posterior) για την σ^2 βρίσκεται με παρόμοιο τρόπο από τον τύπο (1.5).

Το ολοκλήρωμα που εκτιμάει το μ υποχρεωτικά είναι το ολοκλήρωμα μιας κανονικής πυκνότητας και αυτό οδηγεί έμμεσα στον τύπο

$$p(\sigma^2 / y) = (\sigma^2)^{-\frac{n+1}{2}} \exp \left[-\frac{(n-1)s^2}{2\sigma^2} \right]$$

Αυτός ο τύπος μοιάζει με την πυκνότητα της Γάμμα ή χ^2 κατανομής και στην πραγματικότητα είναι σχεδόν συσχετισμένος με αυτές:

Είναι η πυκνότητα scaled inverse- χ^2 με scale factor s^2 και βαθμούς ελευθερίας $n-1$. Η κατανομή scaled inverse- χ^2 είναι μια σημαντική κατανομή για να περιγράψουμε την διακύμανση στην Μπεϋζιανή ανάλυση της κανονικής κατανομής. Γενικά, αν η X είναι μια τυχαία μεταβλητή που ακολουθεί μια χ^2 κατανομή με ν βαθμούς ελευθερίας, τότε η

s^2v/X έχει μια αντίστροφη κλίμακα- X^2 κατανομής με v βαθμούς ελευθερίας και παράγοντα κλίμακας s^2 . Μιλώντας προσεγγιστικά, όσο το x έχει μέσο v , τότε μια κατανομή scaled inverse- X^2 θα αναμενόταν να έχει τιμές τοποθετημένες γύρω από τον scale factor s^2 . Μιλώντας με μεγαλύτερη ακρίβεια, ο μέσος είναι $vs^2/(v-2)$ και η διακύμανση είναι $2/(v-4)$ φορές το τετράγωνο του μέσου. Η καμπύλη της κατανομής είναι θετικά παραμορφωμένη με μια έντονη ουρά.

Η ανάλυση που βασίζεται στην εσφαλμένη prior $1/\sigma^2$ οδηγεί σε αποτελέσματα τα οποία είναι πολύ οικεία λόγω της frequentist ανάλυσης, με το $(\bar{y}-\mu)\sqrt{n}/s$ να ακολουθεί μια t κατανομή με $n-1$ βαθμούς ελευθερίας και το s^2/σ^2 να ακολουθεί μια κατανομή- X^2 με τους ίδιους βαθμούς ελευθερίας. Επιπλέον, αν οι βαθμοί ελευθερίας, $n-1$ δεν υπερβαίνουν το 4, η κατανομή marginal posterior κατανομή για το σ^2 έχει άπειρη διακύμανση ενώ για βαθμούς ελευθερίας ≤ 2 έχει άπειρο μέσο. Αυτό δείχνει ξεκάθαρα ότι για μικρή βάση δεδομένων, οι εσφαλμένες priors μπορεί να οδηγήσουν σε posteriors οι οποίες στερούνται μερικά γνωρίσματα που οι στατιστικοί θεωρούν σημαντικά για την περιγραφή των δεδομένων.

Επεκτείνοντας την ανάλυση σε μια πληροφοριακή και συζυγή prior απαιτείται ο ορισμός μιας πολυσύνθετης κατανομής, της *normal inverse- X^2* (κανονική αντίστροφη- X^2 κατανομή, *NIC*). Η prior για την σ^2 είναι μια inverse- X^2 με ν_0 βαθμούς ελευθερίας και factor scale σ^2 . Η prior για το μ είναι κανονική κατανομή με μέσο μ_0 και διακύμανση σ^2/m_0 . Επισημαίνουμε ότι η κατανομή prior για το μ βασίζεται στο σ^2 , με άλλα λόγια, η κοινή prior κατανομή $p(\mu, \sigma^2)$ καθορίζεται λαμβάνοντας υπόψη την marginal prior κατανομή για το σ^2 και την conditional κατανομή του μ με δεδομένο σ^2 , δηλαδή $p(\mu, \sigma^2) = p(\sigma^2)p(\mu/\sigma^2)$. Οι τέσσερις παράμετροι της *NIC* κατανομής είναι:

- (i) οι βαθμοί ελευθερίας, ν_0 , που καθορίζουν τη διασπορά του σ^2 ανάλογα με την θέση του, με τις μικρότερες τιμές να αντιστοιχούν στις μεγαλύτερες διασπορές
- (ii) ο factor scale σ_0^2 που κυρίως υπολογίζει την θέση του σ^2
- (iii) το μ_0 , που είναι η prior-προσδοκία για τον μέσο και
- (iv) το ισχύων δειγματικό μέγεθος, m_0 .

Η *NIC* είναι η συζευγμένη κατανομή για μια κανονική μεταβλητή. Αν ένα δείγμα μεγέθους n έχει μέσο \bar{y} και διακύμανση s^2 , τότε η posterior κατανομή $p(\mu, \sigma^2/y)$ είναι επίσης μια *NIC* κατανομή και οι τροποποιημένες τιμές των τεσσάρων παραμέτρων είναι:

(1) Νέοι βαθμοί ελευθερίας ίσοι με $n \pm V_o$.

(2) Νέοι factor scale ίσοι με

$$(n + v_o)^{-1} \left[v_o \sigma_o^2 + (n-1)s^2 + \frac{m_o n}{m_o + n} (\bar{y} - \mu_o)^2 \right]$$

(3) Νέες προσδοκίες για τον μέσο ίσες με

$$\mu = \frac{m_o}{m_o + n} \mu_o + \frac{n}{m_o + n} \bar{y}$$

(4) Νέα δειγματικά μεγέθη ίσα με $n + m_o$.

Η κατανομή marginal posterior, για το σ^2 είναι μια scaled inverse- X^2 με βαθμούς ελευθερίας και factor scale που δίνονται από (i) και (ii) αντίστοιχα. Η κατανομή marginal posterior για το μ είναι πάλι ίδια με αυτήν της τυχαίας μεταβλητής $\mu n + kT$, όπου το μ_n δίνεται από το (iii), το k είναι ο δείκτης του (ii) και (iv) και το T ακολουθεί μία σταθερή t κατανομή με βαθμούς ελευθερίας δοσμένους από το (i).

Επιπλέον διατίθεται μια παρόμοια κατανομή για την παλινδρόμηση όπου ο μέσος δίνεται από το $X\beta$. Η παλινδρόμηση για τη διακύμανση του υπόλοιπου σ^2 είναι ακριβώς η ίδια όπως από πάνω, ενώ η κατανομή για το β το οποίο εξαρτάται από το σ^2 είναι μια πολυμεταβλητή κανονική κατανομή. Η ενημέρωση των παραμέτρων από την οπτική σκοπιά των δεδομένων ακολουθείται από εκφράσεις ανάλογες με αυτές στο (i) έως (iv), αλλά με τη χρήση των κατάλληλων δεικτών και όρων.

Αν και η ευκολία αυτής της προσέγγισης την καθιστά ιδιαίτερα ελκυστική, έχει ένα μειονέκτημα. Βασισμένη στο σ^2 , η prior διακύμανση του μ είναι σ^2 / m_o , έτσι η prior αβεβαιότητα στο μ είναι μεγαλύτερη της μεγαλύτερης τιμής του σ^2 . Σε πολλές περιπτώσεις, ιδιαίτερα όταν υπάρχει μικρή prior πληροφορία σχετικά με κάθε παράμετρο, αυτός ο συσχετισμός μεταξύ τους, ίσως είναι μικρής σημασίας (πραγματικά, στο όριο που η m_o και V_o τείνει να είναι 0, το οποίο δίνει μια μη πληροφοριακή prior $1/\sigma^2$, οι δύο παράμετροι γίνονται ανεξάρτητοι).

Παρόλα αυτά, είναι πιθανό να υπάρχει μια χρήσιμη prior πληροφορία σχετικά με τον μέσο αλλά επίσης να υπάρχει και μια σημαντική prior αβεβαιότητα σχετικά με την τιμή της διακύμανσης. Στην περίπτωση αυτή η προηγούμενη εξάρτηση μεταξύ των παραμέτρων φαίνεται λιγότερο ικανοποιητική.

Ένας φανερός τρόπος να παρακάμψουμε το πρόβλημα αυτό είναι να πάρουμε την prior για το μ ως κανονική κατανομή με μέσο μ_0 και prior διακύμανση ω_0^{-2} , η οποία τώρα δεν θα βασίζεται στο σ^2 . Η prior για το σ^2 είναι μια κατανομή scaled inverse- χ^2 με παραμέτρους ν_0 και σ^2 , ανεξάρτητες του μ . Δυστυχώς, αυτή η prior δεν είναι πια συζευγμένη με την πιθανοφάνεια και η posterior κατανομή δεν θα έχει σταθερό τύπο.

Η posterior κατανομή του μ εξαρτημένη από το σ^2 είναι απλά η κανονική κατανομή, με την τιμή του σ^2 να θεωρείται γνωστή. Αυτό είναι ένα παράδειγμα της δεσμευμένης κατανομής υπό προϋποθέσεις και μπορεί να χρησιμοποιηθεί για να απλουστεύσει σημαντικά αναλύσεις που χρησιμοποιούν αυτό το μοντέλο. Το μοντέλο μπορεί να επεκταθεί στην παλινδρόμηση, με την prior στο β να είναι ο πολυμεταβλητός κανονικός τύπος που περιγράφηκε παραπάνω. Η ανεξάρτητη prior για το σ^2 είναι πάλι μια κατανομή scaled inverse- χ^2 με υπερπαραμέτρους ν_0 και σ^2 .

Διατίθενται διάφορες προσεγγίσεις για την ανάλυση αυτού του μοντέλου, δυο εκ των οποίων θα περιγραφούν εν συντομία παρακάτω. Καθώς η posterior κατανομή δεν έχει ένα σταθερό τύπο, δεν μπορούμε να έχουμε αναλυτικές περιγραφές των ιδιοτήτων. Και οι δύο προσεγγίσεις παρέχουν περιγραφές της posterior με το να επιτρέπουν στον αναλυτή να παριστάνει δείγματα, οποιουδήποτε επιθυμητού μεγέθους από την Posterior.

Ας υποθέσουμε ότι έχουμε ένα δείγμα, y , μεγέθους n με μέσο \bar{y} και διακύμανση s^2 , το οποίο έχει παρατηρηθεί από μια κανονική κατανομή με μέσο μ και διακύμανση σ^2 , και το οποίο έχει ανεξάρτητες prior κατανομές του τύπου που περιγράφηκε παραπάνω, και επιθυμούμε να περιγράψει την posterior κατανομή, $p(\mu, \sigma^2 / y)$.

Η πρώτη ανάλυση εξελίσσεται με το δεδομένο ότι η posterior $p(\mu, \sigma^2 / y)$ είναι μια κανονική κατανομή με μέσο

$$\mu_n = \frac{\omega_0^{-2} \mu_0 + n \sigma^{-2} \bar{y}}{\omega_0^{-2} + n \sigma^{-2}} \quad (1.5)$$

και διακύμανση

$$\omega_n^{-2} = (\omega_0^{-2} + n \sigma^{-2})^{-1} \quad (1.6)$$

Ένα ζευγάρι τιμών θα μπορούσε να παριστάνεται από το

$$p(\mu, \sigma^2 / y) = p(\mu / \sigma^2, y) p(\sigma^2 / y)$$

αφού πρώτα προσεγγίσουμε μια τιμή του σ^2 από την κατανομή marginal posterior . Τότε χρησιμοποιούμε την τιμή αυτή του σ^2 στους τύπους (1.5) και (1.6) για να καθορίσουμε την conditional posterior κατανομή $p(\mu / \sigma^2, y)$ από την οποία μπορούμε να προσεγγίσουμε την τιμή του μ .

Εκεί βρίσκεται και η δυσκολία της προσέγγισης της τιμής από το $p(\sigma^2 / y)$, εξαιτίας του ότι δεν έχει σταθερό τύπο. Μια χρήσιμη μέθοδος που μπορεί να χρησιμοποιηθεί είναι να παρατηρήσουμε αυτό που η Μπεϋζιανή θεωρία δίνει:

$$p(\sigma^2 / y) = \frac{p(\mu, \sigma^2 / y)}{p(\mu / \sigma^2, y)}$$

Ο αριθμητής είναι ανάλογος με το γινόμενο των n κανονικών πυκνοτήτων (για τα δεδομένα) και της prior πυκνότητας. Ο παρονομαστής είναι η κανονική κατανομή καθορισμένη από τις σχέσεις (1.5) και (1.6).

Η πυκνότητα της marginal posterior για το σ^2 είναι ανάλογη του ακόλουθου πολύπλοκου τύπου

$$q(\sigma^2) = p(\sigma^2) \cdot \sigma^{-n} \cdot \exp\left[-\frac{(n-1)s^2}{2\sigma^2}\right] \sqrt{\frac{\omega_n^2}{\omega_0^2} \exp\left[-\frac{1}{2}\left(\frac{ny}{\sigma^2} - \frac{\mu_n^2}{\omega_n^2}\right)\right]} \quad (1.7)$$

Μπορούν να παρατηρηθούν διάφορα θέματα που αφορούν αυτήν την πυκνότητα:

- (1) Πρέπει να υπενθυμίσουμε ότι τόσο το ω_n^2 όσο και το μ_n βασίζονται στο σ^2
- (2) Η έκφραση δεν ανταποκρίνεται σε καμία σταθερή συνάρτηση πυκνότητας.

- (3) Αν χρησιμοποιηθεί μια μη πληροφοριακή prior για το μ με $\mu_0 = 0$ και $\omega_0 \rightarrow \infty$, τότε ο τελευταίος εκθετικός παράγοντας του τύπου (1.7) μειώνεται τείνοντας στην μονάδα και ο παράγοντας $\sqrt{\frac{\omega_n^2}{\omega_0^2}}$ προσθέτει έναν περαιτέρω παράγοντα του σ στην πυκνότητα. Η σχέση του συμπεράσματος είναι μια κατανομή scaled inverse- X^2 , έτσι, αν η $p(\sigma^2)$ είναι επίσης μια κατανομή scaled inverse- X^2 , δεσμευμένη πιθανότητα, η σύζευξη αποκαθίσταται.
- (4) Η δειγματοληψία από την πυκνότητα που είναι ανάλογη με αυτήν του (1.7) απαιτεί αριθμητικές μεθόδους. Μια απλή στρατηγική είναι να υπολογίσουμε μια διακριτή προσέγγιση ενός συνόλου σημείων, ας πούμε $\sigma_1, \sigma_2, \dots, \sigma_M$ με την πιθανότητα ότι το σ είναι μικρότερο ή ίσο της του σ_i να δίνεται από τον τύπο

$$p(\sigma \leq \sigma_i) = \frac{\sum_{j=1}^i q(\sigma_j)}{\sum_{j=1}^M q(\sigma_j)} \quad (1.8)$$

Η posterior πυκνότητα είναι $q(\sigma) / \int q(\sigma) d\sigma$, αλλά το ολοκλήρωμα μπορεί να είναι αδύνατο να υπολογιστεί αναλυτικά. Αυτή η διακριτή προσέγγιση έχει το πλεονέκτημα ότι μια έγκυρη συνάρτηση κατανομής μπορεί να φορμαριστεί γνωρίζοντας μόνο την μη κανονική πυκνότητα $q(\cdot)$. Δείγματα τα οποία έχουν προσεγγισθεί από την marginal posterior πυκνότητα μπορούν να επιτευχθούν από την αντίστροφη αθροιστική συνάρτηση κατανομής -CDF, μέθοδος που περιγράφηκε από τον Gelman (1995) με την προσέγγιση να χρησιμοποιείται για το άθροισμα των συναρτήσεων κατανομών.

Η μέθοδος μπορεί εύκολα να επεκταθεί στην περίπτωση της γραμμικής παλινδρόμησης, μολονότι επιβαρυνόμαστε με μερικές καταγεγραμμένες πολυπλοκότητες.

Παρ όλα αυτά, είναι πιθανή μια «κομψή» εναλλακτική προσέγγιση. Η μέθοδος ίσως έχει σοβαρά πρακτικά μειονεκτήματα, αλλά μπορεί να φανεί χρήσιμη όταν ο αριθμός των παραμέτρων είναι μικρός.

ΚΕΦΑΛΑΙΟ 2

Ανάλυση Επιβίωσης (Survival Analysis)

2.1 Εισαγωγή

Σε πολλές μελέτες η μεταβλητή που μας ενδιαφέρει άμεσα είναι το χρονικό διάστημα που περνάει πριν συμβεί κάποιο γεγονός. Το γεγονός αυτό μπορεί να είναι απλά θάνατος ως συμβάν, ή ο θάνατος εξαιτίας κάποιας συγκεκριμένης ασθένειας. Για το λόγο αυτό η ανάλυση τέτοιων δεδομένων αναφέρεται συχνά ως ανάλυση επιβίωσης (survival analysis).

Ένα παράδειγμα μιας τέτοιας μελέτης είναι η κλινική δοκιμή για την θεραπεία του κακοήθους όγκου όπου είναι ανεπαρκής η πρόγνωση. Ο θάνατος ή η υποτροπή του όγκου θα μπορούσε να είναι το σημείο τερματισμού της έρευνας. Τέτοιες μελέτες συχνά περιλαμβάνουν ασθενείς για τους οποίους το γεγονός δεν έχει συμβεί μέχρι τη στιγμή που ξεκινά η ανάλυση. Αν και η στιγμή του γεγονότος για έναν τέτοιο ασθενή είναι άγνωστη, υπάρχουν αρκετές πληροφορίες που αφορούν την εκτίμησή του, αφού είναι γνωστό ότι πρέπει να ξεπερνάει τον παρόντα χρόνο επιβίωσης.

Μια παρατήρηση τέτοιου τύπου καλείται αποκομμένη τιμή (censored value). Οι μέθοδοι των αναλύσεων πρέπει να είναι ικανές να ανταπεξέρχονται σε μελέτες με αποκομμένες παρατηρήσεις. Συχνά, στην αρχή της δοκιμής παρατηρείται ένας αριθμός μεταβλητών και η επιβίωση σχετίζεται με τις τιμές των μεταβλητών αυτών. Κάτι που σημαίνει ότι οι μεταβλητές αυτές είναι προγνωστικές. Οι μέθοδοι των αναλύσεων πρέπει να είναι ικανές να λαμβάνουν υπόψη την κατανομή των προγνωστικών μεταβλητών στις υπό μελέτη ομάδες.

Ο αριθμός των ερευνών του προαναφερθέντος τύπου έχει αυξηθεί κατά τη διάρκεια των τριών τελευταίων δεκαετιών και οι στατιστικές μέθοδοι έχουν βελτιωθεί ώστε να μπορούν να αναλύουν τέτοια δεδομένα. Πολλές από τις μεθόδους αυτές αναπτύχθηκαν κατά τη διάρκεια της δεκαετίας του 1970. Μερικές μέθοδοι θα περιγραφούν στο κεφάλαιο αυτό.

Μία δεύτερη κατάσταση όπου χρησιμοποιεί ευρέως η ανάλυση επιβίωσης είναι η μελέτη της εργασιακής θνησιμότητας, όπου απαιτείται να αξιολογήσουμε αν μια ομάδα εργαζομένων που εκτίθενται σε μια ρυπαντική ουσία υφίσταται μεγάλο αριθμό

γεγονότων θανάτου. Τα άτομα που εισέρχονται στη μελέτη είναι υγιή, και για το λόγο αυτό, μία κοινή μέθοδος ανάλυσης που χρησιμοποιείται είναι η σύγκριση των παρατηρούμενων γεγονότων θανάτου, λαμβάνοντας υπόψη τόσο την χρονική στιγμή όσο και την αιτία από την οποία προκλήθηκε, με το τι θα αναμενόταν εάν η υπό μελέτη ομάδα ήταν μέρος ενός άλλου πληθυσμού με παρόμοιο δείκτη θνησιμότητας. Πολλές από τις μεθόδους των αναλύσεων βασίζονται στην προσέγγιση life-table, την οποία περιγράφουμε παρακάτω.

2.2 Life-Table

Το life-table που αρχικά αναπτύχθηκε αρκετά ικανοποιητικά από τον E.Halley (1656-1742), είναι ένα από τα βασικά εργαλεία της vital στατιστικής και της αναλογιστικής επιστήμης. Η προτυποποίηση που εισάγεται στην υποενότητα 4.3 του 4^{ου} κεφαλαίου είναι μία μέθοδος σύνοψης ενός συνόλου ηλικιακών δεικτών θανάτου, έτσι ώστε να παραχθεί ένα σύνθετο μέτρο εμπειρικής θνησιμότητας για μια ευρύτερη κοινότητα που περιλαμβάνει όλες τις ηλικίες και επιτρέπει την χρήσιμη σύγκριση με την εμπειρική θνησιμότητα άλλων ομάδων. Το life-table είναι μια εναλλακτική συνοπτική διαδικασία με αρκετά παρόμοια γνωρίσματα. Ο σκοπός της είναι να παρουσιάσει το δείγμα της επιβίωσης μιας ομάδας ανθρώπων, καθ' όλη τη διάρκεια της ζωής τους, σ' έναν υπό συζήτηση (ή υπό αμφισβήτηση) δείκτη συγκεκριμένης ηλικίας.

Υπάρχουν δύο διαφορετικοί τρόποι με τους οποίους μπορεί να κατασκευαστεί ένα life-table από τα δεδομένα θνησιμότητας για ένα μεγάλο πλήθος ατόμων. Οι δυο αυτοί τύποι καλούνται συνήθως "current life-table" και "cohort ή generation life-table". Το current life-table¹ περιγράφει το μοντέλο επιβίωσης μιας ομάδας ατόμων, που υπόκειται καθ' όλη τη διάρκεια της ζωής των ατόμων σε έναν συγκεκριμένο ηλικιακό δείκτη θνησιμότητας (age-specific death rate) που έχει προκύψει από πρόσφατη παρατήρηση μιας συγκεκριμένης κοινότητας.

Αυτή η ομάδα είναι απαραίτητα υποθετική. Μια ομάδα ατόμων που τώρα είναι 60 ετών θα βιώσουν κατά προσέγγιση τον επόμενο χρόνο τον πρόσφατο δείκτη

(1) current life-table : κατασκευάζεται από α) την απογραφή πληροφοριών βασισμένων σε άτομα κάθε ηλικίας, για ένα δοσμένο έτος και β) την vital στατιστική που βασίζεται στον αριθμό θανάτων ή αποτυχιών σ' ένα δοσμένο έτος, για κάθε ηλικία. Αυτός ο τύπος life-table αναφέρεται συχνά υπό τον όρο μιας συμμετοχής 100.000 ατόμων

θνησιμότητας που ειδικεύεται στις ηλικίες 60-61, αλλά όσοι από αυτούς επιβιώσουν άλλα 10 χρόνια δεν θα βιώσουν στον 11ο χρόνο τον τωρινό δείκτη θνησιμότητας για τις ηλικίες 70-71, αλλά τον δείκτη που θα επικρατεί σε 10 χρόνια. Το πρόσφατο life-table συνεπώς είναι μια πρακτική σύνοψη της πρόσφατης θνησιμότητας και όχι μια περιγραφή της ακριβούς θνησιμότητας που συμβαίνει σε οποιαδήποτε ομάδα.

Η μέθοδος της κατασκευής των πρόσφατων life-tables δημοσιεύτηκε σε διεθνείς πηγές της vital statistics.

Μια απλοποιημένη προσέγγιση δίνεται περιγραφικά από τους Hill and Hill(1991). Τα βασικά χαρακτηριστικά των life-table φαίνονται στον Πίνακα 2.1 στην αριστερή στήλη του οποίου εμφανίζεται το 10^ο Αγγλικό Life-Table, το οποίο υποδεικνύει την θνησιμότητα των αντρών στην Αγγλία και την Ουαλία μεταξύ 1930-1932. Η δεύτερη στήλη δίνει το q_x , την πιθανότητα ένα άτομο που ζει ακριβώς x χρόνια, να πεθάνει πριν τα επόμενα γενέθλιά του. Η τρίτη στήλη δείχνει l_x , δηλαδή τον αριθμό των ατόμων που μεταξύ ενός αυθαίρετου πλήθους 1000 γεννήσεων, θα επιβιώσουν μέχρι το γενέθλιο έτος τους x . Για να επιβιώσει ένα άτομο μέχρι την ηλικία αυτή πρέπει να επιβιώσει τον πρώτο χρόνο, τον δεύτερο και πάει λέγοντας. Συνεπώς

$$l_x = l_0 p_0 p_1 \dots p_{x-1} \quad (2.1)$$

όπου το $p_x = 1 - q_x$. Αυτός ο τύπος μπορεί να επιβεβαιωθεί από τον πίνακα 2.1 για $x=1$, αλλά όχι και για όλες τις ακόλουθες τιμές γιατί οι τιμές του q_x που δίνονται εδώ αφορούν μόνο τιμές του x που έχουν επιλεγθεί. Ένας τέτοιος πίνακας καλείται *abridged life-table*.

Πίνακας 2.1

Current and cohort life-tables για τους άντρες στην Αγγλία και Ουαλία που έχουν γεννηθεί γύρω στο 1931.					
Age (years)	Current life-tables 1930-32			Cohort life-table, 1931 cohort	
	Probability of death between age x and $x+1$		Life-table survivors	Life-table survivors	
	0^*		l_x	l_x	
A:			Expectation of life e_x		
0	00719		1000	58-7	1000
1	0-0153		928-1	62-2	927-8
5	0-0034		900-7	60-1	903-6
10	0-0015		890-2	55-8	894-8
20	0-0032		872-4	46-8	884-2
30	0-0034		844-2	38-2	874-1
40	0-0056		809-4	29-6	861-8
50	0-0113		747-9	21-6	829-7
60	0-0242		636-2	14-4	—
70	0-0604		433-6	8-6	—
80	0-1450		162-0	4-7	—

Η τέταρτη στήλη που συμβολίζεται με e_x^o , δείχνει την προσδοκία ζωής έως το έτος x . Αυτή είναι η μέση διάρκεια του επιπλέον χρόνου ζωής πέραν από την ηλικία x για όλα τα άτομα l_x που είναι εν ζωή στην ηλικία x .

Σ' έναν ολοκληρωμένο πίνακα το e_x^o μπορεί να υπολογιστεί προσεγγιστικά από τον τύπο

$$e_x^o = (l_{x+1} + l_{x+2} + \dots) / l_x + \frac{1}{2} \quad (2.2)$$

όπου ο όρος της παρένθεσης είναι ο συνολικός αριθμός των ετών που έζησαν τα l_x άτομα πέρα από την ηλικία των x ετών, αν εκείνοι που ήταν ετοιμοθάνατοι μεταξύ της ηλικίας y και $y+1$ πέθαναν αμέσως μετά το y έτος, και το $\frac{1}{2}$ είναι μια διόρθωση που επιτρέπει την πραγματικότητα ότι οι θάνατοι λαμβάνουν χώρα καθ' όλη τη διάρκεια οποιασδήποτε ηλικίας, και το οποίο πολύ χονδρικά προσθέτει μισό έτος στο μέσο χρόνο επιβίωσης.

Το cohort life table² περιγράφει την ακριβή εμπειρία επιβίωσης μιας ομάδας ή την ακολουθία των γεννήσεων την ίδια χρονική στιγμή. Το 1990 για παράδειγμα, οι γεννήσεις αυτές υπόκεινται καθ' όλη την διάρκεια του πρώτου έτους στην ομάδα που υποδεικνύει τον υπολογισμό του δείκτη θνησιμότητας που επικρατεί την περίοδο 1900-1901 για βρέφη κάτω του ενός έτους.

Αν επιζήσουν έως το 10^ο έτος της ηλικίας τους τότε υποδεικνύουν την ηλικιακή θνησιμότητα που επικρατεί την περίοδο 1910-11, και συνεχίζει ομοίως. Τα cohort life-tables συνοψίζουν την θνησιμότητα για τις διαφορετικές ηλικίες τις στιγμές που η cohort θα είναι συμπήξει με τις ηλικίες αυτές. Το δεξί μέρος του πίνακα 2.1 συνοψίζει το l_x για το cohort life-table για τους άνδρες στην Αγγλία και την Ουαλία με το πολύ 5 χρόνια απόκλιση από το 1931.

Όπως αναμενόταν, οι τιμές του l_1 στα δυο life-tables είναι παρόμοιες, όντας ανεξάρτητες όσον αφορά την βρεφική θνησιμότητα σχετικά με τα ίδια ημερολογιακά έτη. Σε μεγαλύτερες ηλικίες οι τιμές του l_x είναι μεγαλύτερες για το cohort table γιατί εξαρτώνται από τον δείκτη θνησιμότητας μεγαλύτερων ηλικιών που έζησαν από το 1931 και οι οποίες είναι χαμηλότερες του δείκτη του 1931. Και οι δύο δείκτες του life-table είναι χρήσιμοι για τις vital statistics και τις επιδημιολογικές μελέτες. Τα πρόσφατα life-tables συνοψίζουν την τωρινή θνησιμότητα και ίσως χρησιμοποιηθούν σαν μια εναλλακτική μέθοδος της προτυποποίησης για σύγκριση μεταξύ μοντέλων θνησιμότητας διαφορετικών κοινοτήτων.

Τα cohort life-tables είναι ιδιαίτερα χρήσιμα σε μελέτες εργατικών θανάτων, όπου μια ομάδα παρακολουθείται για μεγάλη χρονική περίοδο.

2.3 Μελέτες υπό παρακολούθηση

Πολλές ιατρικές έρευνες ασχολούνται με τα μοντέλα επιβίωσης συγκεκριμένων ομάδων ασθενών - για παράδειγμα, εκείνων που υποφέρουν από συγκεκριμένο τύπο κακοήθους ασθένειας. Η επιβίωση ίσως είναι κατά μέσο όρο πολύ μικρότερη από άλλα μέλη του γενικού πληθυσμού. Αφού η ηλικία πιθανό να είναι λιγότερο σημαντικός παράγοντας από την πρόοδο της ασθένειας, είναι φυσικό να μετράμε την επιβίωση από ένα

(2) Cohort life-table: περιγράφει τη θνησιμότητα που βιώνουν από τη γέννηση έως το θάνατο για μια συγκεκριμένη ακολουθία ανθρώπων που γεννιούνται την ίδια στιγμή. Οι άνθρωποι σε κίνδυνο στην αρχή του διαστήματος είναι όσοι επέζησαν το προηγούμενο διάστημα.

συγκεκριμένο στάδιο του ιστορικού της ασθένειας, όπως η ημερομηνία στην οποία έγινε μια συγκεκριμένη επέμβαση.

Η εφαρμογή των μεθόδων life-table στα δεδομένα που προέρχονται από τις υπό μελέτη ομάδες θα μας απασχολήσουν τώρα λεπτομερειακά. Στην θεωρία οι μέθοδοι είναι εφαρμόσιμες στις περιπτώσεις στις οποίες το σημείο τερματισμού δεν είναι ο θάνατος, αλλά κάποια μη θανάσιμα γεγονότα, όπως η επανεμφάνιση κάποιων συμπτωμάτων και τα σημάδια μετά από μια υποτροπή, όπου βέβαια ίσως να μην είναι δυνατόν να καθοριστεί η ακριβής στιγμή της επανεμφάνισης των συμπτωμάτων, όπως στην περίπτωση της στιγμής του θανάτου που συνήθως καθορίζεται με ακρίβεια. Στην πραγματικότητα το γεγονός ίσως είναι περισσότερο επιθυμητό απ' ότι ανεπιθύμητο. Η εξαφάνιση των συμπτωμάτων μετά την αρχή μιας θεραπείας είναι ένα παράδειγμα. Η παρακάτω συζήτηση γίνεται υπό τον όρο της επιβίωσης μετά από μια επέμβαση.

Τη στιγμή της ανάλυσης μιας τέτοιας υπό παρακολούθηση μελέτης οι ασθενείς είναι πιθανόν να έχουν παρατηρηθεί για διάφορες χρονικές περιόδους, κάποιοι εκ των οποίων έχουν κάνει την επέμβαση πολύ μεγάλο διάστημα πριν, και άλλοι έχοντας κάνει την επέμβαση πρόσφατα.

Κάποιοι ασθενείς θα έχουν πεθάνει, σε στιγμές που συνήθως μπορούν να εξακριβωθούν με κάποια σχετική ακρίβεια. Άλλοι είναι γνωστό ότι είναι εν ζωή τη στιγμή της ανάλυσης. Άλλοι ίσως έχουν χαθεί από την παρακολούθηση για διάφορους λόγους μεταξύ μιας εξέτασης και της επόμενης. Άλλοι ίσως έχουν αποσυρθεί από τη μελέτη για ιατρικούς λόγους. Ίσως από την παρεμβολή μιας άλλης ασθένειας ή ενός αιφνίδιου θανάτου.

Αν δεν υπήρχαν τέτοιες επιπλοκές σαν τις προαναφερθείσες, και αν κάθε ασθενής ήταν υπό παρακολούθηση έως τη στιγμή του θανάτου του, η κατασκευή του life-table για την χρονική περίοδο μετά την επέμβαση θα ήταν πολύ απλούστερη. Ο δείκτης επιβίωσης l_x στο life-table, είναι l_0 φορές η αναλογία του χρόνου επιβίωσης, η οποία είναι μεγαλύτερη του x . Έτσι το πρόβλημα θα ήταν απλά στο να επιτευχθεί η κατανομή της περιόδου επιβίωσης - μία εργασία πολύ στοιχειώδης. Στην περίπτωση μας όμως που το πρόβλημα είναι αρκετά πολύπλοκο κατασκευάστηκε ένας πίνακας όπως ο πίνακας 2.2. ώστε να ξεπεραστούν οι επιπλοκές των ελλιπώς συμπληρωμένων δεδομένων,

Πίνακας 2.2

Πίνακας 2.2 Life-table υπολογισμοί για γονείς με έναν ιδιαίτερο τύπο κακοήθους ασθένειας προσαρμοσμένη από τους Berkson και Gage (1950).

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Interval since operation (years)	Last reported during this interval		Living at start of interval	Adjusted number at risk	Estimated probability of death	Estimated probability of survival	Percentage of survivors after x years
x to $x + 1$	Died d_x	Withdrawn w_x	n_x	n_x	l_x	P_x	l_x
0-1	90	0	374	374-0	0-2406	0-7594	100-0
1-2	76	0	284	284-0	0-2676	0-7324	75-9
2-3	51	0	208	208-0	0-2452	0-7548	55-6
3-4	25	12	157	151-0	0-1656	0-8344	42-0
4-5	20	5	120	117-5	0-1702	0-8298	35-0
5-6	7	9	95	90-5	0-0773	0-9227	29-1
6-7	4	9	79	74-5	0-0537	0-9463	26-8
7-8	1	3	66	64-5	0-0155	0-9845	25-4
8-9	3	5	62	59-5	0-0504	0-9496	25-0
9-10	2	5	54	51-5	0-0388	0-9612	23-7
10-	21	26	47	-	-	-	22-8

Αυτός ο πίνακας υιοθετήθηκε από τους Berkson and Gage (1950) σε μία από τις πρώτες δημοσιεύσεις που περιγράφουν την μέθοδο. Στα πρωτότυπα δεδομένα, τα διαστήματα περιόδου μετρήθηκαν από τη χρονική στιγμή του εξιτηρίου από το νοσοκομείο, αλλά για τους σκοπούς της ευκολότερης παρουσίασης έχουμε αλλάξει αυτά σε διαστήματα που ακολούθησαν μετά την επέμβαση. Οι στήλες (1)–(8) φορμαρίστηκαν ως εξής:

- (1) Η επιλογή των χρονικών περιόδων βασίζονται στην φύση των δεδομένων. Οι εκτιμήτριες στην παρούσα μελέτη χρειάστηκαν τους δείκτες επιβίωσης για ακέραιους αριθμούς ετών, έως 10, μετά την επέμβαση. Αν υπήρχε μεγάλο ενδιαφέρον για την επιβίωση πέραν των 10 ετών, τότε τα διαστήματα εύκολα θα μπορούσαν να επεκταθούν πέραν των 10 ετών. Στην περίπτωση αυτή, για να αποφύγουμε το να γίνει ο πίνακας υπερβολικά δυσανάλογος, ίσως να είναι χρήσιμο να χρησιμοποιήσουμε διαστήματα 2 ετών για τουλάχιστον κάποιες από τις ομάδες. Άνισα διαστήματα δεν αποτελούν πρόβλημα. Παράδειγμα μπορούμε να δούμε από τους Merrell and Shulman (1955).

- (2) & (3) Οι ασθενείς της μελέτης κατατάσσονται σύμφωνα με το πώς ήταν η κατάσταση της υγείας τους στην τελευταία αναφορά που έγινε για αυτήν. Αν η αναφορά ήταν ένας θάνατος, ο ασθενής υπολογίζεται στη στήλη (2), ενώ οι ασθενείς που ήταν εν ζωή στην τελευταία αναφορά υπολογίζονται στη στήλη (3). Συνεπώς ο όρος «withdraw» (αποσύρομαι) περιλαμβάνει ασθενείς που πρόσφατα καταγράφηκαν εν ζωή και οι οποίοι θα συνεχίσουν να παρακολουθούνται σε μελλοντικές ελεγχόμενες (υπό παρακολούθηση) εξετάσεις, κι επιπλέον ασθενείς που έχουν χαθεί από την παρακολούθηση για οποιονδήποτε λόγο.
- (4) Το πλήθος των ασθενών που επιβιώνει στην αρχή των διαστημάτων επιτυγχάνεται αθροίζοντας το πλήθος των ασθενών των στηλών (2) και (3) όπως εμφανίζονται στην τελευταία γραμμή. Επομένως, ο αριθμός των εν ζωή ασθενών τα τελευταία δέκα χρόνια είναι $21 + 26 = 47$. Ο αριθμός των εν ζωή τα 9 τελευταία χρόνια περιλαμβάνει αυτούς τους 47 και επιπλέον τους $2 + 5 = 7$ που πέθαναν ή αποσύρθηκαν στο διάστημα 9-10 χρόνων. Το αποτέλεσμα είναι δηλαδή $47 + 7 = 54$.
- (5) Ο αναπροσαρμοσμένος υπό κίνδυνο αριθμός κατά τη διάρκεια του διαστήματος $x, x + 1$ είναι

$$nx' = nx - \frac{1}{2}W_x \quad (2.3)$$

Σκοπός αυτού του τύπου είναι να παράγει έναν παρονομαστή για την επόμενη στήλη. Η αιτιολογία περιγράφεται παρακάτω.

- (6) Η εκτιμώμενη πιθανότητα του θανάτου κατά τη διάρκεια του διαστήματος $x, x-1$ είναι

$$q_x = d_x / n'_x \quad (2.4)$$

Για παράδειγμα, στην πρώτη γραμμή, έχουμε

$$q_0 = 90 / 374,0 = 0,2406$$

Η αναπροσαρμογή από το n_x στο n'_x είναι απαραίτητη γιατί οι αποχωρήσεις w_x είναι υπό κίνδυνο μόνο για το διάστημα αυτό. Είναι πιθανό να δείξουμε πολύ μεγάλη ανοχή για τις αποχωρήσεις, ιδιαίτερα αν το σημείο του διαστήματος που έγινε η αποχώρηση είναι γνωστό. Όπως επίσης, είναι συχνά επαρκές το να θεωρήσουμε ότι οι αποχωρήσεις έχουν την ίδια επίδραση όπως θα συνέβαινε αν οι μισές από αυτές ήταν υπό κίνδυνο για όλη την περίοδο.

Ένας εναλλακτικός ισχυρισμός είναι ότι αν οι w_x ασθενείς δεν είχαν αποχωρήσει, ίσως αναμέναμε περίπου $\frac{1}{2}q_x w_x$ επιπλέον θανάτους. Ο συνολικός αριθμός θανάτων θα ήταν τότε $d_x + \frac{1}{2}q_x w_x$ και θα είχαμε εκτιμημένο δείκτη θνησιμότητας

$$q_x = \frac{d_x + \frac{1}{2}q_x w_x}{n_x} \quad (2.5)$$

Το (2.5) είναι ισοδύναμο του (2.3) και (2.4).

(7) Ισχύει η ισότητα $p_x = 1 - q_x$

(8) Η εκτιμημένη πιθανότητα της επιβίωσης,ας πούμε για παράδειγμα 3 χρόνια μετά την επέμβαση, είναι $p_0 p_1 p_2$. Οι καταχωρήσεις δεδομένων της τελευταίας στήλης, που συχνά καλούνται *life-table survival rates*, επιτυγχάνεται από διαδοχικούς πολλαπλασιασμούς της στήλης (7), με έναν αυθαίρετο πολλαπλασιαστή $l_0=100$. Επίσημα,

$$l_x = l_0 p_0 p_1 \dots p_{x-1} \quad (2.6)$$

όπως στην (2.1).

Τη βάση αυτών των υπολογισμών αποτελούν δύο σημαντικές εικασίες. Πρώτον, θεωρείται ότι όσοι αποχωρούν υπόκεινται στις ίδιες πιθανότητες θανάτου με αυτούς που δεν αποχωρούν. Αυτή είναι μια δικαιολογημένη υπόθεση για τους αποχωρούντες που είναι ακόμη υπό παρακολούθηση και είναι διαθέσιμοι για μελλοντικές παρακολουθήσεις. Ίσως όμως είναι μια ριψοκίνδυνη εικασία για ασθενείς που εξαφανίστηκαν κατά την διάρκεια της παρακολούθησης, αφού δεν έχουμε την δυνατότητα να ερευνήσουμε αν

κάποιος από τους ασθενείς αυτούς αποχώρησε για οποιοδήποτε λόγο που έχει ίσως να κάνει με την υγεία του. Δεύτερον, οι διάφορες τιμές του p_x επιτυγχάνονται από ασθενείς που εισχωρούν στην μελέτη σε διαφορετικές χρονικές στιγμές. Πρέπει να υποθέσουμε ότι οι πιθανότητες παραμείνουν αμετάβλητες στο χρόνο, διαφορετικά οι υπολογισμοί του life-table αντιπροσωπεύουν ποσότητες με σύνθετη ερμηνεία.

Στον Πίνακα 2.2 οι υπολογισμοί μπορεί να συνεχιστούν πέρα από τα 10 χρόνια. Υποθέτουμε, παρόλα αυτά, ότι το d_{10} και w_{10} ήταν και τα δυο μηδέν, όπως θα ήταν αν κανένας από τους ασθενείς δεν είχε παρατηρηθεί για περισσότερο από 10 χρόνια. Τότε το n_{10} θα ήταν μηδέν, καμιά από τις τιμές των q_{10} και p_{10} δεν θα μπορούσε να υπολογιστεί και γενικά, καμιά τιμή του l_{11} δεν θα ήταν διαθέσιμη εκτός και αν το l_{10} ήταν μηδέν (που θα συνέβαινε αν κάποια από τις τιμές p_0, p_1, \dots, p_9 ήταν 0), όπου και στην περίπτωση αυτή το l_{11} θα ήταν 0. Όλα αυτά μπορούν να ληφθούν περισσότερο σοβαρά με το να πούμε ότι καμιά πληροφορία επιβίωσης δεν θα ήταν διαθέσιμη για περιόδους μακράς παρακολούθησης περισσότερο από την μέγιστη που παρατηρείται στην μελέτη. Αυτό σημαίνει ότι η προσδοκία του χρόνου ζωής δεν μπορεί να υπολογιστεί από μελέτες υπό παρακολούθηση, εκτός και αν η περίοδος της παρακολούθησης, τουλάχιστον για ορισμένους ασθενείς, είναι επαρκώς μεγάλη ώστε να καλύψει σχεδόν ολόκληρο το διάστημα της επιβίωσης. Για το λόγο αυτό ο δείκτης επιβίωσης του life-table (στήλη (8) του life-table 2.2) είναι το πιο γενικό σε χρησιμότητα μέτρο της επιβίωσης. Για μια συμμετρική κατανομή αυτό θα μπορούσε να ισοδυναμεί με το προσδόκιμο ζωής.

Για περαιτέρω συζήτηση των μεθόδων του life-table στις υπό παρακολούθηση μελέτες, παραπέμπουμε στους Berkson και Gage (1950), Merrell και Shulman (1955), Cutler και Ederer (1958) και Newell (1961).

2.4 Δειγματικά σφάλματα στο life-table

Καθεμιά από τις τιμές του p_x σε έναν υπολογισμό του life-table υπόκειται σε δειγματική απόκλιση. Αν δεν πρόκειται για τους αποχωρούντες, η απόκλιση θα μπορούσε να θεωρηθεί ως διωνυμική, με ένα μέγεθος δείγματος n_x . Η επίδραση των ατόμων που αποχώρησαν είναι προσεγγιστικά η ίδια με αυτήν του λόγου της μείωσης του μεγέθους του δείγματος προς το n'_x . Η διακύμανση της l_x δίνεται προσεγγιστικά από τον ακόλουθο τύπο του Greenwood (1926), ο οποίος μπορεί να επιτευχθεί παίρνοντας λογαρίθμους και χρησιμοποιώντας μια επέκταση του.

$$\text{var}(l_x) = l_x^2 \cdot \sum_{i=0}^{x-1} \frac{d_i}{n'_i(n'_i - d_i)} \quad (2.7)$$

Στον πίνακα 2.2 για παράδειγμα όπου $l_4 = 35,0\%$ θα έχουμε

$$\text{var}(l_4) = (35,0)^2 = \left[\frac{90}{374 \cdot 284} + \frac{76}{284 \cdot 208} + \frac{51}{208 \cdot 157} + \frac{25}{(51 \cdot 126)} \right] = 6,14$$

Έτσι προκύπτει $SE(14) = \sqrt{6,14} = 2,48$ και προσεγγιστικά το 95% των ορίων εμπιστοσύνης για το l_4 είναι $35,0 \pm (1,96) \cdot (2,48) = 30,1$ και $39,9$.

Η εφαρμογή του τύπου (2.7) μπορεί να οδηγήσει σε απίθανες τιμές για τα όρια εμπιστοσύνης έξω από το εύρος 0 έως 100%. Μια εναλλακτική περίπτωση που δεν αναφέρθηκε παραπάνω είναι να εφαρμόσουμε μετατροπή με τον διπλό λογάριθμο, $\ln(-\ln l_x)$, στον τύπο (2.6), με το $l_0=1$ έτσι ώστε το l_x να είναι η αναλογία με επιτρεπόμενο εύρος 0 έως 1 (Kalbfleisch & Prentice, 1980). Τότε, ο τύπος του Greenwood τροποποιείται ώστε να δώσει τα 95% όρια εμπιστοσύνης για το l_x ως εξής,

$$l_x^{\exp(\pm 1,965)} \quad (2.8),$$

όπου

$$s = SE(l_x) / (-l_x \ln l_x).$$

Για το παραπάνω παράδειγμα, $l_4 = 0,35$, $SE(l_4) = 0,0248$, $s = 0,0675$, $\exp(1,96s) = 1,14$, $\exp(-1,96s) = 0,876$, και τα όρια είναι $0,35^{1,14}$ έως $0,35^{0,876}$, τα οποία είναι ίσα με $0,302$ και $0,399$ αντίστοιχα. Στην περίπτωση αυτή, όπου χρησιμοποιώντας την σχέση (2.7) τα όρια δεν είναι κοντά σε οποιοδήποτε άκρο επιτρεπόμενου εύρους, ο τύπος (2.8) φαίνεται να δίνει σχεδόν ολόιδιες τιμές με τον (2.7).

Ο Peto (1977) δίνει έναν τύπο για το σφάλμα $SE(l_x)$, ο οποίος είναι ευκολότερος στον υπολογισμό από τον (2.7). Συγκεκριμένα,

$$SE(l_x) = l_x \sqrt{[(1-l_x) / n'_x]} \quad (2.9)$$

Όπως και στον τύπο (2.8), είναι απαραίτητο να χρησιμοποιούμε το l_x σαν μια αναλογία.

Από το παράδειγμα (2.9) προκύπτει ότι $SE(l_4)=0.0258$. Ο τύπος (2.9) είναι συντηρητικός αλλά ίσως και ο πιο κατάλληλος για την περίοδο της αυξανόμενης αβεβαιότητας στο τέλος των life-table, όταν δηλαδή υπάρχουν λίγοι εν ζωή ασθενείς που συνεχίζουν να υπόκεινται σε παρακολούθηση.

Οι μέθοδοι για τον υπολογισμό της δειγματικής διακύμανσης των διαφόρων δεδομένων που έχουν καταχωρηθεί στο life-table, συμπεριλαμβανομένου και του προσδόκιμου ζωής, δίνονται από τον Chiang (1984).

2.5 Ο εκτιμητής Kaplan-Meier

Το εκτιμώμενο life-table όπως εμφανίζεται στον Πίνακα (2.2) υπολογίστηκε μετά από διαχωρισμό της περιόδου παρακολούθησης εντός συγκεκριμένων χρονικών ορίων. Σε μερικές περιπτώσεις επομένως ίσως τα δεδομένα να είναι διαθέσιμα σε τύπους ομάδων και όχι μεμονωμένα, κάτι που συχνά είναι πιο πρακτικό. Όπως φαίνεται κι από τον πίνακα αυτό, το life-table περιλαμβάνει μια αυθαίρετη επιλογή για τα χρονικά διαστήματα. Αυτό θα μπορούσε να αποφευχθεί με τη χρήση μιας μεθόδου βασισμένης στους Kaplan και Meier (1958).

Στην συγκεκριμένη μέθοδο, τα δεδομένα ουσιαστικά θεωρούνται ομαδοποιημένα σε ένα μεγάλο αριθμό χρονικών περιόδων, με κάθε διάστημα τόσο μικρό όσο επιτρέπει η ακρίβεια του ιστορικού που έχει καταγραφεί για κάθε ασθενή. Όμως, αν η επιβίωση καταγράφηκε με ακρίβεια μιας ημέρας τότε θα χρησιμοποιηθούν χρονικά διαστήματα εύρους 1 ημέρας. Υποθέτουμε ότι την χρονική στιγμή t_j υπάρχουν d_j θάνατοι και ακριβώς πριν συμβούν οι θάνατοι αυτοί υπήρχαν n_j άτομα εν ζωή. Τότε η εκτιμώμενη πιθανότητα θανάτου τη χρονική στιγμή t_j είναι

$$q_{t_j} = \frac{d_j}{n_j} \quad (2.10)$$

Ο παραπάνω τύπος είναι ισοδύναμος με τον (2.4). Από σύμβαση, αν οποιοδήποτε υποκείμενο είναι αποκομμένο τη χρονική στιγμή t_j , τότε θεωρείται ότι έχει επιβίωση περισσότερη από την χρονική στιγμή θανάτων t_j , και οι αλλαγές του τύπου (2.3) δεν εφαρμόζονται.

Για τα περισσότερα χρονικά διαστήματα ισχύει $d_j=0$ και άρα $q_{t_j}=0$ και η πιθανότητα επιβίωσης είναι $p_{t_j}(=1-q_{t_j})=1$.

Αυτά τα διαστήματα ίσως δεν λαμβάνονται υπόψη στον υπολογισμό της επιβίωσης του life-table που χρησιμοποιεί τον τύπο (2.6). Τότε η επιβίωση την χρονική στιγμή t , δηλαδή η l_t εκτιμάται από

$$l_t = \prod_j P_{l_j} = \prod_j \frac{n'_j - d_j}{n_j} \quad (2.11)$$

όπου το γινόμενο λαμβάνει υπόψη του όλα τα χρονικά διαστήματα στα οποία έχει συμβεί ένα γεγονός θανάτου, έως την χρονική στιγμή t την οποία και συμπεριλαμβάνει. Καθώς τα χρονικά διαστήματα τείνουν στο μηδέν η εκτίμηση αυτή εκφράζει και την εκτιμήτρια γινομένου-ορίου (*product-limit estimator*) αφού είναι ο οριακός τύπος του γινομένου του τύπου (2.6). Η εκτιμήτρια είναι επίσης η εκτιμήτρια μέγιστης πιθανοφάνειας (*the maximum likelihood estimator*). Οι εκτιμητές που επιτυγχάνονται είναι αναπόφευκτα εκφρασμένες σε γραφική μορφή. Η καμπύλη επιβίωσης αποτελείται από οριζόντιες γραμμές με κατακόρυφα βήματα κάθε φορά που συμβαίνει ένας θάνατος (οι υπολογισμοί φαίνονται στον πίνακα 2.4).

2.6 To Logrank test

Το test που περιγράφεται στην ενότητα αυτή χρησιμοποιείται για τις συγκρίσεις δυο ή περισσότερων ομάδων των δεδομένων επιβίωσης. Το πρώτο βήμα είναι να κατατάξουμε τους χρόνους επιβίωσης, τόσο τις παρατηρούμενες όσο και τις αποκομμένες, σε σειρά (rank order). Υποθέτουμε, για παράδειγμα, ότι υπάρχουν δύο ομάδες, οι A και B . Αν τη χρονική στιγμή t_j υπήρχαν d_j θάνατοι και υπήρχαν αντίστοιχα n'_{j_A} και n'_{j_B} άτομα εν ζωή μόλις πριν τη χρονική στιγμή t_j στις ομάδες A και B , τότε τα δεδομένα θα μπορούσαν να καταταχθούν σ' έναν 2x2 πίνακα ως εξής:

	Died	Survived	Total
Group A	d_{j_A}	$n'_{j_A} - d_{j_A}$	n'_{j_A}
Group B	d_{j_B}	$n'_{j_B} - d_{j_B}$	n'_{j_B}
Total	d_j	$n'_j - d_j$	n'_j

Εκτός από τους ισοδύναμους χρόνους επιβίωσης, ισχύει ότι το $d_j = I$ και καθεμιά από τις d_{j_A} και d_{j_B} είναι 0 ή 1. Σημειώνουμε επίσης ότι αν ένα άτομο είναι αποκομμένο τη χρονική στιγμή t_j τότε το άτομο αυτό θεωρείται υπό κίνδυνο την συγκεκριμένη χρονική στιγμή κι έτσι περιλαμβάνεται στο n'_j .

Στην ψευδή υπόθεση που ο κίνδυνος του θανάτου είναι ίδιος και στις δύο ομάδες, αναμένουμε ότι ο αριθμός των θανάτων σε οποιαδήποτε χρονική στιγμή θα κατανέμεται μεταξύ των δύο ομάδων ανάλογα με το πλήθος των ατόμων που βρίσκονται σε κίνδυνο. Έτσι έχουμε

$$E(d_{j_A}) = n'_{j_A} d_j / n'_j$$

$$Var(d_{j_A}) = \frac{d_j (n'_j - d_j) n'_{j_A} n'_{j_B}}{n_j'^2 (n'_j - 1)} \quad (2.12)$$

Στην περίπτωση που $d_j = I$, οι τύποι (2.12) απλοποιούνται ως εξής:

$$E(d_{j_A}) = p'_{j_A}$$

$$Var(d_{j_A}) = p'_{j_A} (1 - p'_{j_A})$$

όπου $p'_{jA} = n'_{jA} / n'_j$ η αναλογία των εν ζώη ατόμων που ανήκουν στην ομάδα A.

Η διαφορά μεταξύ των d_{jA} και $E(d_{jA})$ είναι μία ένδειξη ενάντια στην ψευδή υπόθεση. Το logrank test είναι ο συνδυασμός των διαφορών αυτών πάνω σε όλες τις χρονικές στιγμές που συμβαίνει ένας θάνατος. Είναι ανάλογο του Mantel-Haenszel test για το συνδυασμό δεδομένων πάνω στα στρώματα και εισήχθηκε πρώτη φορά με τον τρόπο αυτό από τον Mantel (1966).

Αθροίζοντας όλες τις χρονικές στιγμές θανάτου t_j έχουμε

$$\begin{aligned} O_A &= \sum d_{jA} \\ E_A &= \sum E(d_{jA}) \\ V_A &= \sum \text{var}(d_{jA}) \end{aligned} \quad (2.13)$$

Παρόμοια αθροίσματα μπορούν να επιτευχθούν για την ομάδα B και από τον τύπο (2.12) συνεπάγεται ότι $E_A + E_B = O_A + O_B$.

Το E_A ίσως αναφέρεται ως ο "αναμενόμενος" αριθμός θανάτων στην ομάδα A, αλλά επειδή σε κάποιες περιπτώσεις, το E_A ίσως ξεπερνά τον αριθμό των ατόμων της αρχικής ομάδας, μια πιο ακριβής περιγραφή είναι ο βαθμός της έκθεσης στον κίνδυνο του θανάτου (extent of exposure to risk of death, Peto, 1977). Ένα στατιστικό test για την ισοδυναμία του δείκτη θνησιμότητας στις δυο ομάδες είναι

$$X_1^2 = \frac{(O_A - E_A)^2}{V_A} \quad (2.14)$$

Το οποίο προσεγγιστικά είναι μια $X_{(1)}^2$. Ένα εναλλακτικό και απλούστερο test, το οποίο δεν απαιτεί τον υπολογισμό των όρων μεταβλητότητας, είναι το

$$X_2^2 = \frac{(O_A - \hat{A}_A)^2}{E_A} + \frac{(O_B - \hat{A}_B)^2}{E_B} \quad (2.15)$$

Η στατιστική προσεγγιστικά είναι μια $X_{(1)}^2$. Πρακτικά, ο τύπος (2.15) είναι επαρκής τις περισσότερες φορές, αλλά δημιουργεί πρόβλημα στην συντηρητική (ή ηθική) πλευρά (Peto & Pike, 1973).

Το logrank test ίσως μπορεί να γενικευτεί σε περισσότερες από δύο ομάδες. Η επέκταση του τύπου (2.14) συνεπάγεται την αντιστροφή του πίνακα διακύμανσης-συνδιακύμανσης του $O - E$ πάνω στις ομάδες (Peto & Pike, 1973), αλλά η επέκταση του (2.15) γίνεται απευθείας. Το άθροισμα του τύπου (2.15) επεκτείνεται ώστε να καλύψει όλες τις ομάδες, με τα αθροίσματα του τύπου (2.13) να υπολογίζονται για κάθε ομάδα με τον ίδιο τρόπο όπως για τις δύο ομάδες.

Το στατιστικό test θα μπορούσε να έχει $k-1$ βαθμούς ελευθερίας, αν υπήρχαν k ομάδες.

Οι αναλογίες $\frac{O_A}{E_A}$ και $\frac{O_B}{E_B}$ αναφέρονται ως οι σχετικοί δείκτες θνησιμότητας και εκτιμούν την αναλογία του δείκτη θνησιμότητας σε κάθε ομάδα προς τον δείκτη θνησιμότητας που προκύπτει από τον συνδυασμό των δύο ομάδων. Η αναλογία αυτών των δύο σχετικών δεικτών εκτιμάει τον δείκτη θνησιμότητας στην ομάδα A συγκριτικά με αυτόν στην ομάδα B, και μερικές φορές καλείται *αναλογία κινδύνου (hazard ratio)*. Ο *hazard ratio* και η μεταβλητότητα του δείγματος δίνονται από τον τύπο

$$h = \frac{O_A / E_A}{O_B / E_B} \tag{2.16}$$

$$SE[\ln(h)] = \sqrt{\left(\frac{1}{E_A} + \frac{1}{E_B}\right)}$$

Μια εναλλακτική εκτίμηση είναι η

$$h = \exp\left(\frac{O_A - E_A}{V_A}\right) \tag{2.17}$$

$$SE[\ln(h)] = \sqrt{\frac{1}{V_A}}$$

(Machin & Gardner, 1989). Και οι δύο τύποι (2.16) και (2.17) είναι μεροληπτικοί και τα διαστήματα εμπιστοσύνης που βασίζονται σε σταθερά σφάλματα (SE) θα είναι μικρότερα από την ελάχιστη ποσότητα ασφαλείας, όταν το hazard ratio δεν είναι κοντά στη μονάδα. Ο τύπος (2.16) είναι μεροληπτικός και επαρκής για το h λιγότερο από 3, αλλά για μεγαλύτερο hazard ratio ίσως υπολογίζεται ένα προσαρμοσμένο σφάλμα ή ίσως ενδείκνυται μια πιο πολύπλοκη ανάλυση.

Το logrank test μπορεί να επεκταθεί ώστε να λάβει υπόψη μία συμμεταβλητή που χωρίζει την συνολική ομάδα σε στρώματα. Οι ποσότητες στο (2.13) αθροίζονται πάνω στα στρώματα πριν την εφαρμογή των τύπων (2.14) και (2.15). Επομένως συμβολίζοντας τα στρώματα με h , ο τύπος (2.14) γίνεται

$$X_1^2 = \frac{\sum_h O_A - \sum_h E_A}{\sum_h V_A} \quad (2.18)$$

Η στρωματοποίηση αποτελεί συνήθως μία επιλογή όταν η κατασκευή της συμμεταβλητής μπορεί να εκπροσωπηθεί μόνο από λίγα στρώματα.

Όταν υπάρχουν διάφορες μεταβλητές που λαμβάνουμε υπόψη, ή μια συνεχής μεταβλητή η οποία πρακτικά δεν κατηγοριοποιείται, τότε οι μέθοδοι που βασίζονται στη στρωματοποίηση γίνονται δυσλειτουργικοί και αναποτελεσματικοί, και είναι προτιμότερο να χρησιμοποιήσουμε μεθόδους παλινδρόμησης.

Το logrank test είναι ένα μη-παραμετρικό test. Άλλα test μπορούν να επιτευχθούν από την τροποποίηση του Wilcoxon's rank sum test, έτσι ώστε να έχουμε την δυνατότητα να συγκρίνουμε τους χρόνους επιβίωσης για τις δύο ομάδες, στην περίπτωση όπου μερικές στιγμές επιβίωσης έχουν αποκοπεί (Cox & Oakes, 1984).

Το γενικευμένο Wilcoxon test πρωταρχικά προτάθηκε από τον Gehan (1965) και τροποποιήθηκε έτσι ώστε να χρησιμοποιείται για τα σταθμισμένα αθροίσματα του τύπου (2.13).

Η πρόταση του Gehan ήταν ότι ο συντελεστής βαρύτητας να είναι ο συνολικός αριθμός των εν ζώη ατόμων σε κάθε ομάδα. Αυτοί οι συντελεστές βαρύτητας είναι εξαρτημένοι από τις αποκομμένες μεταβλητές και μια εναλλακτική για να αποφύγουμε κάτι τέτοιο είναι να χρησιμοποιήσουμε έναν εκτιμητή για τη συνδυασμένη συνάρτηση

επιβίωσης (*combined survivor function*-Prentice, 1978). Αν καμία από τις παρατηρήσεις δεν ήταν αποκομμένη, τότε το test αυτό θα ήταν ολόιδιο με το Wilcoxon rank sum test. Το logrank test είναι αστάθμιστο- αυτό σημαίνει ότι οι συντελεστές βαρύτητας είναι ίδιοι για κάθε θάνατο.

Συνεπώς, το logrank test δίνει μεγαλύτερη βαρύτητα στον θάνατο προς το τέλος της παρακολούθησης, όπου επιβιώνουν λίγα άτομα, και το γενικευμένο Wilcoxon test έχει την τάση να είναι πιο ευαίσθητο από το logrank test σε καταστάσεις όπου οι αναλογίες κινδύνου (*hazard ratios*) είναι υψηλότερες σε προγενέστερες στιγμές επιβίωσης από ότι μετέπειτα. Το logrank test γίνεται βέλτιστο κάτω από συμμετρικές υποθέσεις κινδύνου (*proportional- hazards assumption*), κάτι που σημαίνει ότι το *hazard ratio* είναι αμετάβλητο (σταθερό) σε όλες τις χρονικές στιγμές επιβίωσης.

Ενδιάμεσα συστήματα συντελεστών βαρύτητας είναι μια δύναμη ξ μεταξύ των τιμών 0 και 1, δηλαδή του αριθμού των εν ζωή ατόμων ή της συνδυασμένης συνάρτησης επιβίωσης. Έτσι για το γενικευμένο Wilcoxon test, το $\xi=1$, για το logrank test το $\xi=0$ και για την τετραγωνική ρίζα, $\xi=\frac{1}{2}$.

Παράδειγμα 1:

Στον Πίνακα 2.3 τα δεδομένα δίνονται από το πλήθος των εν ζωή ασθενών με διάχυτο λέμφωμα (*histiocytic lymphoma*) στο αντίστοιχο στάδιο του καρκίνου. Η επιβίωση μετράται από την ημέρα της εισχώρησης του ατόμου στην κλινική δοκιμή. Υπήρχαν μικρές διαφορές στην επιβίωση μεταξύ των δύο ομάδων θεραπειών, που δεν θα τις λάβουμε υπόψη στο παράδειγμα αυτό.

Οι υπολογισμοί του εκτιμητή γινομένου-ορίου του life-table δίνονται από τον πίνακα 2.4 για την ομάδα που βρίσκεται στο στάδιο 3 ενώ η σύγκριση της επιβίωσης μπορεί να παρασταθεί με γράφημα. Η επιβίωση είναι μεγαλύτερης διάρκειας, κατά μέσο όρο, για τους ασθενείς με καρκίνο στο στάδιο 3 σε σχέση με τους ασθενείς με καρκίνο στο στάδιο 4. Αυτή η διαφορά μπορεί να φανεί ξεκάθαρα με ένα γράφημα ή επίσημα από τα συμπεράσματα του logrank test.

Οι βασικοί υπολογισμοί που είναι απαραίτητοι για το logrank test δίνονται στον Πίνακα 2.5. Εν συντομία, φαίνονται μόνο οι θάνατοι που συνέβησαν στην αρχή και στο τέλος της υπό παρακολούθηση περιόδου. Οι δύο ομάδες υποδηλώνονται με τους δείκτες 3 και 4 αντίστοιχα, αντί A και B που χρησιμοποιήθηκαν στην γενική περιγραφή.

Εφαρμόζοντας τον τύπο (2.14) παίρνουμε

$$\begin{aligned} X_1^2 &= (8-16 \cdot 6870)^2 / 11 \cdot 2471 \\ &= 8 \cdot 6870^2 / 11 \cdot 2471 \\ &= 6 \cdot 71 \quad (p = 0.010) \end{aligned}$$

Για να υπολογίσουμε τον τύπο (2.15) ,υπολογίζουμε πρώτα το E_4 χρησιμοποιώντας την σχέση $O_3 + O_4 = E_3 + E_4$. Επομένως $E_4 = 37.3130$ και

$$\begin{aligned} X_2^2 &= 8 \cdot 6870^2 (1/16 \cdot 6870 + 1/37 \cdot 3130) \\ &= 6 \cdot 54 \quad (p = 0.010) \end{aligned}$$

Οι σχετικοί δείκτες θνησιμότητας είναι $8/16 \cdot 6870 = 0.48$ για την ομάδα στο στάδιο 3 και $46/37 \cdot 3130 = 1.23$ για την ομάδα στο στάδιο 4. Η αναλογία αυτών των δεικτών εκτιμάει τον δείκτη θνησιμότητας στο στάδιο 4 συγκριτικά με αυτόν στο στάδιο 3, δηλαδή $1.23/0.48 = 2.57$. Χρησιμοποιώντας τον τύπο (2.16) , έχουμε $SE[\ln(h)] = 0.2945$ και το 95% διάστημα εμπιστοσύνης για τον hazard ratio είναι $\exp[\ln(2.57) \pm 1.96 \times 0.2945] = 1.44$ με 2.58. Χρησιμοποιώντας τον τύπο (2.17) , ο hazard ratio προκύπτει 2.16 (με 95% διάστημα εμπιστοσύνης 1.21 με 3.88).

Πίνακας 2.3

Πίνακας 2.3 Επιβίωση ασθενών με διάχυτο λέμφωμα hystiocytic αντίστοιχο με το στάδιο του καρκίνου (τα δεδομένα προσαρμόστηκαν από τον McKelvey ., 1976).

Survival (days)	6	19	32	42	42	43*	94	126*
Stage 3	169*	207	211*	227*	253	255*	270*	310*
	316*	335*	346*					
Stage 4	4	6	10	11	11	11	13	17
	20	20	21	22	24	24	29	30
	30	31	33	34	35	39	40	41*
	43*	45	46	50	56	61*	61*	63
	68	82	85	88	89	90	93	104
	110'	134	137	160*	169	171	173	175
	184	201	222	235*	247*	260*	284*	290*
	291*	302*	304*	341*	345*			

*Still alive (censored value).

Πίνακας 2.4

Πίνακας 2.4 υπολογισμός του πολλαπλασιαστή-ορίου εκτιμήτη του life-table για το 3^ο καρκινικό στάδιο, βασισμένο στα δεδομένα του πίνακα 2.3.

Time (days) t_j	Died 4	Living at start of day n'_j	Estimated probability of:		Percentage of survivors at end of day 4
			Death	Survival P'_j	
0	—	19	—	—	100-0
6	1	19	0-0526	0-9474	94-7
19	1	18	- 0-0556	0-9444	89-5
32	1	17	0-0588	0-9412	84-2
42	2	16	0-1250	0-8750	73-7
94	1	13	0-0769	0-9231	68-0
207	1	10	0-1000	0-9000	61-2
253	1	7	0-1429	0-8571	52-5

Πίνακας 2.5

Πίνακας 2.5 Υπολογισμός του logrank test (δεδομένα του πίνακα 2.3) το οποίο συγκρίνει την επιβίωση των ασθενών με ακρίνο στα στάδια 3 και 4.

Days when deaths occured	n'_3	n'_4	d_3	d_4	E(d_3)	var (d_3)
4	19	61	0	1	0.2375	0.1811
6	19	60	1	1	0.4810	0.3606
10	18	59	0	1	0.2338	0.1791
11	18	58	0	— 3	0.7105	0.5278
13	18	55	0	1	0.12466	0.1858
17	18	54	0	1	0.2500	0.1875
19	18	53	1	0	0.2535	0.1892
20	17	53	0	2	0.4857	0.3624
201	10	12	0	1	0.4545	0.2479
207	10	11	1	0	0.4762	0.2494
222	8	11	0	1	0.4211	0.2438
253	7	8	1	0	0.4667	0.2489

2.7 Παραμετρικές μέθοδοι

Σε μελέτες θνησιμότητας η μεταβλητή του ενδιαφέροντος είναι ο χρόνος επιβίωσης. Μια πιθανή προσέγγιση στην ανάλυση είναι να διατυπωθεί μια κατανομή για το χρόνο επιβίωσης και να εκτιμηθούν οι παράμετροι της κατανομής από τα δεδομένα. Αυτή η προσέγγιση συνήθως εφαρμόζεται με το να ξεκινήσουμε με ένα μοντέλο για τον δείκτη θνησιμότητας και σταδιακά να καθορίσουμε τον τύπο της τελικής κατανομής του χρόνου επιβίωσης. Ο δείκτης θνησιμότητας θα ποικίλει συνήθως με τον χρόνο της συμμετοχής στη μελέτη, t , και θα συμβολίζεται από το $\lambda(t)$. Μερικές φορές το $\lambda(t)$ αναφέρεται σαν *συνάρτηση κινδύνου (hazard function)*. Υποθέτουμε ότι η πυκνότητα της πιθανότητας του χρόνου επιβίωσης είναι $f(t)$ και η αντίστοιχη συνάρτηση κατανομής είναι $F(t)$. Τότε, αφού ο δείκτης θνησιμότητας είναι ο δείκτης στον οποίο οι θάνατοι, συμβαίνει να είναι διαχωρισμένοι από το τμήμα του πληθυσμού που είναι εν ζωή, θα έχουμε

$$\begin{aligned}\lambda(t) &= \frac{f(t)}{1-F(t)} \\ &= f(t)/S(t)\end{aligned}\tag{2.19}$$

όπου $S(t) = 1-F(t)$ είναι η αναλογία επιβίωσης και αναφέρεται ως *survivor function*.

Η εξίσωση (2.19) παρέχει τη δυνατότητα στα $f(t)$ και $S(t)$ να εξειδικεύονται, υπό τον όρο ότι γνωρίζουμε το $\lambda(t)$. Η γενική λύση επιτυγχάνεται από την ενσωμάτωση του (2.19) με την εκτίμηση του t και σημειώνουμε ότι το $f(t)$ είναι η παράγωγος της $F(t)$.

Θα θεωρήσουμε βέβαιες υποθέσεις. Ο απλούστερος τύπος είναι αυτός που ο δείκτης θνησιμότητας είναι μια σταθερά, δηλαδή $\lambda(t)=\lambda$ για όλα τα t . Τότε:

$$\lambda t = -\ln[S(t)]\tag{2.20}$$

Αυτό σημαίνει ότι

$$S(t) = \exp(-\lambda t)$$

Ο χρόνος επιβίωσης ακολουθεί μια εκθετική κατανομή με μέσο $\frac{1}{\lambda}$. Αν αυτή η κατανομή είναι κατάλληλη, τότε, από τον τύπο (2.20), η γραφική παράσταση του λογαρίθμου της συνάρτησης επιβίωσης του χρόνου είναι μια ευθεία γραμμή που περνάει από την αρχή των αξόνων.

Τα δεδομένα μίας ομάδας ατόμων αποτελούνται από έναν αριθμό θανάτων με γνωστές συναρτήσεις επιβίωσης κι έναν αριθμό ατόμων εν ζωή για τους οποίους το αποκομμένο μήκος της επιβίωσης είναι γνωστό. Αυτά τα δεδομένα μπορούν να χρησιμοποιηθούν για να εκτιμήσουν το λ , χρησιμοποιώντας τη μέθοδο της μέγιστης πιθανοφάνειας (maximum likelihood).

Για μια συγκεκριμένη τιμή του λ , η likelihood αποτελείται από το γινόμενο των όρων $f(t)$ για τους θανάτους και $S(t)$ για τους επιζώντες. Τόσο η εκτίμηση της μέγιστης πιθανοφάνειας του λ , όσο και το σταθερό σφάλμα της εκτιμήτριας και ένα τεστ σημαντικότητας για οποιαδήποτε υποθετική τιμή επιτυγχάνονται χρησιμοποιώντας την γενική μέθοδο της μέγιστης πιθανοφάνειας, αν και, σ' αυτήν την απλή περίπτωση, η λύση μπορεί να επιτευχθεί απευθείας χωρίς προσέγγιση.

Ο κύριος περιορισμός του εκθετικού μοντέλου είναι η εικασία ότι ο δείκτης θνησιμότητας είναι ανεξάρτητος του χρόνου. Θα ήταν ουτοπία να περιμένουμε ότι αυτή η εικασία μπορεί να διατηρηθεί εκτός από κάποιες μικρές χρονικές περιόδους. Ένας τρόπος να ξεπεράσουμε αυτόν τον περιορισμό είναι να χωρίσουμε τις περιόδους παρακολούθησης σε μικρότερα διαστήματα και να θεωρήσουμε ότι ο δείκτης κινδύνου (hazard rate) είναι σταθερός μέσα σε κάθε διάστημα, αλλά διαφορετικός για τα διαφορετικά διαστήματα (Holford, 1976).

Μια άλλη μέθοδος για να αποφευχθεί η εικασία ότι ο hazard είναι σταθερός είναι να χρησιμοποιήσουμε ένα διαφορετικό παραμετρικό μοντέλο του hazard rate. Ένα τέτοιο μοντέλο είναι το Weibull καθορισμένο από την σχέση:

$$\lambda(t) = a\gamma t^{\gamma-1} \quad (2.21)$$

όπου το γ είναι μεγαλύτερο της μονάδας. Αυτό το μοντέλο έχει αποδειχθεί κατά καιρούς εφαρμόσιμο -ανάλογα με την ηλικία- σε περίπτωση καρκίνου σε ανθρώπους (Cook 1969), και με την χρονική στιγμή καρκινογένεση σε πειράματα ζώων (Pike, 1966).

Ένα τρίτο μοντέλο είναι αυτό που ο κίνδυνος αυξάνεται εκθετικά με την ηλικία, το οποίο είναι:

$$\lambda(t) = a \exp(\beta t) \quad (2.22)$$

Αυτός ο τύπος είναι ο κίνδυνος Gompertz και περιγράφει ξεκάθαρα τον δείκτη θνησιμότητας από όλες τις αιτίες στους ενήλικους. Συγκεκριμένα είναι ένα μοντέλο στο

οποίο οι χρονικές στιγμές θανάτου είναι λογαριθμικά-κανονικά κατανεμημένες και έχουν χρησιμοποιηθεί, αλλά διαθέτουν το μειονέκτημα ο σχετικός δείκτης κινδύνου τους να αρχίζει κάποιες φορές να μειώνεται.

2.8 Αναλογικά μοντέλα κινδύνου και μοντέλα παλινδρόμησης

Θα ήταν ασυνήθιστο να αναλύσουμε μία απλή ομάδα ομοιογενών ατόμων, αλλά η βασική αυτή μέθοδος ίσως επεκτείνεται στο να ανταπεξέλθει σε πιο ρεαλιστικές καταστάσεις με την μοντελοποίηση του δείκτη κινδύνου ώστε να εκπροσωπήσει την εξάρτηση σε μεταβλητές καταχωρημένες για κάθε υποκείμενο καθώς επίσης και για το χρόνο.

Για παράδειγμα, σε μία κλινική δοκιμή θα προτεινόταν ο δείκτης κινδύνου να ήταν εξαρτημένος από τη θεραπεία η οποία θα μπορούσε να εκπροσωπείται από μία ή περισσότερες πλασματικές μεταβλητές. Από την άλλη, αν ο αριθμός των προγνωστικών μεταβλητών ήταν γνωστός, τότε ο δείκτης θνησιμότητας θα μπορούσε να εκφραστεί σαν μια συνάρτηση των μεταβλητών. Γενικά, ο δείκτης θνησιμότητας θα μπορούσε να γραφεί σαν μια συνάρτηση τόσο του χρόνου όσο και των συνδιακυμάνσεων οι οποίες είναι όπως και το $\lambda(t, x)$, όπου το x αντιπροσωπεύει το σύνολο των διακυμάνσεων (x_1, x_2, \dots, x_p) .

Οι Zippin και Armitage (1966) θεώρησαν μια προγνωστική μεταβλητή, x , τον λογάριθμο του αριθμού του white blood count (τεστ για τη μετατροπή των λευκών αιμοσφαιρίων στο αίμα) και μια εκθετική κατανομή επιβίωσης, με

$$\lambda(t, x) = (\alpha + \beta x)^{-1} \quad (2.23)$$

Ο μέσος χρόνος επιβίωσης ήταν συνεπώς γραμμικός ως προς x . Ανάλυση που αποτελείται από την εκτίμηση των α και β . Ένα μειονέκτημα της αναπαράστασης είναι ότι ο δείκτης κινδύνου γίνεται αρνητικός για μεγάλες τιμές του x (γιατί ο β είναι αρνητικός). Ένα εναλλακτικό μοντέλο που αποφεύγει αυτό το μειονέκτημα προτάθηκε από τον Glasser (1967), και είναι το εξής:

$$\lambda(t, x) = \alpha \cdot \exp(\beta x) \quad (2.24)$$

όπου ο λογάριθμος του μέσου χρόνου επιβίωσης ήταν συνεπώς γραμμικός ως προς x .

Τόσο το μοντέλο του (2.23) όσο και του (2.24) εσωκλείουν την εικασία ότι ο δείκτης θνησιμότητας είναι ανεξάρτητος του χρόνου. Γενικά, ο κίνδυνος θα βασίζεται στον χρόνο και σε μια οικογένεια μοντέλων ισοδύναμα γραμμικής από τον τύπο

$$\lambda(t,x) = \lambda_o(t) \cdot \exp(\beta^T \cdot x) \quad (2.25)$$

όπου $\beta^T x$ είναι ένα πλέγμα αναπαράστασης της συνάρτησης παλινδρόμησης $\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$ και $\lambda_o(t)$ είναι ένα χρονικά ανεξάρτητο μέρος του κινδύνου. Ο όρος $\lambda_o(t)$ θα μπορούσε να αναπαραστήσει οποιοδήποτε από τα μοντέλα που θεωρήθηκε στην προηγούμενη ενότητα ή σε άλλες παραμετρικές συναρτήσεις του t . Η εξίσωση (2.25) είναι ένα μοντέλο παλινδρόμησης (regression model) με όρους διακύμανσης. Αναφέρεται επίσης ως ένα αναλογικού κινδύνου μοντέλο αφού οι κίνδυνοι (hazards) για τα διαφορετικά σύνολα μεταβλητών παραμένουν στην ίδια αναλογία για όλα τα t . Τα δεδομένα μπορούν να αναλυθούν παραμετρικά χρησιμοποιώντας τον τύπο (2.25) που παρέχει πληροφορίες από κάποιους ειδικούς τύπους του $\lambda_o(t)$. Οι παράμετροι του $\lambda_o(t)$ και επίσης, οι συντελεστές παλινδρόμησης, β , θα μπορούσαν να εκτιμηθούν. Το συμπέρασμα θα μπορούσε να προκύψει λαμβάνοντας υπόψη την εκτίμηση b του β , ενώ οι παράμετροι του $\lambda_o(t)$ θα μπορούσαν να μην έχουν άμεσο ενδιαφέρον. Ένας άλλος τρόπος αναπαράστασης της επίδρασης των διακυμάνσεων είναι να υποθέσουμε ότι η κατανομή του χρόνου επιβίωσης αλλάζει από τον πολλαπλασιασμό της κλίμακας χρόνου με το $\exp(\beta \alpha^T x)$, το οποίο σημαίνει ότι ο λογάριθμος του χρόνου επιβίωσης αυξάνεται από το $\beta \alpha^T x$. Το hazard θα μπορούσε να γραφεί:

$$\lambda(t,x) = \lambda_o [t \cdot \exp(-\beta \alpha^T x) \cdot \exp(-\beta \alpha^T x)] \quad (2.26)$$

Αυτό αναφέρεται ως *accelerated failure time model*.

Για τις εκθετικές κατανομές, $\lambda_o(t)=\lambda$ οι τύποι (2.25) και (2.26) είναι ισοδύναμοι, όπου $\beta \alpha = -\beta$, με συνέπεια το accelerated failure time model να είναι επίσης ένα αναλογικό μοντέλο κινδύνου. Το ίδιο ισχύει και για το Weibull (2.21), όπου $\beta \alpha = \frac{-\beta}{\gamma}$ αλλά δεν θα μπορούσαμε να ισχυριστούμε ότι γενικά το accelerated failure time model είναι ένα αναλογικό μοντέλο κινδύνου (proportional hazard model).

Παρόλα αυτά, ίσως είναι δύσκολο σε κάποιες περιπτώσεις να καθοριστεί αν είναι περισσότερο κατάλληλο ένα proportional- hazard μοντέλο ή ένα accelerated failure time μοντέλο, αλλά τότε τα δυο μοντέλα ίσως δίνουν παρόμοια συμπεράσματα για την επίδραση των διακυμάνσεων (Solomon, 1984).

Οι διαδικασίες για τα εφαρμοσμένα μοντέλα του τύπου που συζητήθηκε παραπάνω είναι διαθέσιμα σ' έναν αριθμό στατιστικών πακέτων. Για παράδειγμα, ένα εύρος παραμετρικών μοντέλων, συμπεριλαμβανομένων της εκθετικής, Weibull και log-normal, εφαρμόζεται με τη χρήση PROC LIFEREG σε πρόγραμμα SAS.

Αναλογικό μοντέλο κινδύνου του Cox(Cox's proportional-hazards model)

Αφού συμβαίνει συχνά ένας κατάλληλος παραμετρικός τύπος του $\lambda_o(t)$ να είναι άγνωστος, και σε οποιαδήποτε περίπτωση, όχι πρωταρχικού ενδιαφέρον, θα ήταν πιο πρακτικό αν δεν χρειαζόταν να αντικατασταθεί το $\lambda_o(t)$ από οποιονδήποτε άλλο συγκεκριμένο τύπο στην σχέση (2.25). Αυτή η προσέγγιση εισήχθη από τον Cox (1972). Το μοντέλο τότε θεωρείται μη παραμετρικό ως προς τους όρους των διακυμάνσεων. Η εκτίμηση του β και τα συμπεράσματα αναπτύχθηκαν με βάση τις πληροφορίες που παρέχονται σε κάθε χρονική στιγμή που συμβαίνει ένας θάνατος. Έστω ότι ένας θάνατος συμβαίνει τη χρονική στιγμή t_j , και υποθέτουμε ότι υπήρχαν n_j εν ζωή άτομα ακριβώς πριν τον χρόνο t_j , των οποίων οι τιμές του x για εκείνα τα άτομα είναι $x_1, x_2, \dots, x_{n'_j}$ και ότι τα άτομα τα οποία πέθαναν, υποδηλώθηκαν, χωρίς βλάβη της γενικότητας από τον δείκτη 1. Το σύνολο των n'_j ατόμων σε κίνδυνο αναφέρεται ως *the risk set*. Το ρίσκο του θανάτου τη χρονική στιγμή t_j για κάθε άτομο στο σύνολο κινδύνου δίνεται από τον τύπο (2.25). Αυτό δεν παρέχει απόλυτα μέτρα κινδύνου, αλλά παρέχει τον σχετικό κίνδυνο για κάθε άτομο, αφού αν και το $\lambda_o(t)$ είναι άγνωστο, είναι ίδιο για κάθε άτομο. Επομένως, η πιθανότητα ότι αυτός ο παρατηρούμενος θάνατος στο t_j ήταν το άτομο που πέθανε αυτή τη χρονική στιγμή δίνεται από τον τύπο

$$p_j = \exp(\beta^T x_1) / \sum \exp(\beta^T x_i) \quad (2.27)$$

όπου το άθροισμα αναφέρεται σε όλα τα μέλη του συνόλου σε κίνδυνο. Παρόμοιοι όροι διαχωρίζονται για κάθε χρονική στιγμή που ένας θάνατος συμβαίνει και συνδυάζονται

για να φορμάρουν την πιθανοφάνεια. Τεχνικά, αυτό καλείται *partial likelihood*, αφού τα στοιχεία διαχωρίζονται με βάση το χρόνο θανάτου και την σύνθεση των υπό κίνδυνο συνόλων, τις στιγμές αυτές. Οι ακριβείς χρονικές στιγμές στις οποίες συμβαίνουν οι θάνατοι δεν παίζουν τόσο καθοριστικό ρόλο, αφού αυτό που καθορίζει το σύνολο κινδύνου είναι η σειρά των θανάτων και των αποκομμένων ατόμων. Επομένως, η μέθοδος έχει ομοιότητες με τα μη παραμετρικά τεστ. Έχει επιπλέον ομοιότητες με το *logrank test*, το οποίο είναι επίσης εξαρτώμενο από το σύνολο κινδύνου.

Το μοντέλο τροποποιείται από την μέθοδο της μέγιστης πιθανοφάνειας κάτι που το κάνει χρήσιμο σε μεγαλύτερο πλήθος περιπτώσεων, ειδικά σε στατιστικά λογισμικά πακέτα, όπως *PROC PHREG* σε *SAS*.

ΚΕΦΑΛΑΙΟ 3

Κλινικές δοκιμές (clinical trials)

3.1 Εισαγωγή

Οι κλινικές δοκιμές ελέγχουν πειράματα με σκοπό να συγκρίνουν την αποτελεσματικότητα και την ασφάλεια ασθενών όταν υπάρχει ιατρική παρέμβαση. Αυστηρά ο όρος κλινική (*clinical*) υπαινίσσεται ότι οι ασθενείς υποφέρουν από κάποια ή περισσότερες συγκεκριμένες ασθένειες και πράγματι οι περισσότερες κλινικές δοκιμές διεξάγονται με την συμμετοχή των ασθενών και συγκρίνονται θεραπείες με σκοπό να βελτιωθούν οι συνθήκες τους. Αλλά με μία πιο ευρεία έννοια μπορούμε να συμπεριλάβουμε στις κλινικές δοκιμές ελεγχόμενες προληπτικές δοκιμές όπως για παράδειγμα εμβόλια σε άτομα που δεν υποφέρουν ακόμα από την υπό μελέτη ασθένεια, καθώς και δοκιμές σε διοικητικές περιπτώσεις που αφορούν την ιατρική φροντίδα όπως οι επιλογή της οικιακής ή νοσοκομειακής φροντίδας για έναν συγκεκριμένο τύπο ασθενών.

Αφού λοιπόν μία κλινική δοκιμή είναι ένα πείραμα υπόκειται σε βασικές αρχές πειραματισμών όπως η τυχαιοποίηση, η αναπαραγωγή και ο έλεγχος της μεταβλητότητας. Συνεπώς, η αλήθεια ότι όταν οι πειραματικές μονάδες είναι ανθρώπινες οντότητες απαιτούνται ειδικές εικασίες και δημιουργούν ανάγκη για την βελτίωση πολλών μοναδικών προβλημάτων που προκύπτουν. Αρχικά στις κλινικές δοκιμές οι ασθενείς συμμετέχουν συνήθως από μία χρονική περίοδο και πάνω και οι σχετικές παρατηρήσεις προκύπτουν σταδιακά. Η πραγματικότητα αυτή περιορίζει την ευκαιρία να εκμεταλλευτούμε περισσότερο πολύπλοκους τύπους πειραματικού σχεδίου (*experimental design*). Δεύτερον υπάρχουν καλύτερες πιθανότητες για την εκτίμηση της ανταπόκρισης στην θεραπεία από τα εργαστηριακά πειράματα. Τρίτον και σημαντικότερον, οποιαδήποτε πρόταση για κλινική δοκιμή πρέπει να είναι εξουχιστικά περιεργασμένη από ηθικής πλευράς, έτσι ώστε για κανέναν γιατρό να μη τεθεί το ζήτημα αν θα επιτρέψει σε ασθενή που βρίσκεται υπό την φροντίδα του να λάβει θεραπεία η οποία μπορεί να είναι και υποδεέστερη αυτής που λάμβανε μέχρι τώρα. Υπάρχουν όμως περιπτώσεις στις οποίες τα σχετικά πλεονεκτήματα της θεραπείας δεν προέρχονται από ξεκάθαρες μεθόδους. Στην περίπτωση αυτή οι γιατροί συμφωνούν σε τυχαίο καταμερισμό τουλάχιστον μέχρι να επιλυθεί το θέμα.

Οι δοκιμές επιβάλλονται σαν επιτακτικές ερευνητικές μελέτες με τυχαία ανάθεση και αναφέρονται ως *Phase III*. Στην ανάπτυξη των φαρμάκων οι *Phase I* μελέτες είναι έρευνες κυμαινόμενης δόσης, συχνά με υγιείς εθελοντές. Οι *Phase II* δοκιμές είναι μικρές εξεταστικές μελέτες σε ασθενείς, σχεδιασμένες με τέτοιο τρόπο ώστε να επιλέγουν ασθενείς επαρκώς υποσχόμενους στο να διασφαλίσουν την δημιουργία μιας μεγαλύτερης *Phase III* δοκιμής. Οι *Phase IV* μελέτες ασχολούνται με την παρακολούθηση της ασφάλειας του φαρμάκου μετά την κυκλοφορία του στην αγορά.

Η οργάνωση μιας κλινικής δοκιμής απαιτεί προσεκτικό σχεδιασμό. Αυτό ισχύει κυρίως για τις πολυκεντρικές μελέτες οι οποίες εφαρμόζονται ευρέως για τις μελέτες χρόνιων ασθενειών όπου απαιτείται πολύ μεγάλος αριθμός ασθενών.

Οι σκοποί και οι μέθοδοι των δοκιμών πρέπει να περιγράφονται με λεπτομέρεια σε έγγραφο που ονομάζεται πρωτόκολλο (*protocol*). Το πρωτόκολλο περιέχει πολλές ιατρικές και διοικητικές λεπτομέρειες για το συγκεκριμένο υπό μελέτη πρόβλημα. Επιπλέον στο πρωτόκολλο πρέπει να δηλώνεται ξεκάθαρα ο σκοπός της δοκιμής, οι τύποι των ασθενών που συμμετέχουν και τα θεραπευτικά μέτρα που χρησιμοποιούνται. Επίσης πρέπει να αναφέρεται το πλήθος των ασθενών, η διάρκεια της περιόδου εισροής ασθενών στην κλινική δοκιμή και όπου κριθεί απαραίτητο η διάρκεια της παρακολούθησης.

3.2 Κλινικές δοκιμές Phase I και Phase II

Η χορήγηση ενός νέου φαρμάκου σε ασθενείς συνοδεύεται πάντα από ένα μεγάλο αριθμό ερευνών και βελτιώσεων, συμπεριλαμβανομένων των φαρμακολογικών και τοξικολογικών μελετών σε ζώα. Οι μελέτες αυτές καθιστούν ικανούς τους ερευνητές να προβλέψουν τον τύπο και την έκταση της τοξικότητας που αναμένεται ύστερα από την χορήγηση συγκεκριμένων δόσεων σε ασθενείς. Παρέχεται επιπλέον η δυνατότητα να γίνουν κλινικές φαρμακολογικές μελέτες και να παρατηρηθούν τυχόν τοξικές επιδράσεις ώστε να καθοριστεί τελικώς η κατάλληλη δοσολογία.

Οι *Phase I* μελέτες συχνά εφαρμόζονται σε ανθρώπους που συμμετέχουν εθελοντικά, αλλά στην προσπάθεια βελτίωσης φαρμάκων για θεραπείες βέβαιων καταστάσεων όπως ο καρκίνος κάτι τέτοιο δεν είναι επαρκές. Στις περιπτώσεις αυτές είναι ίσως απαραίτητο να συμπεριλάβουμε ασθενείς (άτομα δηλαδή που έχουν ήδη νοσήσει) αφού οι τοξικές αντιδράσεις τους διαφέρουν από τα υγιή άτομα. Βασικός σκοπός στον σχεδιασμό μιας *Phase I* δοκιμής είναι να εκτιμήσουμε την δόση που να ανταποκρίνεται σ' ένα μέγιστο αποδεκτό επίπεδο τοξικότητας (μέγιστη επιτρεπτή δόση-*maximum tolerated dose MTD*). Το *MTD* μπορεί να καθοριστεί ως η αναλογία των ανθρώπων που φαίνεται να είχαν μια συγκεκριμένη αντίδραση, ή σαν το μέσο επίπεδο μιας ποσοτικής μεταβλητής όπως η καταμέτρηση των λευκών αιμοσφαιρίων στο αίμα. Ο αριθμός των ασθενών πιθανόν να είναι μικρός σε εύρος 10-50 ατόμων.

Μία προσέγγιση για να ξεκινήσει ο σχεδιασμός της μελέτης είναι να αρχίσουμε με μία πολύ μικρή δόση, καθορισμένη με βάση τα συμπεράσματα από τα πειράματα σε ζώα ή από ανθρώπινες μελέτες που σχετίζονται με τα συγκεκριμένα φάρμακα. Οι δόσεις, εφαρμοσμένες σε πολύ μικρό πλήθος ατόμων, κλιμακώνονται μέχρις ότου να επιτευχθεί το επίπεδο τοξικότητας που θέλουμε (Storer, 1989). Αυτή η στρατηγική είναι παρόμοια με την μέθοδο πάνω-κάτω (*up-and-down method*) για την κβαντική βιοανάλυση, αλλά οι κανόνες για τη αλλαγή της δόσης πρέπει να εξασφαλίσουν ότι το επίπεδο που στοχεύουμε σπάνια υπερβαίνεται. Αυτός ο τύπος σχεδιασμού παρέχει μία πρόχειρη εκτίμηση του *MTD* η οποία ίσως χρειάζεται τροποποίηση όταν οι περεταίρω μελέτες ολοκληρωθούν.

Μία άλλη προσέγγιση είναι η μέθοδος συνεχούς επαναξιολόγησης (continual reassessment method-CRM) , σύμφωνα με την οποία εφαρμόζονται επιτυχημένες δοσολογίες σε κάθε άτομο χωριστά ,και το MTD εκτιμάται σε κάθε επίπεδο από το στατιστικό μοντέλο συσχετίζοντας την ανταπόκριση με την δόση. Βάση των προηγούμενων παρατηρήσεων επιλέγονται επιτυχημένες δόσεις να είναι κοντά στην εκτίμηση του MTD και συνεπώς θα έχουν την τάση να μαζεύονται γύρω από την πραγματική τιμή (αν και πάλι με ένα τυχαίο σφάλμα).

Σε μια δοκιμή *Phase II* η έμφαση δίνεται στην αποτελεσματικότητα, αν και πάντα λαμβάνεται υπόψη η ασφάλεια. Μία δοκιμή που περιλαμβάνει κάποιες οπτικές για την επιλογή της δοσολογίας καθώς επίσης και για την αξιολόγηση της αποτελεσματικότητας καλείται *Phase I/II* .Οι δοκιμές *Phase II* πραγματοποιούνται με την συμμετοχή ασθενών που υποφέρουν από μια ασθένεια που ταξινομείται βάση των φαρμάκων. Ο σκοπός είναι να καταλάβουμε εάν το φάρμακο είναι αποτελεσματικό στο να διασφαλίσει μια ευρεία κλίμακα για την δοκιμή *Phase III*. Με αυτή την λογική θεωρείται σαν μια διαδικασία για να επιλέξουμε, από έναν αριθμό υποψηφίων φαρμάκων, εκείνα με την ισχυρότερη εικασία ότι θα μας πάνε σε μία δοκιμή *Phase III*.

Οι δοκιμές *Phase II* ολοκληρώνονται σχετικά γρήγορα ,και η αξιολόγηση πρέπει να εκτιμάται από μία γρήγορη απόκριση. Σε καταστάσεις, όπως η θεραπεία του καρκίνου, που το θέμα είναι η επιβίωση του ασθενή, θα είναι απαραίτητο να χρησιμοποιήσουμε πιο γρήγορα διαθέσιμα μέτρα ,όπως για παράδειγμα την έκταση της συρρίκνωσης του όγκου ή την ύφεση των συμπτωμάτων.

3.3 Σχεδιάζοντας μία *Phase III* δοκιμή

Μία δοκιμή *Phase III* συνήθως οργανώνεται και χρηματοδοτείται από μία φαρμακευτική εταιρεία καθώς το τελικό συμπέρασμα υποβάλλεται στην κατάλληλη ρυθμιστική αρχή για να δοθεί έγκριση του φαρμάκου και να προωθηθεί τελικά στην αγορά. Ίσως ,εναλλακτικά , να είναι ένα κομμάτι έρευνας υπό την αιγίδα ενός διεθνούς φαρμακευτικού οργανισμού έρευνας. Ίσως να ενδιαφέρει περισσότερο η ιατρική διαδικασία παρά η χρήση του φαρμάκου. Σε οποιαδήποτε περίπτωση στην εκτίμηση της αποτελεσματικότητας είναι πιθανό να είναι πρωταρχικής σημασίας και η ασφάλεια της διαδικασίας και συνεπώς να απαιτηθεί πολύ προσεκτικό σχέδιο και ομοίως εκτέλεση.

Οι σκοποί έχουν περιγραφεί με πολλούς διαφορετικούς τρόπους. Μία προσέγγιση είναι να θεωρήσουμε μία δοκιμή σαν μία διαδικασία επιλογής, στην οποία οι ερευνητές αναζητούν την επιλογή μιας καλύτερης , ή της καλύτερης θεραπείας από ένα σύνολο πιθανών θεραπειών για μία συγκεκριμένη ασθένεια. Αυτό οδηγεί στην πιθανή χρήση της θεωρίας αποφάσεων , στην οποία προσδιορίζονται οι συνέπειες της επιλογής ή της απόρριψης συγκεκριμένων θεραπειών. Κάτι τέτοιο φαντάζει τόσο απλό , αφού η δημοσιοποίηση των αποτελεσμάτων των δοκιμών σε απευθείας υιοθέτηση της επιθυμητής θεραπείας από την ιατρική κοινότητα , και οι συνέπειες οποιουδήποτε συγκεκριμένου συνόλου αποτελεσμάτων είναι υπερβολικά δύσκολο να εκτιμηθεί.

Ένας λιγότερο φιλόδοξος σκοπός είναι να παρέχουμε αξιόπιστα επιστημονικά στοιχεία των συγκρινόμενων αποτελεσμάτων , έτσι ώστε οι ερευνητές και οι άλλοι επαγγελματίες να ενημερωθούν για τις επιλογές. Μία πολύ χρήσιμη διάκριση μεταξύ

επεξηγηματικών(explanatory) και πραγματικών(pragmatic) δοκιμών για τις κλινικές δοκιμές παρουσιάστηκε από τους Schwartz και Lellouch (1967). Μία επεξηγηματική δοκιμή προτίθεται να είναι σχεδόν ανάλογη με ένα εργαστηριακό πείραμα αλλά με τα πλαίσια του συντάγματος της θεραπείας καθορισμένα πολύ αυστηρά. Μία πραγματική δοκιμή ,αντίθετα, στοχεύει να εφαρμόσει τις περισσότερο σκληρές συνθήκες μιας συνηθισμένης ιατρικής πρακτικής. Η κατανομή έχει σημαντικές συνέπειες για τον σχεδιασμό και την ανάλυση των δοκιμών.

Στις περισσότερες δοκιμές *Phase III* οι θεραπείες συγκρίνονται με παράλληλες ομάδες ασθενών , σε καθεμία από τις οποίες ο ασθενής λαμβάνει μόνο μία από τις θεραπείες που είναι υπό σύγκριση. Αυτό είναι απαραίτητο ξεκάθαρα όταν η ασθένεια και η εκτίμηση της απόκρισης απαιτεί μία μεγάλη χρονική περίοδο παρακολούθησης . Σε μερικές δοκιμές για την άμεση ανακούφιση του πόνου χρόνιων ασθενειών είναι πιθανό να διενεργήσουμε μία διασταυρωμένη μελέτη ,στην οποία οι ασθενείς θα λαμβάνουν διαφορετικές θεραπείες για διαφορετικές περιπτώσεις.

Σε μερικές κλινικές μελέτες που καλούνται ισοδύναμες δοκιμές (equivalent trials) ,ο σκοπός δεν είναι να εντοπίσουμε πιθανές διαφορές στην αποτελεσματικότητα , αλλά το να δείξουμε ότι οι αυτές οι θεραπείες είναι ,εντός ισχυρά περιορισμένων ορίων, εξίσου αποτελεσματικές. Στις μελέτες *Phase I* και *Phase II* η ερώτηση είναι αν διαφορετικές (χημικές) συνθέσεις που χορηγούνται στο ίδιο άτομο παράγουν επίπεδα serum τα οποία είναι ουσιαστικά ίδια. Σε μία δοκιμή *Phase III* ένα νέο φάρμακο συγκρίνεται με ένα καθιερωμένο φάρμακο που ήδη χρησιμοποιείται, με την ελπίδα ότι η κλινική απόκριση είναι παρόμοια ή τουλάχιστον όχι χειρότερη και ότι υπάρχουν λιγότερο σοβαρές αρνητικές επιδράσεις.

Το πρωτόκολλο

Οι ερευνητές πρέπει να συντάξουν, εκ των προτέρων, ένα λεπτομερειακό σχέδιο μελέτης, που θα τεκμηριώνεται στο πρωτόκολλο το οποίο πρέπει να καλύπτει τουλάχιστον τα παρακάτω θέματα.

- Σκοπός και κίνητρο της δοκιμής.
- Περίληψη της πρόσφατης βιβλιογραφίας που αφορά την ασφάλεια και αποτελεσματικότητα των θεραπειών.
- Κατηγορία (τύπος) των ασθενών που θα εισαχθούν .
- Τα σχέδια θεραπειών που θα χορηγηθούν.
- Οι μεταβλητές που θα χρησιμοποιούνται για συγκρίσεις ασφαλείας και αποτελεσματικότητας.
- Οι διαδικασίες τυχαιοποίησης.
- Το προτεινόμενο πλήθος ασθενών και όπου είναι απαραίτητο το χρονικό διάστημα της παρακολούθησης .
- Συνοπτική περιγραφή της προτεινόμενης ανάλυσης.
- Οι διαδικασίες παρακολούθησης.
- Υπόθεση και περιγραφή τύπου.
- Διακανονισμός που στοχεύει στο να επιτευχθεί η συγκατάθεση του ασθενούς.
- Συμφωνίες για γραπτή αναφορά του και δημοσίευσης.

Ο καθορισμός των ασθενών

Μία κλινική δοκιμή ασχολείται με την θεραπεία ασθενών με κάποιες ιδιαίτερες ιατρικές συνθήκες , η γενική φύση των οποίων είναι συνήθως σ' ένα πρώιμο στάδιο σχεδιασμού. Η λεπτομέρεια της υγείας των ασθενών ίσως να μην είναι τόσο ξεκάθαρη. Πρέπει το φύλο και η ηλικία των ασθενών να είναι περιοριστική; Πρέπει η σοβαρότητα της ασθένειας να είναι αυστηρά καθορισμένη; Αυτά τα κριτήρια εκλεξιμότητας πρέπει να θέτονται εκ των προτέρων για κάθε δοκιμή ,και γενικά, σε συνδυασμό με την αναπαραγωγή του πειραματικού σχεδίου, πρέπει να λαμβάνονται υπόψη. Υπάρχει μία σύγκρουση μεταξύ της επιθυμίας του να επιτευχθεί ομοιογένεια των πειραματικών ατόμων και της επιθυμίας του να καλυφθεί ένα μεγάλο εύρος συνθηκών. Είναι συνήθως προτιμότερο να τείνουμε προς την κατεύθυνση της ανεκτικότητας όσον αφορά τον καθορισμό των κριτηρίων εισόδου. Αυτό όχι μόνο θα αυξήσει τον αριθμό των ασθενών , αλλά θα επιτρέψει να γίνουν ξεχωριστά συγκρίσεις για διάφορες κατηγορίες ασθενών. Ούτως ή άλλως η εισαγωγή ενός μεγάλου φάσματος ασθενών στην μελέτη σε καμία περίπτωση δεν αποτρέπει τον διαχωρισμό τους σε πιο ομοιογενείς υποομάδες για την ανάλυση των συμπερασμάτων. Παρόλα αυτά, οι συγκρίσεις που βασίζονται σε υποομάδες είναι λιγότερο πιθανό να εντοπίσουν πραγματικές διαφορές μεταξύ των επιδράσεων των θεραπειών σε σχέση με αυτές που βασίζονται σε ολόκληρο το σύνολο των δεδομένων. Υπάρχει επιπλέον ο κίνδυνος του ότι αν γίνουν τόσες πολλές συγκρίσεις σε διαφορετικές υποομάδες , ένα ή περισσότερα από τα τέστ θα δώσουν σημαντικά αποτελέσματα καθαρά από τύχη. Πρέπει επομένως οποιαδήποτε υποομάδα με έναν εκ των προτέρων ισχυρισμό για προσοχή να αναφέρεται στο πρωτόκολλο και να δίνεται ιδιαίτερη σημασία στον σχεδιασμό της δοκιμής της.

Ο καθορισμός των θεραπειών

Και σε αυτή την περίπτωση, οι θεραπευτική αγωγές με τις οποίες θα συγκριθεί η υπό μελέτη θεραπεία είναι γνωστές από την αρχή. Πρέπει όμως αυτές να είναι καθορισμένες και τυποποιημένες μέχρι την τελευταία λεπτομέρεια; Σε πολυκεντρικές δοκιμές τα διάφορα κέντρα ίσως επιθυμούν να υιοθετηθούν μικρές αλλαγές στην ίδια φαρμακευτική αγωγή ή να χρησιμοποιήσουν διαφορετική επακόλουθη θεραπεία. Είναι συχνά καλύτερο να επιτρέπονται να εισαχθούν στη μελέτη οι αλλαγές αυτές, ιδιαίτερα όταν στην ιατρική πρακτικά κάτι τέτοιο είναι κοινώς αποδεκτό, παρά να παρουσιαστεί ένας βαθμός τυποποίησης ο οποίος ίσως να μην είναι αποδεκτός ή να διαπιστωθεί κατά την διάρκεια της δοκιμής ότι δεν λειτουργεί.

Με πολλά θεραπευτικά μέτρα ,είναι κοινή πρακτική να ποικίλει το λεπτομερειακό σχέδιο σύμφωνα με τη συνθήκη του κάθε ασθενούς. Η δοσολογία του φαρμάκου ,για παράδειγμα, ίσως να εξαρτάται από την θεραπευτική απόκριση ή από το μέγεθος των επιδράσεων. Στις δοκιμές τέτοιων θεραπειών υπάρχει μια ισχυρή περίπτωση για διατήρηση της ευελιξίας. Σε πολλές δοκιμές έχει διαπιστωθεί μετά το πέρας της θεραπείας ότι η θεραπευτική αγωγή που ακολουθήθηκε ήταν υπερβολικά αυστηρή. Η συμβουλή που δίνεται σε αυτήν και στην προηγούμενη υποενότητα είναι ότι ο καθορισμός τόσο των ασθενών όσο και της θεραπείας πρέπει να έχει την τάση της ευρύτητας παρά της στενότητας λόγω των πραγματικών αναγκών μιας κλινικής δοκιμής, όπως αυτών που αναφέρθηκαν και νωρίτερα.

Βασικές μεταβλητές και μεταβλητές απόκρισης (baseline and response variable)

Μία κλινική μελέτη είναι πιθανό να συμπεριλάβει το ιστορικό κάθε ασθενούς κι έναν μεγάλο αριθμό άλλων μέτρων για αυτόν. Αυτό εμπίπτει σε δύο κατηγορίες.

- 1. βασικές μεταβλητές:** καταγράφουν δημογραφικά και ιατρικά χαρακτηριστικά προτού ξεκινήσει η δοκιμαστική θεραπεία. Είναι χρήσιμες στο να παρέχουν στοιχεία που τοποθετούν το σύνολο των ασθενών σε διαφορετικές θεραπείες που εφαρμόζουν παρόμοιες τεχνικές, έτσι ώστε να συγκριθεί «ο όμοιος με τον όμοιο». Πιο συγκεκριμένα, παρέχεται στους ασθενείς η δυνατότητα να τοποθετηθούν σε ομάδα που λαμβάνει θεραπεία η οποία έχει επιλεγεί με βάση τα χαρακτηριστικά τους μέσω ξεχωριστών αναλύσεων που έχουν γίνει στις διάφορες υποομάδες ή μέσω της χρήση της τεχνικής της συνδιακύμανσης. Τέτοιες αναλύσεις μειώνουν την επίδραση της δειγματικής διακύμανσης και επιτρέπουν την μελέτη τυχόν αλληλεπιδράσεων μεταξύ βασικών μεταβλητών και μεταβλητών απόκρισης.
- 2. μεταβλητές απόκρισης:** μετρούν τις αλλαγές στα χαρακτηριστικά της υγείας κατά την διάρκεια της χορήγησης της θεραπείας και μετά από αυτήν. Ίσως να συμπεριλαμβάνουν ψυχολογικά και βιολογικά tests καταμέτρησης κλινικών ενδείξεων που διαπιστώθηκαν από τον γιατρό, συμπτωμάτων που αναφέρθηκαν από τον ασθενή και , όπου είναι απαραίτητο, χρόνους επιβίωσης. Στις ελεγχόμενες μελέτες είναι ίσως επιθυμητό να εκτιμάται η ποιότητα της ζωής από την μέτρηση της λειτουργικής ικανότητας και της ευημερίας του ασθενούς.

Το πλήθος των μεταβλητών που καταγράφονται πιθανόν να είναι πολύ μεγάλο ιδιαίτερα όταν οι ασθενείς επιστρέφουν για συνηθισμένες ελεγχόμενες επισκέψεις. Ο πειρασμός του να καταγράψουμε όσες περισσότερες μεταβλητές μπορούμε, στην περίπτωση που κάτι τέτοιο φαίνεται να είναι χρήσιμο, πρέπει να περιοριστεί. Μία περίσσεια πληροφοριών ίσως να σπαταλήσει πόρους που θα ήταν καλύτερο να χρησιμοποιηθούν για την επάνδρωση της μελέτης με περισσότερους ασθενείς και είναι πιθανόν να αποσπάσουν την προσοχή από την ακρίβεια του ιστορικού και να μειώσουν τον ενθουσιασμό και το επίπεδο επίδοσης των ερευνητών. Ένας ιδιαίτερος κίνδυνος που αποδίδεται στην καταγραφή των τόσων πολλών μεταβλητών απόκρισης είναι να εμφανιστούν πολύ εύκολα από τύχη εικονικές σημαντικές θεραπευτικές επιδράσεις.

Αν η μεταβλητή απόκρισης που εμφανίζει την πιο σημαντική διαφορά απομονωθεί για να μελετηθεί διεξοδικά, πρέπει ληφθεί υπόψη και το πλήθος όλων των πιθανών μεταβλητών. Παρόλα αυτά , η κατάλληλη διευθέτηση βασίζεται κατά έναν πολύπλοκο τρόπο στην συσχέτιση μεταξύ των μεταβλητών. Στην απίθανη περίπτωση που οι μεταβλητές είναι ανεξάρτητες ,κατάλληλη φαίνεται να είναι η διόρθωση κατά Bonferroni, σύμφωνα με την οποία η χαμηλότερη τιμή P , η οποία συμβολίζεται P_{min} , αναπροσαρμόζεται σε μία νέα μεταβλητή που δίνεται από τον τύπο

$$P' = 1 - (1 - P_{min})^K$$

όπου κ είναι το πλήθος των μεταβλητών που μας απασχολούν. Αν οι μεταβλητές συσχετίζονται, κάτι που είναι υπερβολικά συντηρητικό, δημιουργείται ανάγκη για μία πάρα πολύ μεγάλη διόρθωση. Ο κίνδυνος της εκβάθυνσης γύρω από τα δεδομένα συνήθως μειώνεται με την βοήθεια μιας μεταβλητής απόκρισης, ή ενός πολύ μικρού μεγέθους μεταβλητών, ή μιας πρωτεύουσας παραμέτρου (*primary endpoint*), αντικατοπτρίζοντας έτσι το βασικό σκοπό της δοκιμής. Οι διαφορές στις μεταβλητές αυτές μεταξύ θεραπειών λαμβάνονται ως δεδομένες. Άλλες μεταβλητές δηλώνονται σαν δευτερεύουσες παράμετροι (*secondary endpoints*). Οι διαφορές στην περίπτωση αυτή θεωρούνται ως σημαντικές αλλά δεν καθορίζονται το ίδιο ξεκάθαρα, ίσως γιατί υπόκεινται σε μία πολλαπλή διόρθωση ή γιατί προωθούν υποψήφιους ασθενείς για περαιτέρω δοκιμές.

Παραπάνω έγινε αναφορά στην εκτίμηση της ποιότητας της ζωής στις υπό παρακολούθηση μελέτες. Η επιλογή των μετρήσιμων μεταβλητών θα εξαρτάται από το είδος της δυσλειτουργίας που αναμένεται σαν μία φυσική συνέπεια της υπό μελέτη ασθένειας ή σαν μία πιθανή αρνητική επίδραση της θεραπείας. Ίσως είναι πιθανό να προσδιοριστεί το εύρος των συνθηκών, όλες εκ των οποίων πρέπει να τεθούν υπό παρακολούθηση, και με κάποιον τρόπο να προσδιοριστεί και ο τύπος ενός απλού δείκτη από τον συνδυασμό των μετρήσεων. Ο σκοπός της μέτρησης της ποιότητας της ζωής είναι να αποδεχτούμε ένα ισοζύγιο που παρουσιάζεται μεταξύ του πιθανού οφέλους από την θεραπεία και των πιθανών μειονεκτημάτων που προκαλούνται από την πλευρά των επιδράσεων. Σε μια δοκιμή που αφορά χρόνιες παθήσεις, όπως ο καρκίνος, η διάρκεια ζωής αποτελεί την πρωταρχική παράμετρο, αλλά μία μετριοπαθής παράταση μπορεί να μην είναι επιθυμητή αν συνοδεύεται από αύξηση της δυσλειτουργίας ή του πόνου. Σε κάποιες μελέτες είναι πιθανόν να χρησιμοποιηθούν προσαρμοσμένοι-ποιοτικά χρόνια επιβίωσης, σύμφωνα με τους οποίους κάθε χρόνο οι επιβιώσαντες σταθμίζονται από ένα μέτρο της ποιότητας της ζωής πριν τις καθιερωμένες μεθόδους της ανάλυσης επιβίωσης που εφαρμόστηκαν. Αυτές οι μέθοδοι υιοθετήθηκαν από την ανάλυση επιβίωσης για τις καρκινικές δοκιμές, όπου το σχετικό μέτρο ίσως είναι ο χρόνος χωρίς συμπτώματα και τοξικότητα (*TWiST*), που οδηγεί στην έκδοση προσαρμοσμένης-ποιότητας.

Μέγεθος δοκιμής

Η ακρίβεια των συγκρίσεων μεταξύ θεραπειών είναι φανερά επηρεασμένη από τον αριθμό των ασθενών που την λαμβάνουν. Η καθιερωμένη προσέγγιση είναι να επιλεγεί μία τιμή δ_1 για μία παράμετρο δ , αντιπροσωπεύοντας έτσι μια επαρκώς σημαντική διαφορά στην αποτελεσματικότητα (για κάποιον που δεν θέλει να την χάσει). Οι τιμές σε ένα τεστ επιλέγονται από το επίπεδο σημαντικότητας α όπου η ψευδής υπόθεση ισχυρίζεται ότι $\delta=0$, και για την δύναμη $1-\beta$ έναντι της εναλλακτικής τιμής ότι $\delta=\delta_1$. Με κάποιες εικασίες σχετικά με την κατανομή της εκτίμησης του δ , συγκεκριμένα της διακύμανσης, το δειγματικό μέγεθος δίνεται από τον εξής τύπο,

$$n > 2 \left[\frac{(z_{2\alpha} + z_{2\beta})\sigma}{\delta_1} \right]^2$$

Ο καθορισμός του δειγματικού μεγέθους είναι μία ανακριβής επιστήμη, γιατί η επιλογή των δ , α και β είναι σε ένα βαθμό αυθαίρετη, και η μεταβλητότητα του στατιστικού τεστ ίσως είναι δύσκολο να εκτιμηθεί εκ των προτέρων. Ωστόσο, η χρήση του δίνει στους ερευνητές κάποια εικόνα της ευαισθησίας που πιθανόν να παρουσιαστεί από μία δοκιμή οποιουδήποτε συγκεκριμένου μεγέθους.

3.4 Η ανάθεση της θεραπείας

Τυχαιοποίηση

Πολλοί ερευνητές του 18^{ου} και 19^{ου} αιώνα συνειδητοποίησαν ότι οι θεραπείες χρειάζεται να συγκριθούν σε ομάδες ασθενών με παρόμοια προγνωστικά χαρακτηριστικά, αλλά η έλλειψη των τεχνικών μέσων που υπήρχε δεν έδινε την δυνατότητα για να επιτευχθεί κάτι τέτοιο. Μία εξαιρετική έρευνα δόθηκε από τον Bull (1959).

Κατά την διάρκεια του τέλους του 19^{ου} αιώνα και της αρχής του 20^{ου} διάφοροι ερευνητές δοκιμών όρισαν δυο θεραπείες που εναλλάσσονται σε σειρά. Αυτό το σύστημα είχε το μειονέκτημα ότι η ανάθεση θεραπείας για κάθε ασθενή είναι γνωστή εκ των προτέρων, κάτι που ίσως οδηγεί σε μεροληψία, είτε κατά την εκτίμηση της απόκρισης είτε κατά την επιλογή των ασθενών. Το τελευταίο ενδεχόμενο, της μεροληψίας δηλαδή στην επιλογή, ίσως εμφανιστεί εξαιτίας του γεγονότος ότι η γνώση ή η υποψία της θεραπείας που θα χρησιμοποιηθεί για τον επόμενο ασθενή μπορεί να επιδράσει στην απόφαση του ερευνητή για το εάν ή όχι θα εισάγει τον ασθενή αυτό στην δοκιμή.

Η τυχαιοποίηση αρχικά διεξαγόταν από πίνακες ή τυχαίους αριθμούς, αλλά στις μέρες μας πραγματοποιείται κυρίως με την χρήση υπολογιστών, ενώ η ατομική ανάθεση θεραπείας καθορίζεται μετά την εισχώρηση του ασθενούς στην ομάδα και δίνεται σε καθέναν ένας ατομικός αναγνωριστικός αριθμός (κωδικός).

Ισορροπία

Μια τυχαιοποιημένη ανάθεση θα εξασφαλίσει ότι δεν υπάρχει συστηματική αιτία μεταξύ των υπό θεραπεία ομάδων για οποιαδήποτε δοσμένη βασική μεταβλητή. Αλλά εντελώς από τύχη, μπορεί να υπάρχει κάποιος βαθμός ανισορροπίας, κι ένας ακραίος βαθμός ανισορροπίας στην πρόγνωση μιας βασικής μεταβλητής για μία μεταβλητή απόκρισης ίσως οδηγήσει σε μία παραπλανητικά σημαντική εκτίμηση μιας θεραπευτικής επίδρασης, ακόμα κι αν στην πραγματικότητα δεν υπάρχει καμία. Παρόλα αυτά είναι επιθυμητό να υπάρχει κάποια προστασία έναντι των γνωστών προγνωστικών μεταβλητών.

Αυτό μπορεί να γίνει είτε μέσω ανάλυσης, αναλύοντας δηλαδή χωριστά τα δεδομένα σε κάθε υποομάδα (καθορισμένη από προγνωστικές μεταβλητές), είτε από ένα μοντέλο όπως η ανάλυση της διακύμανσης, ή ακόμα και κατά τον σχεδιασμό της μελέτης περιορίζοντας την τυχαιοποίηση ώστε να προωθηθεί η ισορροπία που απαιτείται.

Σε κάποιες πρόωρες δοκιμές, η ισορροπία μπορεί να επιτευχθεί από την μέθοδο της περιορισμένης τυχαιοποίησης (*restricted randomization*) ή της αντιμετάθεσης blocks(*permuted blocks*). Εδώ τα επίπεδα καθορίζονται από προγνωστικές μεταβλητές ή καθορισμένους σχεδιασμούς αυτών όπως για παράδειγμα : γυναίκα ,ηλικίας 30-39, με διάρκεια συμπτωμάτων λιγότερο του ενός έτους. Για κάθε στρώμα κατασκευάζεται χωριστά ένας τυχαιοποιημένος σχηματισμός, έτσι ώστε οι αριθμοί που κατανέμονται σε διαφορετικές θεραπείες να ισοσταθμίζονται με συνηθισμένα διαστήματα πχ. ανά δεκάδες. Όταν ολοκληρωθεί η είσοδος των ασθενών ο αριθμός τους δεν θα είναι ίδιος για διαφορετικές θεραπείες , αλλά πιθανόν να είναι αρκετά κοντά σε σχέση με μία απλή τυχαιοποίηση.

Ένα μειονέκτημα των αντιμεταθέσεων των blocks είναι ότι η ανάθεση θεραπείας για έναν νέο ασθενή είναι προκαθορισμένη αν αυτός ο ασθενής τύχει να εισέλθει στο τέλος ενός block. Ένα άλλο πρόβλημα είναι ότι , αν χρησιμοποιούνται τόσες πολλές μεταβλητές , πολλά επίπεδα θα είναι πολύ μικρά ώστε να επιτρέψουν επαρκή ισορροπία των θεραπειών.

Τώρα είναι περισσότερο συνηθισμένοι κάποιοι τύποι ελαχιστοποίησης (minimization-Taves,1974, Begg & Iglewitz ,1980). Ο σκοπός εδώ είναι να προσδιοριστούν οι ασθενείς με τέτοιο τρόπο ώστε να ελαχιστοποιηθεί η πρόσφατη ανομοιότητα μεταξύ των ομάδων, λαμβάνοντας υπόψη ταυτόχρονα μια ποικιλία προγνωστικών μεταβλητών. Η ελαχιστοποίηση μπορεί να διεξαχθεί από το προσωπικό προγραμματιστών στο συντονιστικό κέντρο δοκιμών , έτσι ώστε μία ανάθεση να μπορεί να καθοριστεί σχεδόν απευθείας όταν το κέντρο κοινοποιεί τις τιμές των βασικών μεταβλητών για κάθε ασθενή. Μία πολυκεντρική μελέτη είναι σύνηθες να κανονίζει ότι το κλινικό κέντρο είναι ένας από τους παράγοντες που εξισορροπείται, έτσι ώστε κάθε κέντρο να μπορεί να ισχυριστεί ότι έχει ικανοποιητική ισορροπία στην περίπτωση που χρειαστεί μία ξεχωριστή ανάλυση για κάποιον ασθενή στο συγκεκριμένο κέντρο.

Κατανομή εξαρτώμενη από τα δεδομένα

Στις περισσότερες κλινικές δοκιμές , οι ασθενείς καθορίζονται έτσι ώστε να είναι σε ίσες αναλογίες για διαφορετικές ομάδες θεραπειών, ή σε απλούς λόγους όπως 2:1, οι οποίοι διατηρούνται σταθεροί σε όλη την δοκιμή. Για παράδειγμα, σε μία δοκιμή με μία ελεγχόμενη θεραπεία και διάφορες νέες θεραπείες είναι περισσότερο χρήσιμο να καθοριστεί μεγαλύτερη αναλογία για την ελεγχόμενη ομάδα.

Εδώ θέτονται και κάποια ηθικά ζητήματα όπως το αν οι ερευνητές θα θεωρήσουν σωστό να δώσουν οποιαδήποτε από τις αντίπαλες θεραπείες σε οποιονδήποτε ασθενή. Οπωσδήποτε αν τα στοιχεία που συλλέγονται κατά την πορεία της δοκιμής υποδείξουν ότι μία θεραπεία είναι κατώτερη από μία ή περισσότερες άλλες, το ηθικό αντίβαρο των ερευνητών ίσως αποδυναμωθεί. Παρόλα αυτά, κάποιοι ερευνητές έχουν συμφωνήσει ότι, αντί της εφαρμογής ίσων προκαθορισμένων αναλογιών σε όλες τις θεραπείες και μέχρι να συμβεί ένας ξαφνικός τερματισμός, θα ήταν ηθικά προτιμότερο να επιτραπούν αναλογικές αναθέσεις καθ' όλη την διάρκεια της δοκιμής με τέτοιο τρόπο που οι περισσότεροι ασθενείς να τοποθετούνται βαθμιαία σε περισσότερο επιτυχημένες θεραπείες.

Η συνέπεια οποιουδήποτε τέτοιου σχεδίου είναι ότι, αν υπάρχουν έντονες διαφορές μεταξύ των θεραπειών, το πλήθος των ασθενών που τελικά θα τοποθετηθούν σε καλύτερη θεραπεία θα υπερβαίνει τον αριθμό των ασθενών της υποδεέστερης θεραπείας, ακόμη και στην περίπτωση που το στοιχείο που υπέδειξε την ύπαρξη της επίδρασης δεν είναι ισχυρό. Επιπλέον, ίσως εμφανιστούν δυσκολίες κατά την διάρκεια της θεραπείας, αν υπάρχει κάποια χρονική τάση στα προγνωστικά χαρακτηριστικά του ασθενούς. Οι ασθενείς που εισχωρούν προς το τέλος της δοκιμής, και συνεπώς τοποθετούνται στην επιθυμητή θεραπεία, ίσως έχουν καλύτερη (ή χειρότερη) πρόγνωση από αυτούς που εισήλθαν νωρίτερα, και οι οποίοι τοποθετήθηκαν με ίσες αναλογίες σε διαφορετικές θεραπείες. Μία επιπλέον πρακτική δυσκολία είναι ότι οι γιατροί συνήθως είναι απρόθυμοι να τοποθετήσουν έναν στους δέκα, για παράδειγμα, ασθενείς σε μία χειρότερη θεραπεία, και ίσως προτιμούν ένα σύστημα από το οποίο το ηθικό τους αντίβαρο να διατηρηθεί μέχρι να υπάρξει ισχυρή ένδειξη που αφορά μία επίδραση.

Η κατανομή που εξαρτάται από δεδομένα τέτοιου τύπου καλείται *outcome-dependent allocation* ώστε να την διαχωρίζουμε από σχεδιασμούς όπως η ελαχιστοποίηση που συζητήθηκε στην προηγούμενη ενότητα. Στην τελευταία, η κατανομή εξαρτάται από τιμές βασικών μεταβλητών για τον ασθενή αλλά όχι από τα αποτελέσματα όπως μετριούνται από μία μεταβλητή απόκρισης. Επίσης χρησιμοποιείται ο όρος 'προσαρμοστική ανάθεση' (*adaptive assignment*). Σπάνια βέβαια χρησιμοποιούνται προσαρμοστικοί σχεδιασμοί στην πράξη.

Τυχαιοποιημένη συναίνεση

Το σύνηθες είναι αφού ενημερωθεί ο ασθενής και πριν εισέλθει στην κλινική δοκιμή να επιδιώκεται η συγκατάθεση του. Αν κάποιοι ασθενείς δεν δίνουν την συναίνεση τους ή αν οι γιατροί βρίσκουν δύσκολο να ζητήσουν από τον ασθενή να συμφωνήσει στην θεραπεία που θα καθοριστεί από τυχαία διαδικασία, μάλλον καθίσταται αδύνατο να επανδρωθεί η δοκιμή με έναν επαρκή αριθμό. Ο Zelen (1979) πρότεινε μία διαδικασία η οποία είναι αρκετά χρήσιμη σε αρκετές δοκιμές όπου δηλαδή μία νέα θεραπεία N, πρόκειται να συγκριθεί με μία καθιερωμένη θεραπεία S.

Στον σχεδιασμό του Zelen οι υποψήφιοι ασθενείς είναι τοποθετημένοι τυχαία σε δύο ομάδες, την N και S. Στους ασθενείς της S χορηγείται η θεραπεία S χωρίς να ζητηθεί η συγκατάθεση τους, Στους ασθενείς της N ζητείται η συναίνεση για να λάβουν την θεραπεία N. Αν συμφωνήσουν λαμβάνουν την θεραπεία N, αν όχι λαμβάνουν την θεραπεία S. Η άρνηση της συναίνεσης από την ομάδα ακόμα κι αν οι ασθενείς αυτοί λάβουν μία καθιερωμένη θεραπεία, προκαλεί κάποιες φορές διαμάχες γιατί ο σχεδιασμός δεν μπορεί να χρησιμοποιηθεί χωρίς έγκριση για λόγους ηθικούς.

Στη συνέχεια γίνεται μία ικανοποιητική σύγκριση από την προσέγγιση *τάση-για-θεραπεία* (*intention-to-treat*), βασισμένοι στο σύνολο των ομάδων N και S, ακόμα κι αν δεν λαμβάνουν όλοι οι ασθενείς της N την θεραπεία N. Αν υπάρχει διαφορά στην επίδραση των δύο θεραπειών υπάρχει η τάση να υποτιμάται, λόγω της διαφοράς στους μέσους απόκρισης των ομάδων N και S αφού υπάρχει παραπλανητική επίδραση των ασθενών που δεν έλαβαν την θεραπεία N. Υπάρχει πιθανότητα να εκτιμηθεί η

πραγματική διαφορά ανάμεσα στις ομάδες των ατόμων που συναίνεσαν, αλλά αυτό θα ήταν μάλλον ανεπαρκές αν η αναλογία των ασθενών αυτών είναι χαμηλή.

3.5 Η αξιολόγηση της απόκρισης

Η τυχαιοποίηση, αν εκτελεστεί σωστά, εξασφαλίζει την απουσία της μεροληψίας χάρη στην τοποθέτηση των ασθενών με διαφορετικά προγνωστικά χαρακτηριστικά σε διαφορετικές θεραπευτικές ομάδες. Είναι σημαντικό να αποφευχθεί η μεροληψία η οποία ίσως προκύπτει όταν εφαρμοστούν διαφορετικές προϋποθέσεις των καταχωρημένων αποκρίσεων για διαφορετικές θεραπείες.

Στις περισσότερες δοκιμές συγκρίνεται μία ενεργή θεραπεία με μία άλλη. Επιπλέον μερικές φορές είναι επιθυμητό να συγκριθεί μία νέα θεραπεία με την απουσία της ίδιας της θεραπείας, αν και οι δύο θεραπευτικές ομάδες λαμβάνουν ταυτόχρονα, μία καθιερωμένη θεραπεία. Σε μία τέτοια δοκιμή προκύπτει η απλή γνώση ότι μία επιπρόσθετη παρέμβαση που έγινε σε μία ομάδα παρέχει ίσως μόνο ένα φαινομενικό όφελος, ανεξάρτητο του ουσιαστικού πλεονεκτήματος αυτής της παρέμβασης. Για τον λόγο αυτό, στους ασθενείς της υπό μελέτης ομάδας ίσως δοθεί placebo, ένας αδρανής δηλαδή τύπος θεραπείας πανομοιότυπης με την υπό έλεγχο θεραπεία. Σε μία δοκιμή φαρμάκων ,για παράδειγμα, το placebo θα είναι μία αδρανής ουσία κατασκευασμένη με τέτοιο τρόπο ώστε να έχει το ίδιο σχήμα με αυτό του νέου φαρμάκου και θα χορηγηθεί σύμφωνα με την ίδια φαρμακευτική αγωγή. Με τον τρόπο αυτό, μπορεί να διαχωριστεί η φαρμακευτική επίδραση του νέου φαρμάκου από την ψυχολογική επίδραση της γνώσης της χρήσης του.

Οι βασικές αρχές που αφορούν την κάλυψη της ταυτότητα της θεραπείας επεκτείνονται σε δοκιμές στις οποίες συγκρίνονται δύο ή περισσότερες πιθανές ενεργές θεραπείες. Ο κύριος σκοπός εδώ είναι να εξασφαλιστεί ότι το μέτρο της μεταβλητής απόκρισης δεν επηρεάζεται από την γνώση της χορήγησης της ειδικής θεραπείας. Αν η σχετική απόκριση είναι επιβίωση ή θάνατος , είναι σχεδόν βέβαιο ότι θα καταγραφεί αντικειμενικά και με ακρίβεια ,και δεν πρόκειται να συμβεί καμία μεροληψία. Οποιοδήποτε άλλο μέτρο κατά την πορεία της θεραπείας , όπως η αναφορά των συμπτωμάτων από τον ασθενή, η αναφορά ενδείξεων από τον γιατρό, το ιστορικό της επιδείνωσης της ασθένειας ή ακόμη και το ιστορικό των βιολογικών ελέγχων μέτρησης ,ίσως επηρεάζονται από την γνώση της θεραπείας που λαμβάνεται. Αυτό συμπεριλαμβάνει την γνώση από τον ασθενή, από τον γιατρό ή από οποιοδήποτε άλλο υπάλληλο του τεχνικού προσωπικού. Είναι συνεπώς σημαντικό να κανονίζονται να χορηγούνται για τις θεραπείες κάποιοι τύποι ‘κάλυψης’. Ένας απλός τύπος κάλυψης είναι όταν η ταυτότητα της θεραπείας δεν φανερώνεται στον ασθενή (*single-masked*). Στους περισσότερο συνηθισμένους πιο σύνθετους καλυμμένους σχεδιασμούς, η ταυτότητα της θεραπείας κρύβεται τόσο από τον γιατρό όσο και από οποιονδήποτε άλλο υπάλληλο προσωπικού συμμετέχει στην αξιολόγηση της απόκρισης(*double-masked*). Σε κάποιες περιπτώσεις βέβαια είναι απαραίτητο ο γιατρός να είναι ενήμερος για την ταυτότητα της θεραπείας, ιδιαίτερα σε πολύπλοκες παρεμβάσεις, αλλά πιθανόν για την σωστή αξιολόγηση της απόκρισης κάτι τέτοιο να μην γίνει.

Η κάλυψη επιτυγχάνεται από την εξασφάλιση ότι οι σχετικές θεραπείες σχεδιάζονται με τον ίδιο ακριβώς τρόπο. Αν δύο φάρμακα πρέπει να χορηγηθούν με διαφορετικό τρόπο, για παράδειγμα με ταμπλέτα ή κάψουλα, πιθανόν να χρησιμοποιηθεί μία διπλή-ψεύτικη τεχνική (*double-dummy technique*). Για να συγκρίνουμε ένα φάρμακο Α με ταμπλέτα κι ένα φάρμακο Β με κάψουλα, οι δύο ομάδες θα λάβουν

Ενεργό Α ταμπλέτες, συν placebo Β κάψουλες

ή

Placebo Α ταμπλέτες, συν ενεργό Β κάψουλες

Όταν αποφασιστεί η χορήγηση κάποιας θεραπείας για έναν ασθενή, η ταμπλέτα ή η κάψουλα κτλ. πρέπει να δοθεί ένας κωδικός συγκεκριμένα για αυτόν τον ασθενή. Ένα εναλλακτικό σύστημα που μερικές φορές χρησιμοποιείται, σύμφωνα με το οποίο μία συγκεκριμένη θεραπεία δίνει ένα συγκεκριμένο γράμμα ως κωδικό, πχ Α, και όλοι οι συνδυασμοί που περιέχουν αυτό το φάρμακο καταχωρούνται επίσης σαν Α. Αυτό το σύστημα έχει το μειονέκτημα ότι η ταυτότητα των Α είναι γνωστή ή μπορούν να την υποπτευθούν, όπως για παράδειγμα μέσω της αναγνώρισης των συμπτωμάτων. Τότε ο κωδικός επαγωγικά θα αποκαλυφθεί για όλους τους ασθενείς που λαμβάνουν την θεραπεία αυτή.

Η χρήση λοιπόν του placebo ίσως είναι δύσκολη στη πράξη, είτε γιατί μία θεραπεία έχει εύκολα εντοπίσιμες αρνητικές επιδράσεις είτε γιατί η φύση της παρέμβασης είναι τέτοια που δεν μπορεί να εξομοιωθεί με μία παρόμοια πλασματική. Η τελευταία περίπτωση θα ήταν, για παράδειγμα, μία εγχείρηση. Οι μηχανισμοί κάλυψης, συνεπώς, αν και έχουν υψηλή αποτελεσματικότητα στις περισσότερες περιπτώσεις, δεν θα πρέπει να θεωρούνται σαν πανάκεια για τη αποφυγή μεροληπτικής απόκρισης. Θα ήταν επίσης σοφό να ελέγχουμε την αποτελεσματικότητα θέτοντας ερωτήσεις στους υποτιθέμενους ασθενείς και γιατρούς που δεν γνωρίζουν την θεραπεία που χορηγήθηκε και να συγκρίνουμε τις εικασίες που προκύπτουν με την πραγματική κατάσταση.

Παράδειγμα 1:

Ο πίνακας (1) δείχνει τα αποτελέσματα από μία δοκιμή της βιταμίνης Α, ένα συμπλήρωμα που έχει σκοπό να μειώσει την θνησιμότητα κατά τη περίοδο των πρώτων 8 μηνών σε παιδιά προσχολικής ηλικίας στην επαρχιακή Ινδονησία, όπως αναφέρθηκαν από τον Tarwotjo (1987) και συζητήθηκαν από τους Sommer και Zeger (1991).

πίνακας (1) αποτελέσματα από μία δοκιμή της βιταμίνης Α σε παιδιά προσχολικής ηλικίας στην Ινδονησία (με άδεια από τις τοπικές αρχές)

	Control group			Treatment group		
	Compliance			Compliance		
	No(1)	Yes(2)	Total(3)	No(4)	Yes(5)	Total(6)
Alive	?(m ₀₀)	?(m ₀₁)	11514(m _{0.})	2385(n ₀₀)	9663(n ₀₁)	12048(n _{0.})
Dead	?(m ₁₀)	?(m ₁₁)	74(m _{1.})	34(n ₁₀)	12(n ₁₁)	46(n _{1.})
Total		?(m _{.1})	11588(M)		9675(n _{.1})	12094(N)

Ερμηνεία πίνακα:

- οι στήλες (1) και (2) αναφέρονται σε άτομα που θα ήταν ευπαθή ή μη-ευπαθή αν ήταν στην υπό θεραπεία ομάδα και φυσικά ο αριθμός τους δεν είναι γνωστός. Παρόλα αυτά οι συγκεκριμένοι αριθμοί μπορούν να εκτιμηθούν. Με την εικασία ότι η αναλογία των θανάτων θα ήταν η ίδια για όλα τα μη-συμβατά άτομα των στηλών (1) και (4), αφού καμία υποομάδα δεν έχει λάβει την θεραπεία, και αφού οι αναμενόμενες αναλογίες των μη-συμβατών ατόμων θα έπρεπε να είναι ίδια στην υπό θεραπεία και στην υπό έλεγχο ομάδα, οι εισχωρήσεις στην στήλη (1) μπορούν να υπολογιστούν ως $m_{00} = (M / N)n_{00}$ και $m_{10} = (M / N)n_{10}$ και εκείνες της στήλης (2) από $m_{01} = m_{0.} - m_{00}$ και το $m_{11} = m_{1.} - m_{10}$. Οι προσαρμοσμένες αναλογίες θανάτων ανάμεσα στα υπό θεραπεία συμβατά άτομα στην υπό έλεγχο ομάδα εκτιμάται τώρα από τον τύπο $p_C = m_{11} / m_{1.}$, και συγκρίνεται με την παρατηρούμενη αναλογία στην υπό θεραπεία ομάδα, $p_{TC} = n_{11} / n_{1.}$
- στο συγκεκριμένο παράδειγμα έχουμε, $p_C = 41.4 / 9270.2 = 0.447\%$, και $p_{TC} = 12 / 9675 = 0.124\%$, μία μείωση της τάξης του 72%. Στην ΙΤΤ ανάλυση οι συνολικές αναλογίες θανάτων είναι $p_C = 74 / 11588 = 0.693\%$ και $p_T = 46 / 12094 = 0.380\%$, μία μείωση της τάξης του 41%.

Παρατηρήσεις

Πρέπει να καταγράψουμε κάποια σημαντικά σημεία της ανάλυσης που περιγράφηκε στο παραπάνω παράδειγμα.

1. Η εικασία ότι τα μη συμβατά άτομα θα ανταποκριθούν στην θεραπεία με τον ίδιο ακριβώς τρόπο αν τοποθετηθούν στην υπό έλεγχο ομάδα είναι πολύ κρίσιμη. Ίσως είναι βάσιμο για το παράδειγμα (1) ,όπου η απόκριση είναι επιβίωση ή θάνατος, αλλά θα ήταν λιγότερο αυταπόδεικτο γεγονός σε μία δοκιμή με περισσότερο υποκειμενικές αποκρίσεις. Η εμπειρία της ανάθεσης ή της άρνησης μιας ενεργής θεραπείας ίσως επηρεάσει την απόκριση, προκαλώντας μια συμπεριφορά τυπική ή μία συναισθηματική αντίδραση που δεν θα υπήρχε αν ο ασθενής δεν τοποθετούταν στην υπό έλεγχο ομάδα.
2. Το μοντέλο που χρησιμοποιήθηκε στο παράδειγμα (1) θα μπορούσε να προσαρμοστεί σε μία τυχαιοποιημένη δοκιμή με συναίνεση, όπου η συγκατάθεση για την θεραπεία στην τελευταία περίπτωση θα ανταποκρινόταν στην συμβατότητα της προηγούμενης. Στην πράξη μία ΙΤΤ ανάλυση συνήθως προτιμάται για τυχαιοποιημένες δοκιμές με συναίνεση ,ώστε να αποφεύγονται εικασίες ανάλογες με αυτές που περιγράφηκαν στην 1, εννοώντας, ότι όποιος δεν συγκατατίθεται θα μπορούσε να ανταποκριθεί με τον ίδιο τρόπο σε οποιαδήποτε ανάθεση θεραπείας.
3. Τα αποτελέσματα των δύο ομάδων στο παραπάνω παράδειγμα έχουν συγκριθεί με τον λόγο αναλογιών του θανάτου. Αυτό είναι λογικό, αφού ο απόλυτος δείκτης θνησιμότητας θα εξαρτάται από μία αυθαίρετη διάρκεια παρακολούθησης. Σε άλλες δοκιμές ίσως είναι προτιμότερο να παραθέσουμε την διαφορά στις αναλογίες των θετικών αποκρίσεων.
4. Το δειγματικό σφάλμα της εκτιμημένης αναλογίας απόκρισης p_c διαφέρει από την συνηθισμένη διωνυμική διακύμανση που είναι κατάλληλη για μια απευθείας παρατήρηση από ένα τυχαίο δείγμα μεγέθους m_1 .
5. Σε πολλά παραδείγματα της χρήσης αυτού του μοντέλου ,η υπό έλεγχο ομάδα ίσως είναι επίσης υποδιαιρεμένη , με γνωστούς αριθμούς των συμβατών και μη-συμβατών ατόμων στην ελεγχόμενη θεραπευτική αγωγή. Αυτοί οι αριθμοί δεν χρειάζονται για την ανάλυση, αφού τα χαρακτηριστικά των ασθενών που βρίσκονται υπό θεραπεία πιθανόν να διαφέρουν από εκείνων που βρίσκονται στην ομάδα με την υπό έλεγχο θεραπευτική αγωγή. Για τον λόγο αυτό θα ήταν λάθος να συγκρίνουμε επιτυχημένες αναλογίες στις δύο υποομάδες (εκείνες που είναι συμβατές με την θεραπεία και συμβατές με την ελεγχόμενη θεραπεία αντίστοιχα).

Το μοντέλο που χρησιμοποιήθηκε στο παράδειγμα (1) είναι κορεσμένο υπό την έννοια ότι έχει ίδιο αριθμό παραμέτρων, συγκεκριμένα τέσσερις, όπως και για τις τυχαίες παρατηρήσεις. Οι παράμετροι μπορεί να επιλεγούν να είναι: θ η πιθανότητα ο ασθενής να είναι συμβατός με μία θεραπεία. Οι πιθανότητες θανάτου για τους συμβατούς και μη-συμβατούς με την θεραπεία στην υπό έλεγχο ομάδα, π_1 και π_2 αντίστοιχα. Και δ , η μείωση της πιθανότητας του θανάτου που οφείλεται στην θεραπεία, ανάμεσα στα συμβατά με την θεραπεία άτομα. Οι ανεξάρτητες παρατηρήσεις συμβολίζονται με t , από

τις οποίες προκύπτει και η παρατηρούμενη αναλογία των συμβατών ατόμων στην υπό θεραπεία ομάδα. Και p_T , p_{TC} και p_{TNC} , οι αναλογίες θανάτου στην υπό έλεγχο ομάδα και ανάμεσα σε συμβατούς και μη συμβατούς με την υπό θεραπεία ομάδα, αντίστοιχα.

Σ' ένα τέτοιο μοντέλο, η εκτιμήτρια μέγιστης πιθανοφάνειας των παραμέτρων είναι τέτοια ώστε οι αναμενόμενες συχνότητες να είναι ολόιδιες με αυτές που παρατηρήθηκαν, και οι παράμετροι ίσως είναι πιο εύκολο να εκτιμηθούν εξισώνοντας τις παρατηρούμενες και τις αναμενόμενες τιμές. Οι υπολογισμοί στο παράδειγμα (1) ακολουθούν αυτήν την προσέγγιση.

Συνεπώς, η εκτίμηση της θεραπευτικής επίδρασης δ είναι

$$\delta = \frac{p_C - (1-t)p_{TNC}}{t}$$

Αντικαθιστώντας από τον πίνακα (1) στον παραπάνω τύπο το $t=9675/12094$, το $p_C = 74/11588$, το $p_{TC} = 12/9675$ και το $p_{TNC} = 34/2419$, βρίσκουμε ότι $\delta = 0.00323$ που έρχεται σε συμφωνία με την εκτίμηση που δόθηκε νωρίτερα στο παράδειγμα (1) και σύμφωνα με την οποία το ποσοστό μείωσης είναι 0.447% - 0.124%.

3.6 Ο έλεγχος των δεδομένων

Σε οποιαδήποτε μεγάλη δοκιμή, οι ερευνητές πρέπει να στήσουν ένα σύστημα παρακολούθησης της χορήγησης του φαρμάκου, ώστε να ελέγχουν τα υψηλά επίπεδα που διατηρούνται κατά την διάρκεια της δοκιμής. Ένα τέτοιο σύστημα θα ελέγχει αν ο δείκτης της επιδιωκόμενης επάνδρωσης της δοκιμής με τους κατάλληλους ασθενείς επιτεύχθηκε, διαπιστώνοντας επιπλέον παραβιάσεις στα κριτήρια εισόδου και παρακολούθησης της ακρίβειας των πληροφοριών που καταχωρήθηκαν. Ο έλεγχος των δεδομένων ίσως αποκαλύπτει χαρακτηριστικά του πρωτοκόλλου που δεν ικανοποιήθηκαν και οδηγούν έτσι στην επανέκδοση του. Αν ο δείκτης της επάνδρωσης είναι χαμηλής προσδοκίας τότε οι ερευνητές είτε θα αναζητήσουν συνεργασία κι από άλλα κλινικά κέντρα, είτε θα ελευθερώσουν με κάποιον τρόπο τα κριτήρια εισόδου στην δοκιμή.

Ο έλεγχος των δεδομένων κανονικά δεν πρέπει να χρησιμοποιεί τα εξαγόμενα συμπεράσματα για τους ασθενείς της δοκιμής. Αντίθετα, ο έλεγχος των δεδομένων πρέπει να ασχολείται με τα στοιχεία που προκύπτουν από τα δεδομένα που έχουν συλλεχθεί και αφορούν την ασφάλεια και την αποτελεσματικότητα των θεραπειών που βρίσκονται υπό δοκιμή.

Η ασφάλεια είναι ένα από τα σημαντικότερα θέματα σε οποιαδήποτε σχεδόν δοκιμή. Οι περισσότερες ιατρικές θεραπείες και διαδικασίες παράγουν ασήμαντες επιδράσεις, οι οποίες ούτως ή άλλως αναμένονταν λόγω των αποτελεσμάτων προηγούμενων ερευνών και πιθανόν να μην προκαλούν καμία ιδιαίτερη ανησυχία. Από την άλλη πρέπει να ελέγχονται εξονυχιστικά σοβαρά γεγονότα που εμφανίζονται να είναι αρνητικά (*serious adverse events-SAEs*), ιδιαίτερα όταν σχετίζονται με την απειλή της

ζωής. Ένα ισχυρό στοιχείο ενός μη αναμενόμενου SAE, που δεν εξισορροπείται ξεκάθαρα από πλεονεκτήματα για την επιβίωση του ασθενούς, μπορεί να οδηγήσει στον τερματισμό της δοκιμής ή το λιγότερο σε τροποποίηση ή εγκατάλειψη της υποτιθέμενης θεραπείας.

Φυσικά διαφορές στην αποτελεσματικότητα μπορεί να προκύψουν κατά την διάρκεια της δοκιμής και αναμενόμενο είναι να αυξήσουν την ανησυχία γύρω από ηθικά θέματα. Οι ερευνητές όμως το πιθανότερο είναι ότι ξεκινώντας την δοκιμή έλαβαν υπόψη το ηθικό αντίβαρο, θεωρώντας όλες τις αντιμαχόμενες θεραπείες ως αποδεκτές. Αν τα στοιχεία που προκύπτουν κατά την διάρκεια της δοκιμής υποδείξουν ότι μία θεραπεία είναι υποδεέστερη από μία άλλη, οι ερευνητές ίσως αισθανθούν ότι πρέπει να σταματήσουν μία δοκιμή ή τουλάχιστον να παραλείψουν τη αμφιλεγόμενη θεραπεία.

Οι μηχανισμοί ελέγχου των δεδομένων αναπτύσσονται στις επόμενες παραγράφους.

Η επιτροπή ελέγχου των δεδομένων

Η ευθύνη πρόωρου τερματισμού μιας δοκιμής ή οι αλλαγές στο πρωτόκολλο εναπόκεινται στους ερευνητές (οι οποίοι σε μία πολυκεντρική ή σε άλλες μεγάλες δοκιμές απαρτίζουν την επιτροπή κατευθυντήριων οδηγιών). Παρόλα αυτά, σε μια δοκιμή με κάλυψη οι ερευνητές δεν μπορούν να ελέγξουν τα αποτελέσματα καθώς είναι ανίκανοι για την χορήγηση της θεραπείας. Για τον λόγο αυτό συνήθως τα άτομα τοποθετούνται σε μία ανεξάρτητη ομάδα, γνωστή ως επιτροπή ελέγχου των δεδομένων (*Data Monitoring Committee-DMC*). Το DMC απαρτίζεται από έναν ή περισσότερους στατιστικούς, κάποιους γιατρούς ειδικούς στην περιοχή που μελετάται και ίσως κάποια άλλα μέλη. Δεν είναι φυσιολογικό να συμπεριλαμβάνονται ερευνητές ή διαφημιστές αν και η άποψη πάνω στο θέμα αυτό ποικίλει.

Για την αξιολόγηση των στοιχείων, το DMC καλείται να αντιμετωπίσει όλες τις δυσκολίες που προκύπτουν από τις επαναλαμβανόμενες αναλύσεις των δεδομένων που έχουν συλλεχθεί. Αυτό θα συζητηθεί σε γενικές γραμμές στις επόμενες υποενότητες, όπου αναλύονται πιο διεξοδικά οι μέθοδοι αναλύσεων.

Διαδοχική Ανάλυση

Διαδοχική έρευνα λέγεται αυτή που οι παρατηρήσεις είναι σε σειρά και η διεξαγωγή, το σχέδιο ή η απόφαση του τερματισμού εξαρτάται από τα δεδομένα που έχουν συλλεχθεί μέχρι τώρα. Εδώ ασχολούμαστε κυρίως με την πιθανότητα ενός πρόωρου τερματισμού. Οι διαδοχικές αναλύσεις παρέχουν μεθόδους για ανάλυση δεδομένων στις οποίες η απόφαση αν θα τερματιστεί σε οποιοδήποτε σημείο η δοκιμή βασίζεται στα δεδομένα που έχουν συλλεχθεί. Αρχικά η μέθοδος αυτή αναπτύχθηκε με την πρωτοποριακή εργασία του A. Wald (1902-50).

Κάθε διαδοχική ανάλυση έχει έμμεσα τον ρόλο της διακοπής των αρχικών αποφάσεων, καθορίζοντας τον τρόπο με τον οποίο η απόφαση για τερματισμό εξαρτάται από τα δεδομένα που ήδη έχουμε. Ένα απλό παράδειγμα στο οποίο μπορεί να χρησιμοποιηθεί η διαδοχική εκτίμηση είναι όταν ο ερευνητής θέλει να εκτιμήσει μία

παράμετρο σε ένα συγκεκριμένο επίπεδο ακριβείας. Με την υπόθεση ότι , σε ένα τυχαίο δείγμα μεγέθους n

από μία κατανομή με μέσο μ και διακύμανση σ^2 ,ο εκτιμώμενος μέσος είναι \bar{x}_n και η εκτιμώμενη σταθερή απόκλιση είναι s_n (ο δείκτης υποδεικνύει ότι τα στατιστικά αποτελέσματα αλλάζουν τυχαία αν το μέγεθος n αυξηθεί). Το εκτιμώμενο σταθερό σφάλμα του \bar{x}_n είναι s_n/\sqrt{n} και , αν και το s_n έχει τυχαίο εύρος ,το σταθερό σφάλμα τείνει να μειώνεται καθώς το μέγεθος του δείγματος αυξάνεται.

Ένας πιθανός κανόνας τερματισμού για την διαδοχική εκτίμηση του μ είναι να συνεχίσουμε την δειγματοληψία μέχρι το σταθερό σφάλμα να πέσει σε κάποια προκαθορισμένη χαμηλή τιμή και τότε να σταματήσει.

Οι καθιερωμένοι μέθοδοι ανάλυσης, όπως αυτοί που περιγράφηκαν σε προηγούμενα κεφάλαια , έχουν υποθέσει ότι το δείγμα είναι σταθερό και μεγέθους n . Η ερώτηση τότε που προκύπτει είναι αν αυτές οι μέθοδοι είναι έγκυρες για διαδοχικές μελέτες όπου το n δεν είναι προκαθορισμένο αλλά εξαρτάται από τα δεδομένα που έχουν συλλεχθεί. Η ερώτηση μπορεί να απαντηθεί σε δύο διαφορετικά επίπεδα. Από μία συνήθης σκοπιά, οι ιδιότητες της στατιστικής διαδικασίας επηρεάζονται από την δειγματοληπτική αλληλουχία, στην οποία με τη πάροδο του χρόνου οι ιδιότητες πρέπει να υπολογίζονται από υποθετικές επαναλήψεις των δεδομένων με τον ίδιο κανόνα τερματισμού παρά με το ίδιο δειγματικό μέγεθος. Παρόλα αυτά, για διαδοχικές εκτιμήσεις ,όπως στην απλή περίπτωση που περιγράφηκε παραπάνω, η επίδραση είναι αρκετά μικρή. Για παράδειγμα, η πιθανότητα ένα διάστημα εμπιστοσύνης βασισμένο σε ένα εκτιμημένο σταθερό σφάλμα να καλύπτει την τιμή της παραμέτρου δεν είναι πολύ επηρεασμένη, ιδιαίτερα σε μεγάλα δειγματικά μεγέθη. Ωστόσο θα δούμε σε επόμενη υποενότητα ότι άλλες διαδικασίες ,όπως τα τεστ σημαντικότητας ,ίσως να επηρεάζονται περισσότερο. Από άλλη οπτική ίσως να επιθυμούμε να βγάλουμε συμπεράσματα από την συνάρτηση μέγιστης πιθανοφάνειας ή, με μία Μπεϋζιανή ερμηνεία , από την μεταγενέστερη κατανομή (*posterior distribution*) με την βοήθεια κάποιας προγενέστερης κατανομής (*prior distribution*) . Εδώ ο κανόνας τερματισμού είναι ασυσχέτιστος, αφού οι πιθανοφάνειες για διαφορετικές τιμές παραμέτρων θεωρούν τον ίδιο λόγο οποιονδήποτε κανόνα τερματισμού κι να έχουν. Το σημαντικό αυτό αποτέλεσμα καλείται *ισχυρή αρχή πιθανοφάνειας* (*strong likelihood principle*).

Στον έλεγχο δεδομένων μιας κλινικής δοκιμής ,είναι πιθανόν να εμφανιστεί η περίπτωση του πρόωρου τερματισμού γιατί υπάρχει μία ισχυρή ένδειξη για μία επίδραση χάρη στην οποία ή ,ενάντια στην οποία , μία θεραπεία ,και ο καθιερωμένος τρόπος για να εξεταστεί η ένδειξη αυτή σε ένα μη διαδοχικό πείραμα είναι ο μέσος από ένα τεστ σημαντικότητας . Υποθέτουμε ότι ένα τεστ σημαντικότητας για να συγκρίνει τις επιδράσεις των μέσων των δύο θεραπειών διεξάγεται επαναλαμβανόμενα βάση των δεδομένων που έχουμε, είτε στις σποραδικές συναντήσεις ενός συμβουλίου DMC , ή πιο συνηθισμένα από τους στατιστικούς της δοκιμής.

Η επίδραση αυτή είναι παρόμοια με την επίδραση των πολλαπλών συγκρίσεων, αν και οι δύο περιπτώσεις είναι εννοιολογικά λίγο διαφορετικές .Ένα υποθετικό παράδειγμα θα δείξει ότι τα επαναλαμβανόμενα τεστ σημαντικότητας δεν είναι αμελητέα.

Υποθέτουμε ότι ,σε μία δοκιμή διπλής-κάλυψης, κάθε ασθενής λαμβάνει δύο αναλγητικά χάπια ,ένα A και ένα B , σε διαδοχικές εβδομάδες, με τυχαία σειρά. Στο τέλος της περιόδου της 2^{n5} εβδομάδας κάθε ασθενής δηλώνει την προτίμηση του στο

φάρμακο που έλαβε την 1^η και 2^η εβδομάδα βασισμένος στην ανακούφιση του πόνου. Τα στοιχεία αυτά αποκωδικοποιούνται σε έναν τύπο προτίμησης για το Α ή Β.

Φαίνεται λογικό το να ελέγχουμε τα συσσωρευμένα αποτελέσματα σε οποιοδήποτε επίπεδο ώστε να καταλάβουμε αν υπάρχει σημαντική υπεροχή του Α ή του Β. Το κατάλληλο συμβατικό τεστ, στο νιοστό επίπεδο, θα μπορούσε να είναι εκείνο που βασίζεται στην διωνυμική κατανομή με μέγεθος δείγματος n και με $\pi = \frac{1}{2}$.

Υποθέτουμε ότι τα τεστ διεξάγονται σε ένα διπλής όψης επίπεδο σημαντικότητας 5%. Ο ερευνητής, προχωρά διαδοχικά, ίσως προδιατεθειμένος να σταματήσει αν σε κάποιο επίπεδο επιτευχθεί αυτό το σημαντικό επίπεδο, και τότε είναι που δημοσιεύει τα αποτελέσματα ισχυριζόμενος μια σημαντική διαφορά 5%. Πράγματι αυτή είναι μία σημαντική αξιολόγηση της ένδειξης σε ένα συγκεκριμένο επίπεδο. Η βασική αρχή πιθανοφάνειας που διατυπώθηκε στην τελευταία υποενότητα δείχνει ότι οι σχετικές πιθανοφάνειες των διαφορετικών παραμετρικών τιμών δεν επηρεάζονται από τον κανόνα τερματισμού. Παρόλα αυτά λαμβάνεται υπόψη κάποια επιλογή στοιχείων. Ο ερευνητής έχει έναν μεγάλο αριθμό ευκαιριών να σταματήσει στο επίπεδο σημαντικότητας 5%. Ακόμη κι να η ψευδής υπόθεση είναι αληθής, υπάρχει μια υποδεέστερη πιθανότητα ότι θα βρεθεί κατά την διάρκεια ένα σημαντικό αποτέλεσμα, κι αυτή η πιθανότητα έχει ως φυσικό επακόλουθο να αυξηθεί ξεκάθαρα η διάρκεια διεξαγωγής της δοκιμής.

Στο παράδειγμα αυτό, ο τύπος της διωνυμικής κατανομής δηλώνει ότι είναι αδύνατο να πάρουμε ένα σημαντικό αποτέλεσμα σε ένα επίπεδο διπλής όψης 5% μέχρι να καταχωρηθεί μέγεθος $n=6$, και τότε το σφάλμα τύπου I μικρότερο του 5% θα είναι ίσο με 0,031. Για $n=50$ το σφάλμα τύπου I πιθανόν να έχει αυξηθεί σε 0,171 και για $n=100$ να είναι 0,227. Πράγματι, καθώς το n αυξάνεται, το σφάλμα τύπου I αυξάνεται συνεχώς προσεγγίζοντας το 1.

Για επαναλαμβανόμενα t τεστ για μία συνεχή μεταβλητή απόκρισης, η οποία θεωρείται ότι ακολουθεί κανονική κατανομή, η επίδραση είναι ακόμα περισσότερο εντυπωσιακή. Για $n=100$ η πιθανότητα σφάλματος τύπου I είναι 0.39.

Για να ελέγξουμε την πιθανότητα σφάλματος τύπου I, σε μία πολύ χαμηλή τιμή, όπως το 5%, και για την αξιολόγηση των αποτελεσμάτων σε ένα οποιοδήποτε επίπεδο, απαιτείται ένα πολύ πιο αυστηρό επίπεδο σημαντικότητας. Υποθέτουμε ότι ο κανόνας τερματισμού είναι να σταματήσει η δοκιμή αν τα αθροιστικά αποτελέσματα σε οποιοδήποτε επίπεδο υποδεικνύουν μία βασική διαφορά στο πλασματικό διπλής όψης επίπεδο $2\alpha'$, ή να σταματήσει ύστερα από N στάδια, αν η δοκιμή δεν έχει διακοπεί νωρίτερα. Τίθεται λοιπόν το ερώτημα: Για να κατορθώσουμε να πάρουμε πιθανότητα σφάλματος τύπου I ίση με 2α , σ' ένα επίπεδο σημαντικότητας 5%, ποια τιμή πρέπει να εξαρτάται από το N ; Όσο μεγαλύτερη η τιμή του N , τόσο μικρότερο πρέπει να είναι το $2\alpha'$. Κάποια αποτελέσματα για τις διωνυμικές αποκρίσεις και για τις κανονικές αποκρίσεις φαίνονται παρακάτω στον Πίνακα (2).

Η επιλογή του N συμβολίζει το μεγαλύτερο δειγματικό μέγεθος σ' ένα διαδοχικό τεστ. Η επιλογή του βασίζεται σε διάφορες υποθέσεις. Συγκεκριμένα, όπως και στο κριτήριο 3 αυτής της παραγράφου, ίσως κάποιος επιθυμεί να επιλέξει ένα σχέδιο αλληλουχίας το οποίο δεν ελέγχει την πιθανότητα σφάλματος τύπου I, αλλά έχει και μία ιδιαίτερη ισχύ, ας πούμε $1-\beta$, στο να παρέχει ένα σημαντικό αποτέλεσμα όταν μία βέβαιη εναλλακτική υπόθεση αληθεύει. Στο διωνυμικό τεστ που περιγράφηκε νωρίτερα μία εναλλακτική υπόθεση επεξηγεί ότι η πιθανότητα της προτίμησης για ένα φάρμακο Α, το

οποίο συμβολίζεται με π , είναι κάποια τιμή π_1 διαφορετική από $\frac{1}{2}$. Αν το σχέδιο αλληλουχίας είναι συμμετρικό, αυτόματα θα παρέχει την ίδια ισχύ για $\pi = \pi_{-1}(=1-\pi_1)$ όπως για το π_1 .

Ο πίνακας (3) δείχνει τα μέγιστα δειγματικά μεγέθη, και τα επίπεδα σημαντικότητας χωριστά για κάθε τεστ, για διωνυμικό σχέδιο αλληλουχίας με πιθανότητα σφάλματος τύπου I να είναι $2\alpha = 0.05$ και ισχύ $1-\beta = 0.95$ έναντι διάφορων εναλλακτικών τιμών του π .

Αυτά είναι παραδείγματα επαναλαμβανόμενων τεστ σημαντικότητας (*repeated significance test plans – RST*).

πίνακας (2) : επαναλαμβανόμενα τεστ σημαντικότητας για αθροιστικές διωνυμικές και κανονικές παρατηρήσεις. Εικονικό επίπεδο σημαντικότητας χρησιμοποιείται για κάθε τεστ χωριστά με πιθανότητα σφάλματος τύπου I $2\alpha = 0.05$.

Number of Stages, N	Nominal significance level (two-sided), $2\alpha'$, for individual tests	
	Bionomial	Normal
1	-	0.050
5	-	0.016
10	0.031	0.010
15	0.023	0.008
20	0.022	0.007
50	0.013	0.005
100	0.008	0.004
150	0.007	0.003

πίνακας (3) : Μέγιστο δειγματικό μέγεθος ,N, και εικονικά επίπεδα σημαντικότητας για κάθε test χωριστά , $2\alpha'$, για δυνωμικό σχέδιο RST με πιθανότητα σφάλματος τύπου I $2\alpha = 0.05$, και ισχύ $1 - \beta = 0.95$ έναντι διαφορών τιμών του π διαφορετικών της ψευδούς τιμής του 0.5.

π_1	$2\alpha'$	N
0.95	0.0313	10
0.90	0.0225	16
0.85	0.0193	25
0.80	0.0147	38
0.75	0.0118	61
0.70	0.0081	100

Ομαδικά σχέδια αλληλουχίας

Ένα ομαδικό σχέδιο αλληλουχίας με καθορισμένη πιθανότητα σφάλματος τύπου I, θα μπορούσε να θεωρηθεί σαν μια ειδική περίπτωση επαναλαμβανόμενων test σημαντικότητας με μία μικρή τιμή του N, γνωρίζοντας τώρα ότι καθεμία από τις N παρατηρήσεις είναι μία στατιστική προερχόμενη από ένα σύνολο ξεχωριστών παρατηρήσεων. Ωστόσο χρειάζεται να θεωρήσουμε ένα πιο γενικό πλαίσιο. Το σχέδιο για τα επαναλαμβανόμενα test σημαντικότητας που είδαμε στους πίνακες (2) και (3) ο όρος ότι το εικονικά σημαντικό επίπεδο είναι το ίδιο για όλα τα test χωριστά, δεν είναι αναγκαστικά περιοριστικός.

Παράδειγμα 2:

Ο πίνακας (4) που δίνεται παρακάτω δίνει λεπτομερώς το σχέδιο αλληλουχίας των τριών ομάδων ,που επιτυγχάνουν να έχουν την ίδια ισχύ. Όλα απαιτούν πέντε ισοδύναμα διαστήματα ελέγχου των δεδομένων. Ο πίνακας δείχνει την σταθερή κανονική απόκλιση (τιμή z), υπολογισμένη από την σύγκριση των μέσων των μέχρι τώρα δεδομένων, τα οποία υποδεικνύουν τον τερματισμό της δοκιμής σε κάθε επίπεδο.

Τα χαρακτηριστικά των τριών σχεδίων φαίνονται στον πίνακα (4) και στο σχήμα (1) που περιγράφονται παρακάτω.

πίνακας (4) : Σχέδια των τριών διαδοχικών ομάδων για την δοκιμή που περιγράφηκε στο παράδειγμα (2). Όρια για την σταθερή κανονική απόκλιση (τιμή z) σε ενδιάμεσους και τελικούς ελέγχους.

Οι εισχωρήσεις του πίνακα αυτού προέρχονται από τον East για τα Windows (1999) και από τους Geller και Pocock (1987).

	Pocock	O'Brien-Fleming	Haybittle-Peto
Interim inspection			
1	2.41	4.56	3.29
2	2.41	3.22	3.29
3	2.41	2.63	3.29
4	2.41	2.28	3.29
5	2.41	2.04	1.97
Sample size			
Terminal	155	130	126
Mean on H_0	151	129	126
Mean on H_1	101	103	113

Τα όρια του Pocock (Pocock, 1977) βασίζονται σε επαναλαμβανόμενα σημαντικά τεστ σε ένα καθορισμένο επίπεδο, όπως περιγράψαμε νωρίτερα. Τα όρια για το z καθορίζουν την πιθανότητα σφάλματος τύπου I, ενώ η ισχύς ελέγχεται καθορίζεται από το τερματικό μέγεθος δείγματος όπως καθορίστηκε στον πίνακα (4). Αξίζει να σημειώσουμε ότι το τερματικό δειγματικό μέγεθος ,155, υπερβαίνει το καθορισμένο μέγεθος δείγματος των 126, αλλά αν η εναλλακτική υπόθεση είναι αληθής (στην περίπτωση μας μια σταθερή διαφορά της τάξης του 0.5) ο μέσος του δειγματικού μεγέθους μειώνεται αισθητά, σε 101. Επισημαίνουμε επίσης ότι το όριο τερματισμού για το z υπερβαίνει την καθορισμένη τιμή του δειγματικού μεγέθους του 1.96 ,έτσι, αν η δοκιμή συνεχιστεί στο 5^ο στάδιο με μία τελική τιμή του z μεταξύ 1.96 και 2.41, η ερμηνεία θα ήταν δύσκολη. Όπως αναφέρει και ο Piantadosi (1997) «αυτή είναι μία άβολη θέση για τους ερευνητές».

Τα δύο άλλα σχέδια που φαίνονται στον πίνακα (4) αποφεύγουν σε μεγάλο βαθμό αυτή την δυσκολία. Το σχέδιο του O'Brien-Fleming (OBF- O'Brien & Fleming, 1979) χρησιμοποιεί σταθερά όρια για το αθροιστικό υπόλοιπο τα οποία είναι πολύ κοντά σε αυτά που είναι κατάλληλα για ένα καθορισμένο τεστ δειγματικού μεγέθους σε έναν τελικό έλεγχο. Η συνέπεια είναι ότι τα όρια του z στον k έλεγχο ($k=1,2,\dots,K$) είναι ίσα με εκείνα του τελικού ελέγχου ($k=K$) πολλαπλασιασμένα με $\sqrt{K/k}$, όπου στο παράδειγμα αυτό $K=5$. Αυτά αρχικά έχουν μεγάλο εύρος ,όπως φαίνεται στον πίνακα (4), και συγκλίνουν σε τελικές τιμές αρκετά κοντά με εκείνες που είναι κατάλληλες για ένα μη-διαδοχικό τεστ. Όταν η υπόθεση H_0 είναι αληθής, τα αποτελέσματα της δοκιμής οδηγούν συνήθως στον τερματισμό στον K -έλεγχο, και όπως υποδεικνύεται από τον πίνακα (4), το μέσο δειγματικό μέγεθος είναι κοντά στο μέγιστο, χαμηλότερο για το OBF σχέδιο σε σχέση με το σχέδιο του Pocock. Όταν η H_1 είναι αληθής, τα δύο σχέδια έχουν πολύ κοινές ιδιότητες. Σε κάποιες παραλλαγές του OBF, το ακριβές εύρος των

ορίων για $\kappa=1$ σταματά στην καθιερωμένη τυπική απόκλιση αντιστοιχώντας στο επίπεδο 0.001 $z = 3.29$.

Ο Haybittle (1971) και ο Peto (1976) πρότειναν ότι πρέπει να χρησιμοποιείται μία σταθερή, αλλά μεγάλη τιμή του z για όλες τις ενδιάμεσες αναλύσεις, οι οποίες επιτρέπουν (όπως και στο OBF) να χρησιμοποιείται στο τελικό στάδιο μία τιμή κοντά στην προκαθορισμένη τιμή δειγματικού μεγέθους. Για μία πιθανότητα σφάλματος τύπου I 0.05, διαφορετικών παραλλαγών του σχεδίου Haybittle-Peto χρησιμοποιείται για τα ενδιάμεσα όρια το 3.29 (όπως στον πίνακα (4)) ή το 3.00.

Θεωρήσαμε μέχρι τώρα ότι οι παρατηρήσεις ακολουθούν κανονική κατανομή με γνωστή διακύμανση. Αυτό, φυσικά, μπορεί να είναι πιθανό σε κάποιες περιπτώσεις αλλά, η μεθοδολογία που βασίζεται στην κανονική κατανομή συχνά παρέχει μία χρήσιμη προσέγγιση για ένα μεγάλο φάσμα άλλων καταστάσεων. Παρόλα αυτά, έχουν αναπτυχθεί ειδικές μέθοδοι για άλλους τύπους δεδομένων, που συμπεριλαμβάνουν δυαδικές παρατηρήσεις και χρόνους επιβίωσης. Μία εκτενής περιγραφή ομαδικών διαδοχικών μεθοδολογιών δίνεται από τους Jennison και Turnbull (1920) και τον Whitehead (1999). Οι Geller και Pocock (1987) πινακογράφησαν όρια για διάφορα σχέδια. Επιπλέον ένα πολύ χρήσιμο software δόθηκε από τον East για τα Windows (1999) και για το σύστημα Pest(2000).

Ο Whitehead (1997) αναπτύσσει ένα πολύ γενικό σύστημα διαδοχικών σχεδιασμών για συνεχείς ελέγχους, εφαρμόσιμο στο PEST. Για το καθιερωμένο κανονικό μοντέλο με γνωστή διακύμανση, υπάρχουν όρια που είναι γραμμικά για το γράφημα του αθροιστικού συνόλου.

Τυχαία περικοπή (stochastic curtailment)

Τα σχέδια που περιγράφηκαν παραπάνω παρέχουν έναν πρόωρο τερματισμό όταν κάποια ρεαλιστικά στοιχεία δημιουργήσουν μια διαφορά στην ακρίβεια μεταξύ θεραπειών. Το βασικό κίνητρο γι' αυτό είναι η ηθική ανάγκη να αποφευχθεί η χρήση μιας υποδεέστερης θεραπείας.

Μια κάπως διαφορετική κατάσταση μπορεί να παρουσιαστεί όταν τα ενδιάμεσα αποτελέσματα, για παράδειγμα, δύο θεραπειών είναι παρόμοια, και όταν μπορεί να προβλεφθεί ότι η τελική διαφορά θα μπορούσε σχεδόν βέβαια να μην είναι σημαντική. Οι μέθοδοι για περιορισμό μιας δοκιμής κάτω από αυτές τις συνθήκες έχει προταθεί από πολλούς συγγραφείς. (Schneiderman & Armitage, 1962 ; Lan, 1982 και από τον Ware, 1985). Τα όρια που επιτρέπουν τυχαία περικοπή μπορούν να συμπεριληφθούν σε σχέδια που επιτρέπουν πρόωρο τερματισμό για ακριβής επιδράσεις και είναι εύκολα εφαρμόσιμα με EAST ή PEST.

Αν και η προσέγγιση αυτή φαίνεται να είναι χρήσιμη στο να παρέχει την δυνατότητα μεταβολής προς καταλληλότερες κατευθύνσεις κατά την διάρκεια της έρευνας, ωστόσο υπάρχει κίνδυνος στην τοποθέτηση τόσο σημαντικής βαρύτητας στα προβλεπόμενα αποτελέσματα ενός τελικού τεστ σημαντικότητας. Τα δεδομένα μπορεί να υποδεικνύουν μη σημαντικά θεραπευτικά αποτελέσματα τα οποία όμως είναι σημαντικά για την εκτίμηση, ιδιαίτερα στην περίπτωση της μετα-ανάλυσης.

3.7 Ερμηνεία των αποτελεσμάτων της δοκιμής

Αριθμός που χρειάζεται να αναλυθεί (number needed to treat-NNT)

Μία νέα θεραπεία μπορεί να είναι περισσότερο ακριβή ή λιγότερη αποδεχτή από τους ασθενείς σε σχέση με μία καθιερωμένη θεραπεία. Μια γενική αλλά σημαντική ερώτηση είναι εάν τα οφέλη που προφανώς απονέμονται σε μία θεραπεία δικαιολογούν την χρήση της σε με μεγάλο αριθμό ασθενών που εμπίπτουν σε παρόμοια κατηγορία. Σε κάποιο βαθμό κάτι τέτοιο θα έπρεπε να λαμβάνεται υπόψη από τους ερευνητές της δοκιμής κατά τον σχεδιασμό της μελέτης, και να συζητηθεί στην δημοσιοποιημένη αναφορά της δοκιμής. Στις δοκιμές με δυαδική μεταβλητή απόκρισης, όπως τα στοιχεία δυαδικής συμφόρησης εντός τριών ετών, είναι διαθέσιμος ένας χρήσιμος πίνακας που υποδεικνύει την ισορροπία μεταξύ μελλοντικής μεταχείρισης και οφέλους.

Υποθέτουμε ότι οι πιθανότητες μιας δυσμενούς απόκρισης σε ασθενείς που λαμβάνουν μία νέα ή ελεγχόμενη θεραπεία είναι, αντίστοιχα, π_T και π_C , εκτιμημένα από τις δοκιμές p_T και p_C . Τότε ο απαραίτητος αριθμός των ασθενών που χρειάζονται για την θεραπεία ώστε να αποτραπεί ένα δυσμενές συμπέρασμα είναι $1/(\pi_C - \pi_T)$, και αυτό εκτιμάται από το $1/(p_C - p_T)$. Επισημαίνουμε ότι αυτό είναι το αντίστροφο της μείωσης του απόλυτου κινδύνου, και δεν εκφράζεται καθαρά με τον σχετικό κίνδυνο μείωσης.

Μία δοκιμή που δείχνει έναν μεγάλο σχετικό κίνδυνο μείωσης ίσως να έχει παράλα αυτά έναν μεγάλο εκτιμημένο NNT, επειδή και οι δύο αυτές αναλογίες είναι μικρές, και συνεπώς κρίνονται κατάλληλες για να είναι οριακές τιμές. Αντιθέτως (όπως στο παρακάτω παράδειγμα (3)) ο σχετικός κίνδυνος μείωσης ίσως να είναι μικρός, και επιπλέον μπορεί το NNT να φαίνεται να δικαιολογεί την εκτεταμένη χρήση μιας νέας θεραπείας.

Αφού η μείωση του απόλυτου ρίσκου θα ακολουθεί κατά προσέγγιση κανονική κατανομή, εκτός κι αν το πλήθος των γεγονότων είναι πολύ μικρό, το NNT, θα έχει μία κατανομή με πολύ μακριές ουρές, οι οποίες δεν μπορούν να θεωρηθούν ότι προσεγγίζουν την κανονική κατανομή. Οποιοσδήποτε υπολογισμός δειγματικού λάθους, είναι καλύτερα να υπολογιστεί πρώτα από την απόλυτη τιμή της διαφοράς $p_C - p_T$ και μετά να μετατραπεί σε αμοιβαία κλίμακα αν είναι απαραίτητο. Για παράδειγμα, τα όρια των διαστημάτων εμπιστοσύνης για την διαφορά μπορούν να υπολογιστούν με απλές στατιστικές μεθόδους.

Παράδειγμα 3:

Στην τέταρτη διεθνής μελέτη της επιβίωσης από έμφραγμα (The fourth International Study Of Infarct) οι ομάδες συνεργασίας στην αναφορά των αποτελεσμάτων μιας δοκιμής ευρείας κλίμακας θεραπειών για ασθενείς με υποψία οξέος εμφράγματος, συνδύασαν αυτά τα αποτελέσματα με εκείνα από άλλες σχετικές δοκιμές.

Σε μία τέτοια επισκόπηση, οι συγγραφείς ανέλυσαν την θνησιμότητα τον πρώτο μήνα μετά την έναρξη της θεραπείας σε ασθενείς που έλαβαν τροποποιημένο ένζυμο μηχανισμού απενεργοποίησης (CEI) και σε εκείνες στις ομάδες ελέγχου. Ο αριθμός των θανάτων προς τον αριθμό των ασθενών που θεραπεύτηκαν ήταν 3671/50496 (7.27%)

στην ομάδα CEI, και 3903/50467 (7.73%) στην ομάδα ελέγχου. Οι αναλογίες των θανάτων είναι μικρές αλλά παρόμοιες, και ο σχετικός κίνδυνος είναι 0.94, δηλαδή κοντά στην μονάδα. Ο πληθυσμός δοκιμής είναι επαρκώς μεγάλος ώστε να εξασφαλίσει ότι η διαφορά είναι πολύ σημαντική ($p=0.006$).

Για την παραπάνω περίπτωση έχουμε $p_c - p_T = 0.07734 - 0.07270 = 0.00464$ με σταθερό σφάλμα της τάξης του 0.00166. Συνεπώς, η εκτίμηση NNT είναι $1/0.00464=216$. Οι συγγραφείς σχολιάζουν ότι «τέτοια θεραπεία σώζει 5 ζωές ανά 1000 υπό θεραπεία ασθενείς».

Το 95% των ορίων εμπιστοσύνης για την διαφορά στις αναλογίες, χρησιμοποιώντας κανονικές προσεγγίσεις, είναι $0.00464 \pm (1.96)(0.00166) = 0.00139$ και 0.00789 με όρια απόκρισης για το NNT, $1/0.00789=127$ και $1/0.00139=719$ αντίστοιχα.

Επισημαίνουμε την ασυμμετρία και το εύρος των διαστημάτων εμπιστοσύνης για το NNT. Ακόμα και με μια τέτοια μεγάλη βάση δεδομένων, το NNT δεν μπορεί να καθοριστεί επακριβώς.

3.8 Ειδικό σχεδιασμοί

Διασταυρωμένη δοκιμή (cross-over trial)

Σχεδόν όλες οι προηγούμενες συζητήσεις ασχολήθηκαν με παράλληλες ομάδες, στις οποίες τα άτομα κατανέμονταν σε μία μόνο θεραπευτική ομάδα. Βασισμένοι στις γενικές αρχές του πειραματικού σχεδίου, κάποιος θα αναμένει ότι οι θεραπείες θα μπορούσαν να συγκριθούν με μεγαλύτερη ακρίβεια αν μπορούσαν να χορηγηθούν σε διαφορετικές περιπτώσεις περισσότερες από μία θεραπείες στο ίδιο άτομο. Οι θεραπείες θα μπορούσαν τότε να εκτιμηθούν έναντι ενός ίδιου υποκειμενικού τυχαίου σφάλματος. Ένα επιπρόσθετο πλεονέκτημα θα ήταν το οικονομικό λόγω του το πλήθους των ατόμων, αφού κάθε άτομο θα παρείχε περισσότερες από μία παρατηρήσεις. Ένα σχέδιο στο οποίο οι θεραπείες θα δίνονταν στην σειρά σε κάθε άτομο καλείται διασταυρωμένος σχεδιασμός (cross over design)

Στις περισσότερες, ίσως και στις πιο πολλές κλινικές μελέτες, τέτοιοι σχεδιασμοί είναι μη πραγματοποιήσιμοι, γιατί οι θεραπείες προορίζονται να παράγουν μη αναστρέψιμες αλλαγές για την υγεία του ασθενούς, ή γιατί πρέπει να χορηγηθούν για μία μεγάλη χρονική περίοδο. Ένας διασταυρωμένος σχεδιασμός είναι περισσότερο κατάλληλος για θεραπείες που στοχεύουν σε γρήγορη ανακούφιση των συμπτωμάτων στις χρόνιες ασθένειες, όπου για μεγάλη περίοδο η κατάσταση της υγείας του ασθενούς παραμένει σχεδόν σταθερή. Αυτός ο σχεδιασμός χρησιμοποιείται εκτεταμένα από για τις μελέτες *Phase I* και *Phase II* και για ισοδύναμες δοκιμές.

Ο απλός διασταυρωμένος σχεδιασμός (the simple cross-over design)

Στην απλούστερη μορφή του σχεδιασμού υποθέτουμε ότι συγκρίνουμε δύο θεραπείες A και B όπου μία τυχαία επιλεγμένη ομάδα ασθενών (*group I*) λαμβάνει θεραπείες με την σειρά AB, ενώ μία άλλη ομάδα ασθενών (*group II*) τις λαμβάνει με την σειρά BA. Ίσως το πλήθος των ασθενών στις δύο ομάδες να διαφέρει. Ο σχεδιασμός καλείται απλός διασταυρωμένος σχεδιασμός των δύο θεραπειών των δύο περιόδων (*the simple cross-over design the two-treatments ,two-period design*) ή απλά σχεδιασμός AB/BA. Ο σχεδιασμός έχει το ακόλουθο στήσιμο.

	Run in	Period 1	Wash-out	Period 2
Reading	Z_{ij}	Y_{ij1}		Y_{ij2}
Group I(n_1)	–	A	–	B
Group II(n_2)	–	B	–	A

Εδώ, οι εκτελέσεις είναι στην πραγματικότητα παρατηρήσεις των περισσότερο κατάλληλων μέτρων της συνθήκης του ασθενούς (δηλ. τεστ μετρήσεως για το αναπνευστικό, την συχνότητα της επίθεσης κάποιων τύπων κτλ.). Οι δύο περίοδοι της χορήγησης χωρίζονται από μια περίοδο «καθαρίσματος» (*wash-out period*) ,δηλαδή την περίοδο που καθιστά την συνθήκη του ασθενούς τέτοια ώστε να μπορεί να επιστρέψει καθαρός και σε ένα όσο το δυνατόν πιο ανεπηρέαστο επίπεδο από την προηγούμενη θεραπεία που έλαβε. Η περίοδος εισχώρησης (*run-in*) είναι προαιρετική , αλλά ίσως παρέχει για τον ασθενή την ευκαιρία να χαλαρώσει και για τον ερευνητή να κάνει μία βασική παρατήρηση z_{ij} . Οι αποκρίσεις y_{ij1} και y_{ij2} συχνά θα φτιάχνουν προς το τέλος των περιόδων της θεραπείας, ώστε να επιτρέπουν την μέγιστη επίδραση από τις θεραπείες , αλλά ίσως είναι μέσοι όροι ή μέγιστοι μιας σειράς παρατηρήσεων που πάρθηκαν καθ' όλη την διάρκεια αυτών των περιόδων. Οι παρατηρήσεις ίσως επίσης φτιάχτηκαν στο τέλος της wash-out περιόδου ,απευθείας πριν την 2^η περίοδο ή αμέσως μετά από αυτήν.

Οι συμβολισμοί για τις εκτελέσεις είναι ως ακολούθως:

ο δείκτης i υποδεικνύει την ομάδα-group και μερικές φορές καλείται *sequence* (με τιμές $i=1$ για το group I, $i=2$ για το group 2). Το j προσδιορίζει το άτομο εντός της ομάδας. Ο τελικός δείκτης υποδεικνύει την πρώτη και δεύτερη απόκριση. Οι κρίσιμες μετρήσεις είναι οι ενεργές αποκρίσεις y_{ij1} και y_{ij2} ,αλλά, όπως θα δούμε, η βασική εκτέλεση z_{ij} είναι επίσης χρήσιμη.

Αρχικά θα θεωρήσουμε ένα απλό επιπρόσθετο μοντέλο για τις ενεργές αποκρίσεις.

$$Y_{ijk} = a_{ik} + \xi_{ij} + \epsilon_{jk}$$

όπου ο όρος a είναι κοινός σε όλα τα άτομα μιας συγκεκριμένης ομάδας και μιας συγκεκριμένης περιόδου, και εξαρτάται από την θεραπεία που λαμβάνει ο ασθενής. Ο όρος ξ είναι ένα χαρακτηριστικό κάθε ατόμου , και ο όρος ϵ είναι ένα τυχαίο σφάλμα το

οποίο θεωρούμε ότι ακολουθεί κανονική κατανομή $N(0, \sigma^2)$. Η ερώτηση που τίθεται τώρα είναι πώς ο όρος α αντικατοπτρίζει τις επιδράσεις των ομάδων, των περιόδων και των θεραπειών. Το μοντέλο μας για τον a_{ik} είναι το ακόλουθο.

	Period 1	Period 2
Group I (AB)	$\mu + \tau_A + \pi_1$	$\mu + \tau_B + \pi_2 + \gamma_{AB}$
Group II (BA)	$\mu + \tau_B + \pi_1$	$\mu + \tau_A + \pi_2 + \gamma_{BA}$

Εδώ το μ είναι ένας γενικός μέσος, ο όρος τ αντιπροσωπεύει τις επιδράσεις της θεραπείας, ο όρος π αντιπροσωπεύει τις περιόδους των επιδράσεων και ο όρος γ συμβολίζει το γινόμενο (θεραπεία)*(περίοδο αλληλεπίδρασης). Θα αναφερθούμε στον όρο της αλληλεπίδρασης αργότερα, αλλά προς το παρόν θα χρησιμοποιήσουμε το μοντέλο απλώς με τις κύριες επιδράσεις. Αυτό σημαίνει ότι θα θεωρήσουμε ότι και οι δύο όροι γ είναι μηδέν.

Παράδειγμα 4:

Σε μία κλινική δοκιμή για ένα νέο φάρμακο για την θεραπεία της ακράτειας, σε κάθε έναν από τους 29 ασθενείς δόθηκε ένα φάρμακο για μία χρονική περίοδο 14 ημερών και ένα placebo για μία χωριστή περίοδο 14 ημερών, με την σειρά της χορήγησης να έχει επιλεγεί τυχαία για κάθε ασθενή. Ο πίνακας (5) δείχνει τον αριθμό των «στεγνών» νυκτών κατά την διάρκεια κάθε θεραπευτικής περιόδου.

Ένα τεστ για την σχετική αποτελεσματικότητα ενός φαρμάκου και του placebo θα μπορούσε να επιτευχθεί από τις συμπληρωμένες σειρές των 29 ζευγών παρατηρήσεων. Θα χρειαστούμε τις 29 διαφορές για φάρμακο-placebo. Για την ομάδα I, αυτές οι διαφορές (για περίοδο 1 και περίοδο 2) συμβολίζονται με d_{1j} . Για την ομάδα II θα χρειαστεί να αλλάξουμε το σύμβολο των διαφορών σε d_{2j} . Γράφοντας $d_{1j}^* = d_{1j}$ και $d_{2j}^* = -d_{2j}$, μπορούμε να κάνουμε ένα τεστ βασισμένο στις 29 τιμές των d_{ij}^* . Αυτό δίνει:

$$\bar{d}^* = 2.172$$

$$s^2 = 11.005$$

$$SE(\bar{d}^*) = \sqrt{11.005/28} = 0.616$$

$$t = \bar{d}^*/SE(\bar{d}^*) = 3.53 \text{ με } 28 \text{ βαθμούς ελευθερίας (p=0.001)}$$

Πίνακας (5) : Ο αριθμός των «στεγνών» νυχτών από τις 14 σε θεραπευμένους ασθενείς με ακράτεια μετά την χορήγηση φαρμάκου και placebo

Period:	1	2		
Treatment:	A (drug)	B (placebo)	Difference	Sum
Patient number, j	y_{j1}	u_{j2}	(1-2) d_j	(1+2) $+/$
1	8	5	3	13
2	14	10	4	24
3	8	0	8	8
4	9	7	2	16
5	11	6	5	17
6	3	5	-2	8
7	6	0	6	6
8	0	0	0	0
9	13	12	1	25
10	10	2	8	12
11	7	5	2	12
12	13	13	0	26
13	8	10	-2	18
14	7	7	0	14
15	9	0	9	9
16	10	6	4	16
17	2	2	0	4

<i>Group II</i>				
Period:	1	2		
Treatment:	B (placebo)	A (drug)	Difference	Sum
Patient number, j	y_{ji}	y_{ji}	(1-2) d_{2j}	(1+2) $+/$
1	12	11	1	23
2	6	8	-2	14
3	13	9	4	22
4	8	8	0	16
5	8	9	-1	17
6	4	8	-4	12
7	8	14	-6	22
8	2	4	-2	6
9	8	13	-5	21
10	9	7	2	16
11	7	10	-3	17
12	7	6	1	13

Η παραπάνω ανάλυση υποστηρίζει ισχυρά ότι το φάρμακο αυξάνει τον αριθμό των «στεγνών» βραδιών σε σύγκριση με το placebo. Παρόλα αυτά, είναι ανεπαρκές γιατί δεν έλαβε υπόψη την σειρά με τη οποία χορηγήθηκαν τα φάρμακα στους ασθενείς. Κάτι τέτοιο μπορεί να θεωρηθεί ιδιαίτερα σημαντικό αν υπήρχε μια συστηματική περίοδος αντίδρασης. Για παράδειγμα, αν οι ασθενείς έχουν την τάση να αισθάνονται μεγαλύτερη ανακούφιση την περίοδο 1 σε σχέση με την περίοδο 2.

Οι μέσες αποκρίσεις στις δύο περιόδους για τις δύο ομάδες είναι οι εξής

Period	1	2	Difference (1-2)
Group I : Treatment	A	B	
	$\bar{y}_{11} = 8.118$	$\bar{y}_{12} = 5.294$	$\bar{d}_1 = 2.824$
Group II : Treatment	B	A	
	$\bar{y}_{21} = 7.667$	$\bar{y}_{22} = 8.917$	$\bar{d}_2 = -1.250$

Για να ελέγξουμε την παρουσία μιας θεραπευτικής επίδρασης μπορούμε να εξετάσουμε τις διαφορές μεταξύ \bar{d}_1 και \bar{d}_2 , αγνοώντας τους όρους αλληλεπίδρασης $E(\bar{d}_1 - \bar{d}_2) = 2(\tau_A - \tau_B)$. Αυτή η σύγκριση περιέχει ένα t-test δύο δειγμάτων, δίνοντας

$$t = (\bar{d}_1 - \bar{d}_2) / SE(\bar{d}_1 - \bar{d}_2) \text{ με } n_1 + n_2 - 2 \text{ DF}$$

Όπου $SE(\bar{d}_1 - \bar{d}_2) = \sqrt{S_d^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$ και S_d^2 η εκτίμηση της διακύμανσης d_{ij} που συνδέεται

και με τις δύο ομάδες. Στο παράδειγμα αυτό $\bar{d}_1 - \bar{d}_2 = 4.074 / 1.237 = 3.29$ με 27 DF ($p=0,003$). Ένα αποτέλεσμα που πλησιάζει αυτό που δόθηκε νωρίτερα από την ανάλυση.

Αν υπάρχει μια θεραπευτική επίδραση η προσδοκία της διαφοράς $\bar{d}_1 - \bar{d}_2$ θα είναι $2(\tau_A - \tau_B)$. Η θεραπευτική επίδραση $\tau_A - \tau_B$ έχει εκτιμηθεί συνεπώς ως $\frac{1}{2}(\bar{d}_1 - \bar{d}_2) = 2.037$ «στεγνές» νύχτες από τις 14 συνολικά, και το σταθερό σφάλμα του ως $\frac{1}{2}SE(\bar{d}_1 - \bar{d}_2) = 0.618$. Το 95% διάστημα εμπιστοσύνης για την θεραπευτική επίδραση είναι επομένως $2.037 \pm (2.052)(0.618) = 0.77$ σε 3.31 ημέρες. Ο πολλαπλασιαστής εδώ είναι $t_{27,0.05} = 2.052$.

Για να ελέγξουμε την περίοδο επίδρασης χρησιμοποιούμε $\bar{d}_1 + \bar{d}_2$, αφού, αγνοώντας τους χρόνους επίδρασης έχουμε, $E(\bar{d}_1 + \bar{d}_2) = 2(\pi_1 - \pi_2)$. Το ίδιο σταθερό σφάλμα

χρησιμοποιείται για $\bar{d}_1 - \bar{d}_2$. Στο παράδειγμα αυτό το αποτέλεσμα του τεστ δίνει $t=1.574/1.237=1.27$ με 27DF ($p=0.21$) μην παρέχοντας στοιχεία για την περίοδο επίδρασης.

Θεραπεία × περίοδος αλληλεπίδρασης και διασταύρωσης

Η ανάλυση που έγινε στο παράδειγμα (4) βασίζεται στην υπόθεση ότι η απόκριση επηρεάζεται επιπρόσθετα από την θεραπεία και την περίοδο των επιδράσεων, δηλαδή ότι δεν υπάρχει θεραπεία×περίοδος αλληλεπίδρασης (TP). Μια τέτοια αλληλεπίδραση στο παραπάνω παράδειγμα συμβολίζεται με τους όρους γ_{AB} και γ_{BA} . Επισημαίνουμε ότι μόνο η διαφορά τους έχει σημασία: αν υπήρχαν δύο όροι που είναι ίσοι θα μπορούσαν να απορροφηθούν από την παράμετρο π_2 της περιόδου και θα μπορούσαμε να επιστρέψουμε σε ένα μοντέλο μη-αλληλεπίδρασης. Υπάρχουν διάφορες πιθανές αιτίες για μία αλληλεπίδραση.

1. Αν η περίοδος «κάθαρσης» μεταξύ των δύο θεραπειών είναι πολύ μικρή, ίσως επηρεαστεί η απόκριση στην περίοδο 2 από την θεραπεία που έλαβε ο ασθενής τόσο την περίοδο 1 όσο και την περίοδο 2.
2. Ακόμα και αν η περίοδος «κάθαρσης» είναι αρκετά μεγάλη ώστε να καθαρίσει ο οργανισμός από το φάρμακο της πρώτης περιόδου, οι ψυχολογικές και σωματικές επιδράσεις ίσως παραμείνουν κατά την περίοδο 2.
3. Αν υπάρχει μία ισχυρή περίοδος επίδρασης, τότε αλλάζοντας το γενικό επίπεδο της απόκρισης των περιόδων 1 και 2, υπάρχει περίπτωση να αλλάξει η θεραπευτική περίοδος γιατί εξαρτάται άμεσα από την κλίμακας μέτρησης.

Ένα τεστ που ελέγχει το αν υπάρχει αλληλεπίδραση TP μπορεί να επιτευχθεί και με τον ακόλουθο τρόπο. Το άθροισμα των δύο αποκρίσεων για το άτομο j στην ομάδα i δίνεται από τον τύπο

$$s_{ij} = y_{ij1} + y_{ij2}$$

Αν δεν υπάρχει αλληλεπίδραση TP, η προσδοκία του s_{ij} για αυτό το άτομο είναι, σύμφωνα με το παράδειγμα (3), $2\mu + \xi_{ij} + (\tau_A + \tau_B) + (\pi_1 + \pi_2)$. Αφού τα άτομα έχουν κατανομηθεί τυχαία, το ξ_{ij} μπορεί να ακολουθήσει τυχαία κατανομή με τον ίδιο μέσο (ο οποίος μπορεί να θεωρηθεί 0 λόγω του ότι ο μέσος υπολογίζεται από το μ), και μία διακύμανση σ_0 . Οι προσδοκίες των s_{1j} και s_{2j} είναι συνεπώς ίσες. Από την άλλη πλευρά, αν υπάρχει μία αλληλεπίδραση TP, οι προσδοκίες θα διαφέρουν. Για παράδειγμα αν $\gamma_{AB} > \gamma_{BA}$ -όπως συμβαίνει αν οι δύο θεραπείες έχουν διαφορετικές διασταυρωμένες επιδράσεις- η προσδοκία του s_{1j} θα υπερβαίνει αυτήν του s_{2j} . Παρόλα αυτά μπορούμε να ελέγξουμε τις διαφορές ανάμεσα στις δύο τιμές μέσων των s_{ij} , \bar{s}_1 και \bar{s}_2 , χρησιμοποιώντας πάλι ένα t-test δύο δειγμάτων (two-sample t-test). Στην περίπτωση

αυτή, η εκτίμηση διακύμανσης εντός της ομάδας ,που συμβολίζεται με s_s^2 , είναι η εκτίμηση του s_{ij} , ενώ προηγουμένως ήταν του d_{ij} . Τέλος, εφόσον η s_{ij} ποικίλει λόγω μεταβλητότητας ανάμεσα στα άτομα, πρέπει να αναμένουμε ότι $s_s^2 > s_d^2$.

Παράδειγμα 4(συνέχεια):

Οι τιμές των s_{ij} δίνονται στον πίνακα (5). Οι δύο μέσοι είναι $\bar{s}_1 = 13.412$ και $\bar{s}_2 = 16.583$. Η συζευγμένη εκτίμηση διακύμανσης (*pooled variance estimate*) είναι $s_s^2 = 41.890$ και το σφάλμα δίνεται από τον τύπο,

$$SE(\bar{s}_1 - \bar{s}_2) = \sqrt{[(41.890)(\frac{1}{17} + \frac{1}{12})]} = 2.440$$

Όμως, $t = (13.412 - 16.583) / 2.440 = -1.30$ με 1 DF ($p = 0.20$). Δεν υπάρχει ξεκάθαρο στοιχείο μιας αλληλεπίδρασης TP, και συνεπώς η προηγούμενη ανάλυση φαίνεται δικαιολογημένη.

Συνήθως για την ύπαρξη μιας TP αλληλεπίδρασης θα υπάρχει μία μικρή ένδειξη ελέγχοντάς και εκτιμώντας την θεραπευτική επίδραση από το σύνολο των δεδομένων. Η περίοδος 1 παρέχει ένα έγκυρο τεστ της A έναντι της B, αφού τα άτομα κατανεμήθηκαν τυχαία στις δύο θεραπείες. Η περίοδος 2 είναι αμφιλεγόμενης τιμής, καθώς τα άτομα έχουν υποβληθεί σε άλλες θεραπείες προτού εισέλθουν στην περίοδο 2.

Μία λογική διαδικασία φαίνεται να είναι το να εκτιμηθεί η θεραπευτική επίδραση από τη διαφορά ανάμεσα στους δύο μέσους $\bar{y}_{11} - \bar{y}_{21}$. Δυστυχώς, υπάρχουν πολλές δυσκολίες σχετικές με την προσέγγιση αυτή. Συγκεκριμένα:

1. Η σύγκριση των μέσων των δύο περιόδων υπόκειται στην μεταξύ τους διακύμανση, και είναι συνεπώς λιγότερο ακριβής από την συνολική cross-over προσέγγιση.
2. Το τεστ για τον έλεγχο ύπαρξης αλληλεπίδραση TP, βασισμένο στην διαφορά $\bar{s}_1 - \bar{s}_2$, επηρεάζεται από την διαφορά που παρουσιάζουν τα άτομα μεταξύ τους. Έτσι μπορεί να υπάρχει μία σημαντική αλληλεπίδραση TP αλλά να μην μπορεί να εντοπιστεί. Η ακρίβεια της χασούρας, τόσο εδώ όσο και στην σύγκριση θεραπειάς στο 1 μπορεί να μετριαστεί από την χρήση των z_{ij} , που προέκυψαν από τις παρατηρήσεις που έγιναν κατά τη διάρκεια των θεραπευτικών περιόδων.
3. Βάση της εικασίας που έχει γίνει πρέπει να εφαρμοστεί μία διαδικασία δύο επιπέδων: Αρχικά πρέπει να ελεγχθεί η αλληλεπίδραση TP και αν είναι σημαντική η θεραπευτική επίδραση τότε ελέγχεται και εκτιμάται από την περίοδο 1. Αν η TP δεν είναι σημαντική τότε χρησιμοποιείται η συνηθισμένη crossover ανάλυση. Ο Freeman επισήμανε ότι αυτό δίνει ένα πολύ σημαντικό συμπέρασμα. Αν η ψευδής υπόθεση είναι αληθής, δηλαδή οι θεραπευτικές και οι TP επιδράσεις

είναι μηδέν, και όλα τα τεστ γίνονται ,για παράδειγμα, σε ένα επίπεδο σημαντικότητας 5%, τότε ολόκληρη η πιθανότητα σφάλματος τύπου I μπορεί να είναι μεγάλη της τάξης του 9.5%. Αν όμως οι TP επιδράσεις δεν είναι μηδέν η αύξηση του σφάλματος τύπου I μπορεί να είναι πολύ μεγαλύτερη.

4. Για τον ίδιο λόγο, αν υπάρχει μία μη μηδενική TP επίδραση, η εκτίμηση της θεραπευτικής επίδρασης θα έχει εμφανίσει συστηματικό σφάλμα στο υποσύνολο των περιπτώσεων όπου χρησιμοποιήθηκε η εκτίμηση της περιόδου 1. Αν δεν υπάρχει επίδραση TP, το σφάλμα εξαφανίζεται.

Άλλες μεταβλητές απόκρισης

Σε οτιδήποτε έχουμε αναφερθεί μέχρι τώρα, έχουμε υποθέσει ότι η μεταβλητή απόκρισης είναι συνεχής και ότι ακολουθεί κανονική κατανομή. Και γενικά η συγκεκριμένη εικασία γίνεται στις περισσότερες έρευνες εκτεταμένων σχεδίων. Τα πιο πολλά από τα σημεία που αφορούν την προαναφερθείσα στρατηγική σχεδιασμού, μπορούν να εφαρμοστούν επιπλέον και για μη-κανονικά δεδομένα απόκρισης, αν και υπάρχουν πια ειδικές μέθοδοι για τέτοιες καταστάσεις. Ο Tudor και ο Koch (1994) περιγράφουν τις μη-παραμετρικές μεθόδους και οι Kenward και Jones (1994) περιγράφουν μοντέλα για δυαδικές και σειριακές κατηγορίες δεδομένων, βασισμένοι κυρίως σε μετατροπές logistic.

Δοκιμές n-of-1

Σε οποιαδήποτε κλινική μελέτη, οι ασθενείς μπορεί να εμφανίσουν διαφορετικές αντιδράσεις στην θεραπεία. Κάτι τέτοιο συνήθως συμβαίνει γιατί υπάρχει μία αλληλεπίδραση άτομο×θεραπεία. Σε μία δοκιμή παράλληλων ομάδων όμως, παρά το γεγονός ότι είναι πιθανή η εμφάνιση αλληλεπιδράσεων μεταξύ θεραπειών και κάποιων υποομάδων ατόμων, δεν μπορεί να απομονωθεί η αλληλεπίδραση από το τυχαίο σφάλμα.

Από την άλλη σε έναν εκτεταμένο cross-over σχεδιασμό, ίσως είναι πιθανό να καταλάβουμε αν οι θεραπευτικές επιδράσεις διαφέρουν μεταξύ των ατόμων, αν και αυτό δεν συμβαίνει συχνά. Σε αυτό τον απλό cross-over σχεδιασμό, δεν φαίνεται να συμβαίνει κάτι τέτοιο, καθώς η αλληλεπίδραση ασθενή×θεραπεία ενσωματώνεται εντός του όρου του σφάλματος του ατόμου.

Η δοκιμή n-of-1 (Guyatt, 1990) είναι μία προσπάθεια να επιτύχουμε χωριστά για κάθε άτομο μια αξιόπιστη εκτίμηση της θεραπευτικής επίδρασης. Στις πιο πολλές περιπτώσεις δοκιμών αυτό δεν είναι δυνατόν να γίνει χωρίς τον κατάλληλο αριθμό επαναλήψεων, κάτι που συνεπάγεται μία αρκετά μεγάλη σειρά αλλαγών στην θεραπεία. Αυτό έχει ως αποτέλεσμα να αυξάνεται η πιθανότητα να αλλάξουν οι θεραπευτικές επιδράσεις με τον χρόνο και να αυξηθεί το ρίσκο να αποχωρήσουν ασθενείς από την δοκιμή.

Το απλούστερο και ίσως πιο ικανοποιητικό σχέδιο είναι το να χορηγηθούν θεραπείες με μία τυχαία σειρά. Τότε, καθεμία από τις καταγραφές που έχουν γίνει μπορεί να θεωρηθεί ως ανεξάρτητη και μπορούν να αναλυθούν από κλασσικές μεθόδους, όπως για παράδειγμα με ένα two-sample t test. Αν η σειρά των παρατηρήσεων αποτελείται από

άτομα που έχουν την τάση να συσχετίζονται, τότε η επίδραση από την αλλαγή της σειράς θα αντικατοπτρίζεται από μία αύξηση του τυχαίου σφάλματος. Αν κριθεί απαραίτητο μπορούμε να μειώσουμε την επίδραση της παραπάνω τάσης με έναν συστηματικό σχεδιασμό όπως τα εναλλασσόμενα blocks (permuted blocks), καθένα από τα οποία περιέχει μία τυχαία τροποποίηση των θεραπειών. Τέτοιοι σχεδιασμοί όμως καθιστούν αρκετά δύσκολο το να καλυφθεί η ταυτότητα της θεραπείας, και αυτό είναι πολύ σημαντικό σε μία δοκιμή ενός μόνο ασθενούς που απαιτείται σε ένα μεγάλο βαθμό η συνεργασία του.

Γενικά, δεν φαίνεται πολύ πιθανό να κατασκευάζονται ευρέως n-of-1 δοκιμές που παρέχουν προσωπικές οδηγίες για την επιλογή της θεραπείας για όλους τους ασθενείς. Και είναι συνεπώς πιο συνηθισμένο η επιλογή να γίνει από τις βασικές ενδείξεις που παρέχονται από τις καθιερωμένες δοκιμές στις ομάδες ασθενών.

Ισοδύναμες δοκιμές

Ο σκοπός μιας ισοδύναμης δοκιμής δεν είναι τόσο το να εντοπίσει πιθανές διαφορές στην αποτελεσματικότητα, όσο το να δείξει ότι εντός στενά καθορισμένων ορίων οι θεραπείες είναι, εξίσου αποτελεσματικές. Αυτή η ανάγκη εμφανίζεται σε δύο διαφορετικά κείμενα. Μία *βιοϊσοδύναμη δοκιμή* είναι μία Phase I μελέτη, που το πιθανότερο είναι να έχει κατασκευαστεί από κάποια φαρμακευτική εταιρεία, με σκοπό να συγκρίνει δύο διαφορετικούς τρόπους ή μεθόδους για την χορήγηση κάποιου φαρμάκου. Η τάση είναι να θεωρούνται οι μέθοδοι συμβατές. Η μεταβλητή απόκριση θα βασίζεται σε φαρμακοκινητικές μελέτες, στις οποίες για κάθε άτομο γίνεται μέτρηση της συγκέντρωσης του φαρμάκου στο αίμα, κατά την διάρκεια μιας μικρής περιόδου μετά την χορήγηση, και καταγράφονται κάποια σημαντικά χαρακτηριστικά της καμπύλης συγκέντρωσης-χρόνου, όπως η περιοχή κάτω από την καμπύλη. Συχνά τέτοιες μεταβλητές εμφανίζονται να είναι διαστρεβλωμένα κατανεμημένες, γι αυτό εφαρμόζεται λογαριθμική τροποποίηση για να απλοποιηθεί η ανάλυση. Οι απαιτήσεις συνήθως σχετίζονται με την αναλογία των φαρμακευτικών αποκρίσεων για τους διαφορετικούς τρόπους, και αυτό μπορεί να εκφραστεί χονδρικά λαμβάνοντας υπόψη τις διαφορές μεταξύ των μέσων των λογαριθμικών μετατροπών, οι οποίες απαιτείται να βρίσκονται εντός προκαθορισμένων ορίων. Οι βιοϊσοδύναμες δοκιμές τις περισσότερες φορές περιλαμβάνουν πολύ μικρούς αριθμούς ατόμων, και συχνά επίσης ακολουθούν έναν απλό σχεδιασμό cross-over, με τις κατάλληλες βέβαια προφυλάξεις που θα βοηθήσουν το αν αποφευχθούν οι μεταφορές.

Ένα άλλο σενάριο είναι μία δοκιμή να συγκρίνει τις αποτελεσματικότητες της αγωγής των δύο θεραπειών, όπου μία καινούργια θεραπεία, T, συγκρίνεται με μία καθιερωμένη θεραπεία, S.

Το ενδιαφέρον των κέντρων στρέφεται στο αν η μεταβλητή απόκριση για το S διαφέρει από αυτήν του T, λιγότερο από κάποια στενά όρια που έχουν τεθεί. Για τον λόγο αυτό γίνεται one-side σύγκριση, στην οποία ένα μικρό μειονέκτημα της T είναι ανεκτό, γιατί παράγει λιγότερες σοβαρές επιδράσεις σε σχέση με την S, ή με άλλα λόγια είναι αποδεκτή, και κανένα όριο δεν χρειάζεται να αντικατασταθεί για το οποιοδήποτε πιθανό

πλεονέκτημα της T σε σχέση με την S. Συχνά για τέτοιες καταστάσεις χρησιμοποιούμε τον όρο *non-inferiority* δοκιμή.

Μία σημαντική επισήμανση είναι το ότι ένα τεστ της ψευδούς υπόθεσης της μηδενικής διαφοράς μπορεί να υποδεικνύει ότι οι θεραπείες είναι τελείως ασυσχέτιστες είτε γιατί μία μικρή μη μηδενική πραγματική διαφορά είναι ίσως αποδεκτή αν βρίσκεται μέσα στο επιτρεπτό εύρος της ισοδυναμίας, αλλά επίσης και γιατί μία μη σημαντική διαφορά δεν δίνει απαραίτητα κάποια ένδειξη για το εάν ή όχι η πραγματική διαφορά βρίσκεται εντός των επιτρεπτών διαστημάτων.

Υποθέτουμε ότι η παράμετρος θ μετρά την διαφορά στην αποτελεσματικότητα μεταξύ των S και T με τις περισσότερες περιπτώσεις να δείχνουν προτίμηση στην T. Σε μία διπλής όψης ισοδύναμη δοκιμή (two-sided test), τα όρια της ισοδυναμίας ίσως συμβολίζονται με $\theta_L < 0$ και $\theta_U > 0$. Ο δείκτης είναι ικανός να ισχυριστεί ισοδυναμία αν μπορεί να επιβεβαιώσει ότι η θ βρίσκεται εντός του εύρους (θ_L, θ_U) . Μία απλή προσέγγιση για να επιβεβαιωθεί ισοδυναμία είναι αν ένα κατάλληλο διάστημα εμπιστοσύνης για το θ βρίσκεται ολόκληρο μέσα σε αυτό το εύρος.

Για τον σκοπό αυτό, υποθέτουμε ότι χρησιμοποιούμε $100(1-2\alpha)\%$ όρια εμπιστοσύνης. Δηλαδή 95% όρια που έχουν $\alpha = 0.025$. Μία ισοδύναμη προσέγγιση θα ήταν να επιλέξουμε τεστ σημαντικότητας που χρησιμοποιούν στις ψευδείς υποθέσεις καθένα από τα ισοδύναμα όρια.

1. Ελέγχουμε την ψευδή υπόθεση H_L στην οποία $\theta = \theta_L$ έναντι της μονόπλευρης (one-sided) εναλλακτικής υπόθεσης $\theta > \theta_L$, με μονόπλευρο επίπεδο σημαντικότητας α .
2. Ελέγχουμε την ψευδή υπόθεση H_U στην οποία $\theta = \theta_U$ έναντι της μονόπλευρης (one-sided) εναλλακτικής υπόθεσης $\theta < \theta_U$, με μονόπλευρο επίπεδο σημαντικότητας α .

Η ισοδυναμία επιβεβαιώνεται αν και μόνο αν απορριφθούν και οι δυο αυτές υποθέσεις. Για μία *non-inferiority* δοκιμή, μόνο το τεστ 1 είναι απαραίτητο και ανταποκρίνεται στην απαίτηση ότι ένα μονόπλευρο $100(1-2\alpha)\%$ διάστημα εμπιστοσύνης, επεκταμένο στο άπειρο, πρέπει να πέφτει ολόκληρο στο δεξί όριο του θ_U .

Τυχαιοποιημένα συμπλέγματα (cluster randomization)

Μέχρι τώρα έχουμε υποθέσει ότι κάθε άτομο χωριστά έχει τοποθετηθεί τυχαιοποιημένα σε διαφορετική θεραπευτική ομάδα. Σε κάποιες καταστάσεις ίσως είναι πιο βολικό να θεωρήσουμε συμπλέγματα (clusters) μέσα στα οποία τα άτομα θα πέφτουν φυσικά. Ακολουθούν μερικά σχετικά παραδείγματα

1. Σε μία σύγκριση των σχεδιασμών εμβολιασμού σε μία αγροτική κοινότητα, ίσως είναι πιο πρακτικό να εφαρμόσουμε το ίδιο σχέδιο για όλους τους κατοίκους του συγκεκριμένου χωριού. Ένα παρόμοιο σχέδιο υιοθετήθηκε για μία δοκιμή του συμπληρώματος βιταμίνης A σε παιδιά προσχολικής ηλικίας στην Ινδονησία.

2. Σε μία σύγκριση συγκεκριμένων στρατηγικών για παιδιά σχολικής ηλικίας σε αστικές περιοχές, ίσως είναι περισσότερο πρακτικό να εφαρμόσουμε την ίδια στρατηγική για όλα τα παιδιά ενός συγκεκριμένου σχολείου, ή τουλάχιστον να χρησιμοποιήσουμε την τάξη ως το cluster της συγκεκριμένης περίπτωσης.
3. Μία προσέγγιση για την θεραπεία σεξουαλικά μεταδιδόμενων ασθενειών σε αγροτικές περιοχές περιλαμβάνει ένα πρόγραμμα βελτίωσης της εκπαίδευσης του προσωπικού, της βελτίωσης των ιατρικών μηχανημάτων καθώς επίσης και τον εφοδιασμό περισσότερο αξιόπιστων φαρμάκων. Κατά την σύγκριση αυτού του προγράμματος με ένα ελεγχόμενο σχέδιο, το φυσικό cluster μπορεί να θεωρηθεί το κέντρο υγείας.
4. Η Κοινότητα Παρεμβατικής Δοκιμής για την Διακοπή του Καπνίσματος (The Community Intervention Trial for Smoking Cessation-COMMIT) τυχαιοποίησε 22 Νότιες Αμερικανικές Κοινότητες σε μία βασική κοινότητα δοκιμής.

Η παραπάνω προσέγγιση καλείται *cluster randomization, group randomization ή community-based intervention*. Η ποικιλία στον σχεδιασμό αυτών των μελετών συμπεριλαμβάνει την χρήση αντίστοιχων cluster σχεδιασμών, που βασίζονται στην ομοιότητα των βασικών μεταβλητών ή κάποιων εξ αυτών, και επιπλέον στον περιορισμό των μέτρων απόκρισης για ένα τυχαίο δείγμα ατόμων σε κάθε cluster, παρά σε ολόκληρο το cluster του πληθυσμού.

Το πιο σημαντικό σημείο της cluster randomization είναι ότι, αν και τα μέτρα απόκρισης φτιάχτηκαν πάνω σε ξεχωριστά άτομα, οι θεραπευτικές συγκρίσεις είναι λιγότερο ακριβείς σε σχέση με τις αντίστοιχες στις οποίες κάθε άτομο θα ήταν τυχαιοποιημένο. Η αιτία είναι ότι οι αποκρίσεις των ατόμων που ανήκουν στο ίδιο cluster είναι πιθανόν να υποδείξουν θετικό συσχετισμό. Στην ακραία περίπτωση του τέλει συσχετισμού, χάνεται το πλεονέκτημα αναπαραγωγής εντός των cluster, με συνέπεια το αποτελεσματικό δειγματικό μέγεθος να είναι ο αριθμός των clusters και όχι ο αριθμός των ατόμων.

Υποθέτουμε ότι υπάρχουν $2K$ clusters, που είναι τοποθετημένα τυχαία σε δύο θεραπείες, με K clusters σε καθεμία από αυτές, και ότι το i^{th} cluster περιέχει n_i άτομα, με ένα συνολικό μέγεθος δείγματος $N = \sum n_i$ για κάθε θεραπεία. Μπορεί να αποδειχθεί ότι η διακύμανση για ολόκληρο τον μέσο απόκρισης σε κάθε ομάδα είναι «παραφουσκωμένη» εξαιτίας του παράγοντα $1 + p[(\sum n_i^2 / N) - 1]$, που καλείται επίδραση του σχεδιασμού, και όπου p είναι ο συσχετισμός μεταξύ των αποκρίσεων για άτομα που ανήκουν στο ίδιο cluster. Αν όλα τα clusters έχουν το ίδιο μέγεθος n , ο παράγοντας παραπάνω παράγοντας γίνεται $1 + p(n-1)$. Αν το $p=1$, το σχέδιο επίδρασης γίνεται n , και η διακύμανση ολόκληρου του μέσου είναι ανάλογη του $1/K$ και όχι του $1/N$, όπως επισημάνθηκε στην προηγούμενη παράγραφο. Στο στάδιο σχεδιασμού, τα δειγματικά μεγέθη καθορίζονται από τις μεθόδους που περιγράφηκαν νωρίτερα πολλαπλασιασμένες με την επίδραση σχεδιασμού.

Μία απλή προσέγγιση της ανάλυση της cluster-randomized δοκιμής είναι να αντικαταστήσουμε τις αποκρίσεις σε κάθε cluster από τον μέσο τους και να χρησιμοποιήσουμε τους μέσους αυτούς σε μία κλασική two-sample sized ανάλυση, για παράδειγμα σε ένα two-sample t test ή σε ένα μη παραμετρικό Mann-Whitney test. Ένα θεωρητικό μειονέκτημα της προσέγγιση αυτής είναι ότι οι μέσοι των clusters θα έχουν

διαφορετικές διακυμάνσεις αν τα n_i διαφέρουν. Κάτι τέτοιο βέβαια δεν αποτελεί σοβαρό πρόβλημα, εκτός και αν τα n_i διαφέρουν υπερβολικά.

Μία πιο επίσημη προσέγγιση θα ήταν να παραστήσουμε τα δεδομένα με ένα μοντέλο τυχαίας επίδρασης, με διακύμανση σ_B^2 για την απόκλιση μεταξύ των clusters, και σ_W^2 εντός αυτών. Αυτές οι διακυμάνσεις υπολογίζονται από τα δεδομένα.

Ένας πιο διεξοδικός υπολογισμός των cluster-randomization δοκιμών δίνεται από τον Murray (1998).

ΚΕΦΑΛΑΙΟ 4

Στατιστικές μέθοδοι στην επιδημιολογία (statistical methods in epidemiology)

4.1 Εισαγωγή

Ο τομέας της ιατρικής επιστήμης που καλείται *επιδημιολογία* έχει καθοριστεί ως η *μελέτη των κατανομών και των προσδιορισμών των καταστάσεων που σχετίζονται με την υγεία ή των γεγονότων που συμβαίνουν σε συγκεκριμένους πληθυσμούς, και η εφαρμογή των μελετών αυτών με σκοπό να ελεγχθούν τα προβλήματα υγείας*. Χονδρικά, η επιδημιολογία ασχολείται με την κατανομή της ασθένειας στον πληθυσμό, ή μιας σωματικής συνθήκης, ή των παραγόντων που επηρεάζουν την κατανομή. Στο πεδίο της ανήκει και η μελέτη χρόνιων ασθενειών, καθώς επίσης και μεταδοτικών ασθενειών που ωθούν στην αύξηση της επιδημίας των κλασικών τύπων ασθενειών. Το άτομο εμπίπτει σε κάποιον βαθμό στην *κοινωνική ιατρική ή δημόσια υγεία*, και κάποιος από τους λόγους που καθιστούν αναγκαίο το να μελετηθεί σχετίζονται με κοινωνικά θέματα και θέματα διαχείρισης, όπως η πρόληψη και η χρήση υπηρεσιών υγείας.

Η επιδημιολογία ασχολείται κυρίως με χαρακτηριστικά των ομάδων παρά με τα άτομα, και αναπόφευκτα αυξάνει τα στατιστικά προβλήματα. Κάποια από αυτά εννοιολογικά είναι ίδια με τα στατιστικά προβλήματα που δημιουργούνται σε άλλα παρακλάδια της ιατρικής επιστήμης αλλά και σε μη ιατρικές επιστήμες, και μπορούν να προσεγγιστούν από μεθόδους αναλύσεων που περιγράφηκαν σε προηγούμενα κεφάλαια. Άλλα μεθοδολογικά προβλήματα στην επιδημιολογία, αν και στατιστικού ενδιαφέροντος, είναι αλληλένδετα με εικασίες μη στατιστικής φύσης, για τις οποίες όμως δεν θα γίνει καμία αναφορά στην συνέχεια. Η ερμηνεία των στατιστικών δεδομένων που αφορούν τη ζωή και την υγεία και τα οποία απαιτούν μία πολύ καλή γνώση των διαδικασιών διαχείρισης για την καταγραφή τέτοιων δεδομένων και για την ταξινόμηση των ασθενειών και των αιτιών του θανάτου, αποτελούν κλασικά παραδείγματα της επιδημιολογίας (Benjamin, 1968 ; World Health Organization, 1992). Όπως επίσης και η κατάλληλη χρήση και η σταδιακή ανάπτυξη της ιατρικής καταγραφής διαφόρων δεδομένων (Acheson, 1967). Υπάρχει επί πλέον μία αξιολογική συλλογή βιβλιογραφίας που ασχολείται με την μαθηματική θεωρία των επιδημιολογικών ασθενειών. Ο Bailey (1975) και ο Becker (1989) παρέχουν χρήσιμες συνοπτικές πληροφορίες σχετικά με αυτή την εργασία.

Τέλος στο κεφάλαιο αυτό θα δούμε εν συντομία κάποια βασικά προβλήματα που εμφανίζονται στην επιδημιολογική έρευνα και στις στατιστικές μεθόδους που έχουν αναπτυχθεί για την αντιμετώπισή τους.

4.2 Ο σχεδιασμός των ερευνών

Είναι χρήσιμο να μπορούμε να κάνουμε διάκριση μεταξύ των *περιγραφικών ερευνών*, οι οποίες σχεδιάστηκαν για να παρέχουν κάποια απλά χαρακτηριστικά μεταξύ των πληθυσμών, και των *αναλυτικών ερευνών*, που σχεδιάστηκαν να ερευνούν συσχετισμούς μεταξύ βασικών μεταβλητών. Παραδείγματα του πρώτου τύπου αποτελούν οι έρευνες που εκτιμούν την εξάπλωση κάποιων ασθενειών στον πληθυσμό ή την συχνότητα κατανομής ιατρικών συσκέψεων κατά την διάρκεια μιας συγκεκριμένης χρονικής περιόδου. Ένα παράδειγμα του δεύτερου τύπου θα μπορούσε να είναι μία μελέτη του συσχετισμού ανάμεσα στην χρήση ενός κλασικού φαρμάκου και της εμφάνισης μιας αρνητικής επίδρασης. Η διάκριση δεν είναι τόσο ξεκάθαρη όσο φαίνεται να δείχνει. Σε μία έρευνα πληθυσμού μπορεί να ενδιαφερόμαστε επίσης για συσχετισμούς όπως αυτός ανάμεσα στην εξάπλωση μιας ασθένειας και της ηλικίας του ατόμου. Αντιθέτως, σε μία έρευνα συσχετισμού θα μπορούσαμε να θεωρήσουμε τον υπό μελέτη συσχετισμό ως μια πληθυσμιακή παράμετρο και να ερευνήσουμε το πώς εμφανίζεται τέτοια διαφορά από πληθυσμό σε πληθυσμό. Παρόλα αυτά στις πληθυσμιακές μελέτες η κύρια έμφαση δίνεται στην πρόβλεψη αξιόπιστων εκτιμήσεων για τα χαρακτηριστικά ενός καλά καθορισμένου πληθυσμού. Σε αναλυτικές μελέτες (που θα δούμε και παρακάτω), ο καθορισμός ενός πληθυσμού είναι λιγότερο σημαντικός από την σχέση μεταξύ μεταβλητών.

Το στατιστικό ενδιαφέρον στις μελέτες του πληθυσμού εμφανίζεται αρχικά από τις μελέτες που γίνονται με δειγματοληπτικές μεθόδους. Κάποιες έρευνες φυσικά, έχουν πραγματοποιηθεί από συνολική απογραφή. Το βασικά πλεονεκτήματα της δειγματοληψίας έγκεινται στο χαμηλό της κόστος, στο ανθρώπινο δυναμικό και στον χρόνο. Επιπλέον οι δειγματοληψίες μικρότερης κλίμακας διερευνούν το πώς θα γίνουν περισσότερο αξιόπιστες παρατηρήσεις.

Συνήθως το πρώτο βήμα είναι να καθοριστεί το *δειγματικό πλαίσιο* (sampling frame) που είναι ουσιαστικά μια λίστα ή ένας τύπος ταυτοποίησης των ατόμων στον πληθυσμό που έχει συλλεχθεί. Για παράδειγμα, αν ο σκοπός είναι να συλλέξουμε για την μελέτη ενήλικους κατοίκους συγκεκριμένης χώρας τότε ένας χρήσιμος τρόπος θα ήταν να καθορίσουμε το δειγματικό πλαίσιο βάση των ατόμων που καταχωρήθηκαν στις εκλογικές λίστες της περιοχής. Αν η πρόθεση είναι να συλλέξουμε άτομα από μία μικρή περιοχή της χώρας, το δειγματικό πλαίσιο ίσως καθοριστεί από τον σχεδιασμό ενός χάρτη με κατάλληλες υποδιαιρέσεις. Την στιγμή που τα άτομα θα συλλεχθούν, το πρόβλημα μεταφέρεται στο ποιος αριθμός από τα άτομα αυτά θα εισαχθούν στο δείγμα.

Πώς επιλέγεται το δείγμα αυτό; Αν κάποια χαρακτηριστικά του πληθυσμού είναι γνωστά, ίσως για παράδειγμα από συμπεράσματα προηγούμενων ερευνών, τότε σε κάποιες περιπτώσεις το δείγμα αποτελείται από άτομα που έχουν επιλεγεί σκόπιμα ενώ τα χαρακτηριστικά των ατόμων του δείγματος μπορεί να συμφωνούν ακριβώς ή σχεδόν ακριβώς με αυτά του συνολικού πληθυσμού. Για παράδειγμα σε δειγματοληψία για έρευνα αγοράς κάποιοι οργανισμοί χρησιμοποίησαν αναλογίες δειγματοληψίας, δηλαδή μια μέθοδο δειγματοληψίας βάση της οποίας κάθε άτομο πριν επιλεγεί δίνει συνέντευξη ατόμου στη οποία ρωτάται για συγκεκριμένα χαρακτηριστικά (όπως ηλικία, φύλο, κοινωνικό στάτους). Έτσι οι αναλογίες που επιλέγονται στις ομάδες επιδιώκεται να ανταποκρίνονται όσο το δυνατόν περισσότερο στις αναλογίες του πληθυσμού. Το πρόβλημα με την μέθοδο αυτή είναι ότι μπορεί να εμφανίσει κάποιες σοβαρές διαφορές

όπου ο πρώτος αριθμός αντιστοιχεί στην κατακόρυφη θέση και ο δεύτερος στην οριζόντια θέση.

Τα συστηματικά δείγματα είναι συχνά λιγότερο ακριβή από τα τυχαία δείγματα. Είναι ιδιαίτερα επικίνδυνα γιατί ακόμα και αν η σειρά των δειγμάτων στο δειγματικό πλαίσιο επιβεβαιώσει μία θετική συσχέτιση ανάμεσα στις μονάδες των οποίων η απόσταση διαφέρει, αυτές θα εξισωθούν με το διάστημα του δείγματος. Στο προηγούμενο γεωγραφικό παράδειγμα, αν προσθέσουμε την πληροφορία ότι υπάρχει μία σειρά των πιο γνωστών κοιλάδων από την ανατολή έως την δύση που αφορούν τις πέντε ξεχωριστές δειγματικές μονάδες, οι επιλεγόμενες μονάδες θα ήταν όλες αρκετά γνωστές. Το δειγματοληπτικό σφάλμα της μονάδας θα ήταν επομένως ασυνήθιστα υψηλό.

Γενικά είναι δύσκολο να εκτιμήσουμε το δειγματοληπτικό σφάλμα από ένα απλό συστηματικό δείγμα εκτός κι αν θεωρήσουμε ότι τέτοιες συσχετίσεις δεν υπάρχουν και το δείγμα μπορεί επομένως να θεωρηθεί ως τυχαία επίδραση. Ένας χρήσιμος μηχανισμός είναι να επιλέξουμε ταυτόχρονα διαφορετικά συστηματικά δείγματα με διαφορετικές αρχικές επιλογές. Θα μπορούσαμε τότε να βγάλουμε συμπεράσματα ξεχωριστά για κάθε συστηματικό δείγμα από μια μέση τιμή ή κάποια άλλη σχετική στατιστική μέθοδο, και η δειγματική διακύμανση ολόκληρου του μέσου θα μπορούσε να επιτευχθεί λαμβάνοντας την διακύμανση κάθε μέσου χωριστά.

Στρωματοποιημένο δείγμα (stratified sampling)

Σ' αυτή την μέθοδο ο πληθυσμός χωρίζεται σε υποομάδες, ή στρώματα (*strata*), καθένα από τα οποία λαμβάνεται τυχαία ως δείγμα με γνωστό δειγματικό μέγεθος. Ας υποθέσουμε ότι επιθυμούμε να εκτιμήσουμε την μέση τιμή μιας καθορισμένης μεταβλητής στον πληθυσμό και αυτό το στρώμα να καθορίζεται με τέτοιο τρόπο έτσι ώστε ο μέσος να διαφέρει σημαντικά από το ένα στρώμα στο άλλο. Σε ένα απλό τυχαίο δείγμα, η κατανομή των παρατηρήσεων πάνω στα στρώματα θα ποικίλει από δείγμα σε δείγμα, και αυτή η έλλειψη ελέγχου θα συμβάλει στην μεταβλητότητα του δειγματικού μέσου. Αντίθετα, σε επαναλαμβανόμενες δειγματοληψίες με ίδια κατανομή πάνω στα στρώματα, το στοιχείο της μεταβλητότητας δεν θα συνδέεται με αυτά. Τα στρώματα τότε ίσως καθορίζονται ποιοτικά. Για παράδειγμα, σε μία διεθνή έρευνα υγείας διαφορετικών περιοχών μιας χώρας ίσως χρειαστεί να πάρουμε στρώματα. Και αυτά ίσως καθορίζονται από μία ή περισσότερες ποσοτικές μεταβλητές – για παράδειγμα ομάδες κατά ηλικία. Το κλάσμα του στρώματος που λαμβάνεται ως δείγμα μπορεί να είναι σταθερό ή μπορεί να ποικίλει από στρώμα σε στρώμα. Τα σημαντικά σημεία είναι : (i) ότι η κατανομή τόσο του πληθυσμού όσο και του δείγματος πάνω στα στρώματα πρέπει να είναι γνωστή, και (ii) ότι η μεταβλητότητα μεταξύ των στρωμάτων θα πρέπει να είναι όσο το δυνατόν πιο μεγάλη, ή ισοδύναμα ότι κάθε στρώμα θα πρέπει να είναι όσο πιο ομοιογενές γίνεται.

Υποθέτουμε ότι υπάρχουν k στρώματα, ότι στο i^{th} στρώμα το μέγεθος του πληθυσμού είναι N_i , και ότι ο μέσος και η διακύμανση της μεταβλητής X είναι μ_i και σ_i^2 αντίστοιχα. Παίρνουμε ένα τυχαίο δείγμα μεγέθους n_i από το i^{th} στρώμα. Το δειγματικό κλάσμα n_i / N_i θα συμβολίζεται με f_i . Ο δειγματικός μέσος και η εκτίμηση

της διακύμανσης στο i^{th} στρώμα είναι \bar{x}_i και s_i^2 αντίστοιχα. Το συνολικό πληθυσμιακό μέγεθος είναι

$$N = \sum_{i=1}^k N_i$$

το συνολικό δειγματικό μέγεθος είναι

$$n = \sum_{i=1}^k n_i$$

Η μέση τιμή του x στον συνολικό πληθυσμό δίνεται ξεκάθαρα από τον τύπο

$$\mu = \frac{\sum N_i \mu_i}{N}$$

Η κατάλληλη εκτίμηση του μ είναι

$$\mu = \frac{\sum N_i \bar{x}_i}{N} \quad (4.1)$$

η οποία εύκολα διαπιστώνεται ότι είναι αμερόληπτη εκτιμήτρια. Η διακύμανση της είναι

$$\begin{aligned} \text{var}(\mu) &= \frac{1}{N^2} \sum N_i^2 \text{var}(\bar{x}_i) \\ &= \frac{1}{N^2} \sum \frac{N_i^2 \sigma_i^2}{n_i} (1 - f_i) \end{aligned} \quad (4.2)$$

Κάτι που εκτιμάται από το δείγμα ως

$$\text{var}(\mu) = \frac{1}{N^2} \sum \frac{N_i^2 s_i^2}{n_i} (1 - f_i) \quad (4.3)$$

Αν όλα τα f_i είναι μικρά, οι όροι $1 - f_i$ μπορούν να αντικατασταθούν από την μονάδα. Γενικά το (4.2) θα είναι μικρότερο από την διακύμανση του μέσου ενός τυχαίου δείγματος γιατί ο όρος σ_i^2 μετράει μόνο την μεταβλητότητα εντός του στρώματος. Η μεταβλητότητα μεταξύ των μ_i είναι ασυσχέτιστη.

Αν καθορισθεί το n και τα f_i είναι μικρά, η $\text{var}(\mu)$ ελαχιστοποιείται εφόσον τα n_i επιλεγούν όσο το δυνατόν πιο ανάλογα με το $N_i \sigma_i$. Όμως αν το σ_i είναι σταθερό από το ένα στρώμα στο άλλο, τότε το n_i θα επιλέγεται έτσι ώστε να είναι ανάλογο του N_i , δηλαδή το δειγματικό κλάσμα θα είναι σταθερό. Το στρώμα με σχετικά μεγάλο σ_i έχει αντίστοιχα αυξανόμενα δειγματοληπτικά κλάσματα. Συνήθως οι μικρές τιμές για το σ_i θα είναι γνωστές πριν διεξαχθεί η έρευνα και η επιλογή ενός σταθερού δειγματικού σφάλματος θα είναι η πιο λογική στρατηγική που θα ακολουθήσουμε στην περίπτωση αυτή.

Αν από την αρχή της έρευνας καθορίστηκε ότι θα πρέπει να εκτιμάται ένα μέρος του π , στον παραπάνω τύπο θα αντικατασταθούν τα \bar{x}_i και μ_i από τα p_i και π_i , που είναι οι παρατηρούμενες και οι πραγματικές αναλογίες στο i^{th} στρώμα, και το σ_i^2 θα αντικατασταθεί από το $\pi_i(1-\pi_i)$ ενώ το s_i^2 από το $n_i p_i(1-p_i)/(n_i-1)$.

Εκτός από την αυξανόμενη πρόβλεψη της εκτιμήτριας του πληθυσμιακού μέσου, η στρωματοποίηση ίσως υιοθετηθεί για να παρέχει ακριβείς εκτιμήσεις του μέσου για καθένα από τα στρώματα. Αυτό οδηγεί στο να μην γίνει η χρήση της βέλτιστης κατανομής που περιγράψαμε παραπάνω, έτσι ώστε κανένα από τα δειγματικά μεγέθη σε καθένα από τα στρώματα να μην γίνει πολύ μικρό.

Παράδειγμα 1:

Είναι επιθυμητό να εκτιμήσουμε την επικράτηση μιας καθορισμένης συνθήκης (για παράδειγμα την αναλογία των ατόμων που έχουν επηρεαστεί) σε έναν πληθυσμό 5000 ατόμων, παίρνοντας ένα δείγμα των 100 ατόμων. Υποθέτουμε ότι είναι γνωστό ότι η επικράτηση συνδέεται με την ηλικία και ότι ο πληθυσμός μπορεί να καταταχθεί σε τρία στρώματα καθορισμένα από την ηλικία, με το ακόλουθο πλήθος σε καθένα από αυτά:

Stratum, i	Age (years)	N_i
1	0-14	1200
2	15-44	2200
3	45-	1600
		5000

Υποθέτουμε ότι η πραγματική επικράτηση, π_i , στα διαφορετικά στρώματα είναι η ακόλουθη:

Stratum, i	π_i
1	0.02
2	0.08
3	0.15

Εύκολα εξακριβώνεται ότι ολόκληρη η επικράτηση, $\pi (= \sum N_i \pi_i / \sum N_i)$, ισούται με 0.088. Ένα απλό τυχαίο δείγμα μεγέθους $n=100$ θα μπορούσε να δώσει μία εκτίμηση p με διακύμανση $(0.088)(0.912)/100=0.000803$.

Για στρωματοποιημένο δείγμα με βέλτιστη κατανομή τα δειγματικά μεγέθη στα στρώματα θα πρέπει να διαλέγονται σε αναλογία με το $N_i \sqrt{[\pi_i(1-\pi_i)]}$. Οι τιμές αυτής της ποσότητας στα τρία στρώματα είναι 168, 597 και 571, που είναι ανάλογες με τις τιμές 0.126, 0.447, και 0.427, αντίστοιχα. Τα βέλτιστα δειγματικά μεγέθη είναι, $n_1=12$, $n_2=45$ και $n_3=43$, ή αριθμοί πολύ κοντά σε αυτούς. Ας καλέσουμε την κατανομή αυτή Α. Τα δειγματικά μεγέθη εξαρτώνται από το άγνωστο π_i , και συνεπώς στην πραγματικότητα δεν μπορούν να χρησιμοποιηθούν εύκολα. Αν γνωρίζαμε έστω και πολύ λίγα σχετικά με την πιθανή διακύμανση π_i , ίσως επιλέγαμε το n_i , αγνοώντας την επίδραση της σταθερά μεταβαλλόμενης απόκλισης. Αυτό για την κατανομή Β θα έδινε, $n_1=24$, $n_2=44$ και $n_3=32$. Τρίτον, αν έχουμε κάποια ιδέα ότι η επικράτηση (και συνεπώς και η απόκλιση) αυξάνεται με την ηλικία, τότε η κατανομή τροποποιείται σχετικά αυθαίρετα στο να δώσει, ας πούμε, $n_1=20$, $n_2=40$ και $n_3=40$ (κατανομή C). Η εκτίμηση του π , σε κάθε περίπτωση, δίνεται από τύπο ισοδύναμο του (4.1).

$$\pi = \frac{\sum N_i p_i}{N}$$

Όπου το p_i είναι η εκτιμώμενη επικράτηση του στρώματος i . Η διακύμανση του π δίνεται από την ισοδυναμία του τύπου (4.2), στον οποίο θα αναφέρεται ο όρος f_i όντας μικρός:

$$\text{var}(\pi) = \frac{1}{N^2} \sum \frac{N_i^2 \pi_i (1-\pi_i)}{n_i}$$

Οι τιμές του $\text{var}(\pi)$ για τις τρεις κατανομές είναι ως εξής:

Κατανομή	$\text{var}(\pi)$
A	0.000714
B	0.000779
C	0.000739

Όπως αναμενόταν, η μικρότερη διακύμανση είναι για το A και η υψηλότερη για το B, η τελευταία εκ των οποίων είναι λίγο μικρότερη από την διακύμανση του τυχαίου δείγματος.

Πολυεπίπεδο δείγμα (multistage sampling)

Στην μέθοδο αυτή το δειγματικό πλαίσιο κατανέμεται σε έναν πληθυσμό δειγματικών μονάδων πρώτου επιπέδου (first-stage sampling units), από το οποίο λαμβάνουμε ένα δείγμα πρώτου επιπέδου. Αυτό είναι συνήθως ένα τυχαίο δείγμα ή ένα συστηματικό ή στρωματοποιημένο δείγμα. Όπως θα δούμε, μπορεί να είναι ένα τυχαίο δείγμα στο οποίο κάποιες μονάδες πρώτου επιπέδου επιτρέπεται να έχουν μεγαλύτερη πιθανότητα να επιλεγούν σε σχέση με άλλες. Καθεμία από τις δειγματικές μονάδες πρώτου επιπέδου υποδιαιρείται σε δειγματικές μονάδες δεύτερου επιπέδου. Η διαδικασία μπορεί να συνεχιστεί σε τόσα επίπεδα όσα κριθούν απαραίτητα.

Υπάρχουν δύο βασικά πλεονεκτήματα του πολυεπίπεδου δείγματος. Πρώτα από όλα παρέχει την δυνατότητα μείωσης του κόστους εφόσον χρησιμοποιούμε αναλογίες του δειγματικού πληθυσμού και όχι ολόκληρο το δείγμα. Δεύτερον, είναι πιο πρακτικό για καταστάσεις στις οποίες δεν είναι διαθέσιμο όλο το δείγμα προτού ξεκινήσει η έρευνα. Απαιτείται μια λίστα από μονάδες πρώτου επιπέδου, αφού οι μονάδες δεύτερου επιπέδου επιλέγονται και ταξινομούνται βάση του πρώτου επιπέδου που επιλέγεται στο δείγμα.

Ένα χαρακτηριστικό παράδειγμα είναι η έρευνα της υγείας των ανδρών που εργάζονται σ' ένα εργοστάσιο. Ίσως δεν υπάρχει ένας ολοκληρωμένος κατάλογος των ανδρών στο εργοστάσιο, αλλά είναι εύκολο να επιτύχουμε να φτιάξουμε μια λίστα των εργοστασίων, που θα μπορούσαν να αποτελέσουν τις μονάδες πρώτου επιπέδου. Από κάθε εργοστάσιο από αυτά που επιλέχθηκαν στο δείγμα πρώτου επιπέδου, θα μπορούσε να επιτευχθεί η καταγραφή των ανδρών που εργάζονται σε αυτά και θα μπορούσε από τη λίστα αυτή να επιλεγεί ένα δείγμα δεύτερου επιπέδου. Εκτός από τα πλεονεκτήματα που έχουμε από την καταγραφή των ανδρών που ανήκουν μόνο στα εργοστάσια που επιλέχθηκαν στο πρώτο επίπεδο, η διαδικασία αυτή θα μπορούσε να οδηγήσει σε σημαντική οικονομία κόστους παρέχοντας την δυνατότητα στην έρευνα να συγκεντρωθεί μόνο στα επιλεγμένα εργοστάσια αντί να εξεταστεί ένα απλό δείγμα ανδρών σε όλα τα διαφορετικά μέρη της χώρας.

Η οικονομία στο κόστος και των πόρων που απαιτούνται για την μελέτη συνοδεύεται δυστυχώς από απώλεια ακρίβειας στην πρόβλεψη σε σύγκριση με την μελέτη που βασίζεται σε ένα απλό τυχαίο δείγμα. Υποθέτουμε τώρα ότι για το παράδειγμα που αναφέραμε επιλέγουμε ένα δείγμα 20 εργοστασίων και κάποια δείγματα δεύτερου επιπέδου 50 ανδρών από καθένα από τα 20 εργοστάσια. Αν υπάρχει μία συστηματική διακύμανση μεταξύ των εργοστασίων, ίσως εξαιτίας των διαφορετικών συνθηκών υγείας στα διαφορετικά μέρη της χώρας ή εξαιτίας των διαφορετικών εργασιακών κινδύνων, αυτή η διακύμανση θα μπορούσε να προκύπτει από ένα δείγμα μόνο 20 μονάδων πρώτου επιπέδου. Ένα τυχαίο δείγμα 1000 ανδρών, από την άλλη, θα μπορούσε να εκπροσωπεί 1000 τυχαίες επιλογές μονάδων πρώτου επιπέδου.

Ένα χρήσιμο εργαλείο στο οποίο βασιζόμαστε για την επιλογή δείγματος δεύτερου επιπέδου είναι το λεγόμενο *self-weighting* (αυτορυθμιζόμενο). Σύμφωνα με αυτό κάθε μονάδα πρώτου επιπέδου δίνεται από μία πιθανότητα επιλογής που είναι ανάλογη με τις μονάδες που περιέχονται στο δεύτερο στάδιο. Τα δείγματα δεύτερου επιπέδου επιλέγονται τότε έτσι ώστε να έχουν το ίδιο μέγεθος. Και ακολούθως κάθε μονάδα δεύτερου επιπέδου σε ολόκληρο τον πληθυσμό έχει μία ισοδύναμη ευκαιρία να επιλεγεί και επομένως οι τύποι που χρειάζονται είναι απλοποιημένοι.

Ομαδικό δείγμα (cluster sampling)

Μερικές φορές, στο τελικό στάδιο ενός πολυεπίπεδου δείγματος έχει γίνει και η ολοκληρωμένη καταμέτρηση των διαθέσιμων μονάδων. Στο παράδειγμα μας συγκεκριμένα, όταν τοποθετηθεί στο εργοστάσιο κάποια ομάδα έρευνας με σκοπό να εξεταστούν όλοι οι άντρες του εργοστασίου τότε αυξάνεται το κόστος ελέγχου των ανδρών του εργοστασίου. Από την άλλη όμως ίσως είναι πραγματικά χρήσιμο ώστε να αποφύγουμε την αδυναμία του να προσκαλέσουμε κάποιους άντρες να συμμετάσχουν και όχι κάποιους άλλους. Όταν δεν υπάρχει δείγμα στο τελευταίο στάδιο, η μέθοδος αναφέρεται ως «cluster sampling». Ο ερευνητής δεν έχει έλεγχο πάνω στο πλήθος των δειγματικών μονάδων στις ομάδες και αυτό σημαίνει, σε σύγκριση με ένα τυχαίο δείγμα, ότι η απώλεια ακριβείας είναι πολύ μεγαλύτερη από ένα πολυεπίπεδο δείγμα.

Επίδραση σχεδιασμού (design effect)

Ο λόγος της διακύμανσης ενός εκτιμητή από ένα δειγματικό σχεδιασμό προς την διακύμανση ενός εκτιμητή από ένα απλό τυχαίο δείγμα με τον ίδιο συνολικό αριθμό δειγματικών μονάδων είναι γνωστή ως *επίδραση σχεδιασμού*, συχνά αναφερόμενη σύντομα ως *Deff*. Στο παράδειγμα (1) η *Deff* για την κατανομή C είναι $0.000739/0.000803=0.92$. Ένας άλλος τρόπος να υπολογίσουμε την *Deff* είναι λαμβάνοντας υπόψη το δειγματικό μέγεθος. Η ίδια ακρίβεια θα μπορούσε να επιτευχθεί και με στρωματοποιημένο δείγμα 92 ανθρώπων με κατανομή C αντί για ενός απλού τυχαίου δείγματος 100 ατόμων. Το στρωματοποιημένο δείγμα είναι περισσότερο αποτελεσματικό ($Deff < 1$) από ένα απλό τυχαίο δείγμα και αυτό μπορεί γενικά να ισχύει υπό την προϋπόθεση ότι υπάρχει μία μικρή διακύμανση μεταξύ των στρωμάτων. Η αποτελεσματικότητα του στρωματοποιημένου δείγματος αυξάνεται με την αύξηση της ετερογένειας ανάμεσα στα στρώματα και την μεγαλύτερη ομοιογένεια εντός των στρωμάτων προκύπτει εξαιτίας της αύξησης των παραπάνω.

Αντίθετα, ένα πολυεπίπεδο και ένα ομαδικό δείγμα θα έχει συνήθως $Deff > 1$. Αυτό σημαίνει ότι, θα απαιτείται ένα μεγαλύτερο δείγμα από αυτό του απλού τυχαίου δείγματος. Για το ομαδικό δείγμα με m μέλη σε κάθε ομάδα και συσχέτιση της τάξης του ρ για την υπό μελέτη διακύμανση εντός των ομάδων, η *Deff* δίνεται από τον τύπο

$$Deff=1+(m-1)\rho$$

ο οποίος πάντα υπερβαίνει το 1 (εκτός από το απίθανο σενάριο της αρνητικής συσχέτισης εντός των ομάδων). Αν το m διαφέρει μεταξύ των ομάδων, τότε, στον παραπάνω τύπο, το m αντικαθιστάται από το $\sum m^2 / \sum m$, το οποίο προσεγγιστικά είναι \bar{m} , υπό την προϋπόθεση ότι ο συντελεστής διακύμανσης του μεγέθους της ομάδας είναι μικρός.

Η παραπάνω συζήτηση που γίνεται για την αποτελεσματικότητα διαφορετικών δειγματικών σχεδιασμών σε σύγκριση με το απλό τυχαίο δείγμα λαμβάνει υπόψη το δειγματικό μέγεθος. Όπως παρατηρήσαμε στην υποενότητα για το πολυεπίπεδο δείγμα, ένα από τα πλεονεκτήματα της μεθόδου αυτής είναι η μείωση του κόστους. Και η συνολική αποτελεσματικότητα ενός δειγματικού σχεδιασμού θα έπρεπε να αξιολογείται λαμβάνοντας υπόψη το κόστος της διαδικασίας μέχρι να ολοκληρωθεί το δείγμα, και όχι μόνο από το τελικό μέγεθος του δείγματος. Αν το κόστος του δειγματικού σχεδιασμού είναι c για κάθε μέλος του δείγματος, αναλογικά με το κόστος κάθε μέλους ενός απλού τυχαίου δείγματος, η συνολική αποτελεσματικότητα του δειγματικού σχεδιασμού είναι

$$\frac{100\%}{c \times Deff}$$

Συνεπώς, ένας δειγματικός σχεδιασμός είναι λίγο έως πολύ, πιο αποτελεσματικός από ένα τυχαίο δείγμα θεωρώντας ότι το γινόμενο $c \times Deff$ είναι μικρότερο, ή μεγαλύτερο, από την μονάδα, αντίστοιχα.

Σημαδεμένα-Ξανασυλληφθέντα δείγματα (mark-recapture sampling)

Το όνομα της τεχνικής αυτής προέρχεται λόγω της εφαρμογής της σε ζώα ή πουλιά μιας περιοχής με σκοπό να εκτιμηθεί το πλήθος τους. Πιο συγκεκριμένα, τοποθετούνται παγίδες και τα ζώα που αιχμαλωτίζονται συνθέτουν ένα δείγμα του πληθυσμού. Στην συνέχεια σημαδεύονται και απελευθερώνονται. Ένα δεύτερο σύνολο παγίδων που τοποθετείται, παράγει ένα δεύτερο δείγμα αποτελούμενο από κάποια μη μαρκαρισμένα ζώα και κάποια μαρκαρισμένα σε περίπτωση που αυτά είχαν αιχμαλωτιστεί από το πρώτο σύνολο παγίδων. Πάλι, τα ζώα σημαδεύονται και απελευθερώνονται και η διαδικασία συνεχίζεται για με διάφορα σύνολα παγίδων.

Το πλήθος των ζώων που αιχμαλωτίζονται σε κάθε περίπτωση, σε συνδυασμό με το πλήθος των ζώων που πιάστηκαν και σημαδεύτηκαν από προηγούμενη αιχμαλωσία, μπορεί να χρησιμοποιηθεί για να εκτιμήσει τον συνολικό πληθυσμό. Οι υποθέσεις σχετίζονται με το εάν ο πληθυσμός είναι κλειστός ή όχι (χωρίς εισχωρήσεις και απώλειες κατά την δειγματοληπτική περίοδο), όπως και με το εάν η πιθανότητα αιχμαλωσίας εξαρτάται από προηγούμενο ιστορικό αιχμαλωσίας, και με το τι ετερογένεια υπάρχει μεταξύ των ζώων που σχετίζεται με την πιθανότητα της αιχμαλωσίας. Όσο μεγαλύτερο όγκο πληροφοριών έχουμε στην διάθεση μας - για παράδειγμα, περισσότερες περιπτώσεις τοποθέτησης παγίδων και ένα σύστημα μαρκαρίσματος που επιτρέπει την ταυτοποίηση στην περίπτωση που κάποιο ζώο είχε αιχμαλωτιστεί ξανά- τόσο πιο εύκολα μπορούν οι υποθέσεις να ελεγχθούν και να τροποποιηθούν.

Η τεχνική αυτή εφαρμόζεται με σκοπό να εκτιμηθεί το μέγεθος ενός πληθυσμού ατόμων με κάποια κοινά χαρακτηριστικά υγείας. Είναι ιδιαίτερα χρήσιμο για σπάνια χαρακτηριστικά, τα οποία θα απαιτούσαν ένα πολύ μεγάλο δείγμα στην περίπτωση που θα χρησιμοποιούσαμε περισσότερο παραδοσιακές μεθόδους. Επιπλέον είναι κατάλληλο για μελέτη συνηθειών που οι άνθρωποι ίσως είναι διστακτικοί να αποκαλύψουν σε ένα ερωτηματολόγιο, όπως η ενδοφλέβια χρήση ναρκωτικών, ή για ομάδες οι οποίες ίσως παραλείπονται λόγω διαφορετικότητας από το δειγματικό πλαίσιο, όπως για παράδειγμα οι άστεγοι σε μία πόλη. Οι λίστες που περιέχουν άτομα από μία τέτοια ομάδα πιθανόν να είναι διαθέσιμες από οργανισμούς που παρέχουν υπηρεσίες για την ομάδα αυτή, και αν διάφορες τέτοιες λίστες είναι διαθέσιμες, τότε καθεμία ίσως μπορεί να θεωρηθεί ότι ανταποκρίνεται και σε μία περίπτωση “παγίδευσης”. Αν υπάρχει επαρκής ταυτοποιημένη πληροφορία, τότε αυτή είναι ικανή να καθορίσει αν ένα άτομο είναι σε περισσότερες από μία λίστες (αιχμαλωσίες). Η διαφορά από τις καταστάσεις αιχμαλωσίας των ζώων είναι ότι δεν υπάρχει χρονική ακολουθία με αποτέλεσμα οι λίστες να αντιμετωπίζονται συμμετρικά ή μία με την άλλη.

Αν υπάρχουν k λίστες, τότε οι παρατηρήσεις μπορούν να ταξινομηθούν σε έναν πίνακα 2^k , δηλώνοντας την παρουσία ή την απουσία κάθε λίστας. Ο αριθμός που αντιστοιχεί στην απουσία στις k λίστες στο κελί φυσικά και δεν είναι γνωστός, και ο σκοπός της μεθόδου είναι να εκτιμήσει τον αριθμό αυτό και συνεπώς και τον συνολικό αριθμό απουσιών. Θα ήταν δικαιολογημένο να θεωρήσουμε ότι η πιθανότητα του να είναι ένας άνθρωπος σε μία τέτοια λίστα εξαρτάται από το εάν ή όχι ένα άτομο είναι και σε κάποια από τις άλλες λίστες. Αυτό σημαίνει ότι, θα υπάρχει μία λίστα εξάρτησης (ανταποκρινόμενη στην αιχμαλωσία που εξαρτάται από την ιστορία προηγούμενης αιχμαλωσίας). Μία μέθοδος να προχωρήσουμε είναι να εφαρμόσουμε ένα *log-γραμμικό μοντέλο* σε $2^k - 1$ κελιά του 2^k πίνακα και να χρησιμοποιήσουμε αυτό το μοντέλο για να εκτιμήσουμε το πλήθος στο μη παρατηρούμενο κελί. Οι εξαρτήσεις μεταξύ των κελιών μπορούν να συμπεριληφθούν σαν όροι αλληλεπίδρασης αλλά η υψηλότερη τάξη αλληλεπίδρασης ανάμεσα στις k λίστες έχει οριστεί να είναι το μηδέν, αφού δεν μπορεί να εκτιμηθεί από μη ολοκληρωμένο πίνακα.

Αναλυτικά η παραπάνω μεθοδολογία βασισμένη στην χρήση του *log-γραμμικού μοντέλου* δόθηκε από τον Seber (1982), και από τον Cormack (1989). Μία σύντομη περιγραφή δόθηκε επίσης από τον Chao (1998). Η μέθοδος δίνει μία εκτίμηση του συνολικού πληθυσμιακού μεγέθους και του σταθερού σφάλματος. Ένα πρόβλημα είναι ότι η εκτίμηση βασίζεται σε μη επαληθεύσιμες εικασίες, με συνέπεια το σταθερό σφάλμα να μην περιγράφει επαρκώς την αβεβαιότητα. Παρόλα αυτά, σε καταστάσεις που δεν μπορούν να εφαρμοστούν πιο ισχυρές μέθοδοι που βασίζονται σε τυχαίο δείγμα, η μέθοδος αυτή παρέχει κάποιες ικανοποιητικές εκτιμήσεις.

Απόδοση (imputation)

Στα υπό έρευνα δείγματα συνήθως υπάρχει απώλεια δεδομένων καθώς κάποιοι άνθρωποι αποτυγχάνουν να απαντήσουν σε κάποιες ερωτήσεις, είτε ακούσια είτε γιατί αρνούνται να απαντήσουν σε συγκεκριμένες ερωτήσεις. Η ανάλυση ολόκληρου του συνόλου των δεδομένων διευκολύνεται αν τα χαμένα δεδομένα αντικατασταθούν από τιμές απόδοσης (imputed values). Οι τιμές απόδοσης επιτυγχάνονται από την χρήση ενός μοντέλου

βασισμένου στις παρατηρούμενες τιμές. Για παράδειγμα μια πολλαπλή παλινδρόμηση του x_1 στα x_2, x_3, \dots, x_p εφαρμόζεται σε όλα τα δεδομένα και χρησιμοποιείται για να εκτιμήσει τις τιμές του x_1 για τα άτομα αυτά, με την συγκεκριμένη διακύμανση να παραλείπεται αλλά λαμβάνοντας υπόψη τις παρατηρούμενες τιμές των x_2, x_3, \dots, x_p . Αυτό δίνει τις καλύτερες προβλεπόμενες τιμές αλλά το να αντικατασταθούν όλες αυτές οι τιμές και το να κάνουμε τόσες εξαιρέσεις στην διακύμανση έχει κόστος στην ακρίβεια. Αντί για αυτό λοιπόν, θα μπορούσαμε χτίσουμε την τυχαία μεταβλητότητα βάση του συνόλου των τιμών απόδοσης, όχι μόνο από την μεταβλητότητα της κατάλληλης παλινδρόμησης αλλά επιπλέον κι από την μεταβλητότητα λόγω της αβεβαιότητας κατά την εκτίμηση των συντελεστών παλινδρόμησης. Κατά συνέπεια οι τιμές απόδοσης περιέχουν τυχαία στοιχεία και δεν υπάρχουν μοναδικά καλύτερα σύνολα τιμών απόδοσης. Ακόμα και σε ένα σύνολο δεδομένων με τυχαία διακύμανση, μια απλή ανάλυση θα δώσει πολύ μικρό σταθερό σφάλμα των εκτιμώμενων παραμέτρων. Κάτι τέτοιο μπορεί να αποφευχθεί από την πολλαπλή απόδοση. Αυτό σημαίνει ότι τα απολεσθέντα δεδομένα αναλύονται με τον ίδιο τρόπο. Οι διακυμάνσεις των εκτιμώμενων παραμέτρων μπορούν τότε να παραχθούν από τον συνδυασμό της εντός-διακύμανσης απόδοσης με την μεταξύ τους διακύμανση απόδοσης.

Η πολλαπλή απόδοση χρησιμοποιείται γενικά σε μελέτη μεγάλων δειγμάτων αλλά εφαρμόζεται ευρέως και όταν υπάρχει κατάλληλο διαθέσιμο λογισμικό. Τεχνικές σχετικές με την μέθοδο αυτή αναφέρονται από τον Rubin (1987) και από τον Barnard (1998).

Ο Barnard (1998) περιγράφει επίσης εφαρμογές της πολλαπλής απόδοσης σε ένα ευρύτερο φάσμα από αυτό της μη-απόκρισης. Αυτά εσωκλείουν την απόκριση της πραγματικής ηλικίας όταν τα δεδομένα που επιλέχθηκαν με βάση την ηλικία δεν επαρκούν ώστε να έχουμε ακριβή συμπεράσματα, ή την απόδοση των ημερομηνιών διάγνωσης του συνδρόμου ανοσοποιητικής ανεπάρκειας (AIDS) σε άτομα με τον ιό HIV που δεν είχαν ακόμη εμφανίσει AIDS, χρησιμοποιώντας ένα μοντέλο βασισμένο στο σύνολο των συνδιακυμάνσεων.

4.3 Δείκτες και προτυποποιήσεις

Από τότε που η επιδημιολογία ασχολήθηκε με την κατανομή μιας ασθένειας στους πληθυσμούς, απαιτήθηκαν σύντομες μετρήσεις που περιγράφουν την ποσότητα του πληθυσμού που νοσεί από τη ασθένεια αυτή. Υπάρχουν δύο βασικά μέτρα, το συμβάν και η εξάπλωση.

Το συμβάν είναι μια μέτρηση του δείκτη σύμφωνα με τον οποίο έχουμε καινούργια περιστατικά ασθενούντων που προκύπτουν σε έναν υγιή πληθυσμό (πληθυσμός που δεν είχε νοσήσει μέχρι τώρα). Συνεπώς, το συμβάν, το οποίο υποδηλώνεται από το I , καθορίζεται από τον τύπο

$$I = \frac{\text{πλήθος νέων περιπτώσεων την συγκεκριμένη χρονική περίοδο}}{\text{πληθυσμός σε κίνδυνο}}$$

Η χρονική αυτή περίοδος έχει τις ίδιες μονάδες με αυτές που έχει και ο δείκτης. Συχνά ο δείκτης πολλαπλασιάζεται με 1000 ή 1000000 ώστε να αποφύγουμε μικρά δεκαδικά κλάσματα. Για παράδειγμα το 1997 υπήρχαν 280 νέες περιπτώσεις καρκίνου του πάγκρεας σε άντρες της Νέας Νότιας Ουαλίας σε πληθυσμό 3.115 εκατομμυρίων αντρών. Το συμβάν δηλαδή ήταν $280/3.115=90$ ανά εκατομμύριο ανά έτος.

Η εξάπλωση, που υποδηλώνεται από το P , είναι ένα μέτρο της συχνότητας της ύπαρξης της ασθένειας σε μία δοσμένη χρονική στιγμή, και καθορίζεται από τον τύπο

$$I = \frac{\text{συνολικός αριθμός περιπτώσεων σε δοσμένη χρονική στιγμή}}{\text{συνολικός πληθυσμός την στιγμή αυτή}}$$

Τόσο το συμβάν όσο και η εξάπλωση εξαρτώνται συνήθως από την ηλικία, και πιθανόν από το φύλο, και από φυλετικά και ηλικιακά χαρακτηριστικά.

Οι δείκτες συμβάντος και εξάπλωσης θα σχετίζονται, αφού μία περίπτωση συμβάντος είναι απευθείας περιστατικό μιας περίπτωσης εξάπλωσης και παραμένει ένα τέτοιο μέχρι τη θεραπεία ή τον θάνατο. Δοσμένης μιας σταθερής κατάστασης λοιπόν, ο σύνδεσμος μεταξύ των δύο μέτρων δίνεται από τον τύπο

$$P=It \tag{4.4}$$

Όπου το t είναι η μέση διάρκεια της ασθένειας. Για μία χρόνια ασθένεια για την οποία δεν υπάρχει θεραπεία, το t θα μπορούσε να είναι η μέση επιβίωση μετά το περιστατικό της ασθένειας.

Προτυποποίηση

Στην vital στατιστική εμφανίζονται συχνά προβλήματα εξαιτίας ανατροπών τα οποία δίνουν ώθηση στην δημιουργία μιας ομάδας μεθόδων που καλείται *προτυποποίηση*. Θα περιγράψουμε πολύ σύντομα μία ή δύο από τις πολύ βασικές μεθόδους.

Η θνησιμότητα σε έναν πληθυσμό συνήθως υπολογίζεται από έναν ετήσιο δείκτη θανάτου-για παράδειγμα το πλήθος των ατόμων που πεθαίνουν κατά την διάρκεια ενός ημερολογιακού έτους, διαιρείται από το εκτιμώμενο μέγεθος πληθυσμού που προκύπτει στα μέσα της χρονιάς. Συχνά ο δείκτης αυτός πολλαπλασιάζεται με μία βολική δύναμη του 10, όπως για παράδειγμα το 1000, ώστε να αποφύγουμε πολύ μικρά δεκαδικά κλάσματα. Τότε ο δείκτης καλείται *ετήσιος δείκτης ανά 1000 πληθυσμό*. Αν ο δείκτης θνησιμότητας που υπολογίστηκε για έναν πληθυσμό καλύπτει ένα ευρύ ηλικιακό φάσμα, τότε καλείται *χονδρικός δείκτης θνησιμότητας (crude death rate)*.

Σε μία σύγκριση της θνησιμότητας ανάμεσα σε δύο πληθυσμούς, ας πούμε, εκείνων που προέρχονται από δύο διαφορετικές χώρες, ο crude δείκτης θνησιμότητας μπορεί να είναι παραπλανητικός. Η θνησιμότητα βασίζεται ισχυρά στην ηλικία. Αν οι δύο χώρες έχουν κατοίκους διαφορετικού ηλικιακού φάσματος, αυτή η αντίθεση από

μόνη της ίσως εξηγεί μία διαφορά στον δείκτη crude. Ένα παράδειγμα που δίνεται στον Πίνακα 4.1 (που δίνεται παρακάτω) ,δείχνει το πλήθος των ατόμων και το πλήθος των θανάτων χωριστά στις διαφορετικές ηλικιακές ομάδες , για τις δύο χώρες. Με **A** υποδηλώνουμε τις αναπτυγμένες βιομηχανικά χώρες , με μία αρκετά μεγάλη αναλογία ηλικιωμένων ατόμων, και με **B** υποδηλώνουμε , μία αναπτυσσόμενη χώρα με μικρό δείκτη ηλικιωμένων ατόμων. Ο δείκτης θνησιμότητας για κάθε ηλικία (που καλείται *ηλικιακά-συγκεκριμένος δείκτης θνησιμότητας, age-specific death rates*) είναι σημαντικά υψηλότερος για την **B** χώρα σε σχέση με την **A**, και επιπλέον ο crude δείκτης θνησιμότητας είναι μεγαλύτερος για την χώρα **A** σε σχέση με την **B**. Μερικές φορές ,παρόλα αυτά, η θνησιμότητα πρέπει να συγκρίνεται για ένα μεγάλο πλήθος διαφορετικών πληθυσμών, και απαιτούνται κάποιοι τροποποιημένοι τύποι για τις ηλικιακές διαφορές. Για παράδειγμα , η θνησιμότητα σε μια χώρα ίσως πρέπει να συγκριθεί για τα διαφορετικά έτη. Επιπλέον ίσως πρέπει να τεθούν υπό μελέτη διαφορετικές περιοχές της ίδιας χώρας. Ή κάποιος ίσως επιθυμεί να συγκρίνει έναν μεγάλο αριθμό διαφορετικών επαγγελματιών. Δύο προφανείς γενικεύσεις είναι :

- i. Σταθεροποιούμε όλους τους παράγοντες εκτός από έναν , ή και σε συνδυασμό ακόμα με την ηλικία, και
- ii. Σε μελέτες νοσηρότητας όπου το κριτήριο μελέτης είναι το επεισόδιο μιας συγκεκριμένης ασθένειας και όχι ενός θανάτου.

Θα συζητήσουμε επίσης την συνηθισμένη κατάσταση-την προτυποποίηση των δεικτών θνησιμότητας δηλαδή για τις ηλικίες.

Η βασική ιδέα στην προτυποποίηση είναι η εισαγωγή ενός *σταθερού πληθυσμού* με καθορισμένη ηλικιακή δομή. Η θνησιμότητα για οποιονδήποτε *ειδικό πληθυσμό* είναι τότε τροποποιημένη ώστε να επιτρέπει τις αισθητές διαφορές στην ηλικιακή δομή ανάμεσα στους σταθερούς και τους ειδικούς πληθυσμούς. Υπάρχουν δύο προσεγγίσεις : Μία *άμεση* και μία *έμμεση* μέθοδος προτυποποίησης. Μία σύντομη περιγραφή δίνεται για τις παραπάνω μεθόδους από το Liddell (1960) , τον Kalton (1968) η τους Hill και Hill (1991).

Θα χρησιμοποιηθεί ο ακόλουθος πίνακας

	Standard			Special		
	(1)	(2)	(3)	(4)	(5)	(6)
Age Group	Population	Deaths	Death Rate (2)/(1)	Population	Deaths	Death Rate (5)/(6)
1	N_1	R_1	P_1	n_1	r_1	p_1
⋮						
i	N_i	R_i	P_i	n_i	r_i	p_i
⋮						
k	N_k	R_k	P_k	n_k	r_k	p_k

Άμεση μέθοδος

Στην άμεση μέθοδο ο δείκτης θνησιμότητας προτυποποιείται βάση της ηλικιακής δομής του σταθερού πληθυσμού. Ο άμεσα προτυποποιημένος δείκτης για τον συγκεκριμένο πληθυσμό είναι, συνεπώς,

$$p' = \frac{\sum N_i p_i}{\sum N_i} \quad (4.5)$$

Επιτυγχάνεται από την εφαρμογή των ειδικών δεικτών θνησιμότητας, p_i , για τα μεγέθη των σταθερών πληθυσμών, N_i . Εναλλακτικά, το p' μπορεί να θεωρηθεί ένας σταθμικός μέσος του p_i , χρησιμοποιώντας τα N_i ως συντελεστή βαρύτητας. Η διακύμανση του p' μπορεί να εκτιμηθεί από τον τύπο

$$\text{var}(p') = \frac{\sum (N_i^2 p_i q_i / n_i)}{(\sum N_i)^2} \quad (4.6)$$

Όπου το $q_i = 1 - p_i$ και αν όπως συμβαίνει στις περισσότερες περιπτώσεις, τα p_i είναι πολύ μικρά, τότε η διωνυμική διακύμανση των p_i , $p_i q_i / n_i$ ίσως αντικατασταθεί από τον Poisson όρο $p_i / n_i (= r_i / n_i^2)$, δίνοντας εναλλακτικά τον τύπο

$$\text{var}(p') \approx \frac{\sum (N_i^2 p_i / n_i)}{(\sum N_i)^2} \quad (4.7)$$

Για να συγκρίνουμε δύο συγκεκριμένους πληθυσμούς, A και B, μπορούμε να υπολογίσουμε έναν τυποποιημένο δείκτη για καθένα (δηλαδή από τα p'_A και p'_B). Θεωρούμε επομένως

$$\bar{d} = p'_A - p'_B$$

$$\bar{d} = \frac{\sum N_i (p_{A_i} - p_{B_i})}{\sum N_i}$$

Η μέθοδος διαφέρει από αυτήν του Cochran's test μόνο στο ότι χρησιμοποιούμε ένα διαφορετικό σύστημα συντελεστών βαρύτητας.

Η διακύμανση δίνεται από

$$\text{var}(\bar{d}) = \frac{\sum N_i^2 \text{var}(d_i)}{(\sum N_i)^2} \quad (4.8)$$

με το $\text{var}(d_i)$ να δίνεται από τον τύπο

$$\text{var}(d_i) = p_{0i} q_{0i} (n_{A_i} + n_{B_i}) / n_{A_i} n_{B_i}$$

Όταν τα p_{0i} είναι μικρά τα q_{0i} μπορούμε προσεγγιστικά να τα βάλουμε ίσα με 1.

Αν απαιτείται να συγκρίνουμε δύο συγκεκριμένους πληθυσμούς χρησιμοποιούμε τις αναλογίες των προτυποποιημένων δεικτών, p'_A / p'_B , και τότε η διακύμανση της αναλογίας μπορεί να επιτυγχάνεται χρησιμοποιώντας τον τύπο (4.6).

Η διακύμανση που δίνεται από τον τύπο (4.7) ίσως να μην είναι ικανοποιητική για την κατασκευή των διαστημάτων εμπιστοσύνης αν το πλήθος των θανάτων για τις ξεχωριστές ηλικιακές ομάδες είναι μικρό, αφού η κανονική προσέγγιση στην περίπτωση αυτή δεν είναι ικανοποιητική και τα Poisson όρια είναι ασύμμετρα. Ο τυποποιημένος δείκτης (4.5) είναι ένα σταθμισμένο άθροισμα των ποσοτήτων Poisson, r_i . Ο Dobson

(1991) έδωσε μία μέθοδο υπολογισμού ενός προσεγγιστικού διαστήματος εμπιστοσύνης βασισμένο στο διάστημα εμπιστοσύνης του συνολικού αριθμού θανάτων.

Παράδειγμα 2:

Στον πίνακα 4.1 θα μπορούσε να υπολογιστεί ένας προτυποποιημένος δείκτης p' για κάθε πληθυσμό. Τι θα έπρεπε όμως να ληφθεί ως σταθερός πληθυσμός; Δεν υπάρχει μοναδική απάντηση στο ερώτημα αυτό. Η επιλογή ίσως να μην έχει σημαντική επίδραση στην σύγκριση των δύο πληθυσμών, αν και σίγουρα θα επηρεάσει τις απόλυτες τιμές των προτυποποιημένων δεικτών. Αν η σύγκριση μεταξύ των ειδικών ηλικιακών δεικτών είναι πολύ διαφορετική στις διαφορετικές ηλικιακές ομάδες, ίσως πρέπει να αναλογιστούμε αν οι προτυποποιημένοι δείκτες επιθυμούμε να εκφράζουν ξεχωριστά την θέση σε καθορισμένα τμήματα της ηλικιακής κλίμακας. Για παράδειγμα θα ήταν επιθυμητό να δώσουμε λιγότερη βαρύτητα σε μεγαλύτερες ηλικιακές ομάδες γιατί σκοπός της μελέτης είναι κατά κύριο λόγο η σύγκριση της θνησιμότητας σε νεαρές ηλικίες, διότι οι πληροφορίες που αφορούν άτομα μεγαλύτερης ηλικίας είναι λιγότερο αξιόπιστες, ή γιατί οι δείκτες θνησιμότητας σε μεγαλύτερες ηλικίες επηρεάζονται από το δειγματικό σφάλμα.

Στην βάση του πίνακα 4.1 δίνονται προτυποποιημένοι δείκτες με τρεις επιλογές σταθερού πληθυσμού.

- (1) πληθυσμός A
- (2) πληθυσμός B
- (3) υποθετικός πληθυσμός C, με ανάλογη κατανομή να είναι ενδιάμεσα από την A και B, δηλ

$$N_{C_i} = \frac{1}{2} \left(\frac{n_{A_i}}{\sum n_{A_i}} + \frac{n_{B_i}}{\sum n_{B_i}} \right)$$

Επισημαίνουμε ότι για την μέθοδο (1) ο τυποποιημένος δείκτης για το A είναι ίδιος όπως και για τον πρακτικό δείκτη. Παρόμοια για την (2) ο προτυποποιημένος δείκτης για το B είναι ο ίδιος όπως ο πρακτικός δείκτης. Αν και οι απόλυτες τιμές των προτυποποιημένων δεικτών είναι διαφορετικές για τις τρεις επιλογές του σταθερού πληθυσμού και οι αντίθετες είναι γενικά ίδιες για κάθε περίπτωση.

Έμμεση μέθοδος

Η μέθοδος αυτή είναι περισσότερο πρακτική από την άποψη της σύγκρισης παρατηρούμενων και αναμενόμενων θανάτων παρά λαμβάνοντας υπόψη τους προτυποποιημένους δείκτες. Στον συγκεκριμένο πληθυσμό ο συνολικός αριθμός θανάτων που παρατηρήθηκαν είναι $\sum r_i$.

Το πλήθος των θανάτων που θα αναμενόταν, αν οι ειδικοί ηλικιακοί δείκτες θνησιμότητας ήταν οι ίδιοι όπως στον σταθερό πληθυσμό, είναι $\sum n_i P_i$. Όλη η εμπειρική θνησιμότητα του σταθερού πληθυσμού ίσως εκφράζεται λαμβάνοντας υπόψη αυτήν του σταθερού πληθυσμού από την αναλογία των παρατηρούμενων θανάτων προς τους αναμενόμενους:

$$M = \frac{\sum r_i}{\sum n_i P_i} \quad (4.9)$$

Όταν ο παραπάνω τύπος πολλαπλασιαστεί με 100 και εκφραστεί σαν ποσοστό, τότε αναφέρεται ως τυποποιημένη αναλογία θνησιμότητας (*standardized mortality ratio-SMR*)

Για να επιτευχθεί η διακύμανση του M μπορούμε να χρησιμοποιήσουμε το συμπέρασμα $\text{var}(r_i) = n_i p_i q_i$ και να θεωρήσουμε τα P_i σαν συνεχείς μεταβλητές χωρίς οποιαδήποτε δειγματική αυξομείωση (αφού συχνά θα θέλουμε να συγκρίνουμε μία SMR με κάποια άλλη χρησιμοποιώντας τον ίδιο σταθερό πληθυσμό. Σε οποιαδήποτε περίπτωση ο σταθερός πληθυσμός θα είναι συχνά πολύ μεγαλύτερος από τον ειδικό πληθυσμό, και η $\text{var}(P_i)$ θα είναι πολύ μικρότερη από την $\text{var}(p_i)$. Αυτό δίνει,

$$\text{var}(M) = \frac{\sum n_i p_i q_i}{(\sum n_i P_i)^2} \quad (4.10)$$

Όπως συνήθως αν το p_i είναι μικρό, το $q_i \approx 1$ και

$$\text{var}(M) = \frac{\sum r_i}{(\sum n_i P_i)^2} \quad (4.11)$$

Τα διαστήματα εμπιστοσύνης για το M που κατασκευάστηκαν χρησιμοποιώντας τον τύπο (4.11) έχουν συνολικό αριθμό θανάτων $\sum r_i$ πολύ μικρό.

Αν ο σκοπός του υπολογισμού της $\text{var}(M)$ είναι το να καταλάβουμε αν το M διαφέρει σημαντικά από την μονάδα, θα μπορούσαμε να το πάρουμε ως $n_i P_i Q_i$, πάνω στην υπόθεση ότι το p_i διαφέρει από τις δειγματικές διακυμάνσεις. Αν πάλι το P_i είναι μικρό, το $Q_i \approx 1$, και έχουμε την ισοδυναμία των αναμενόμενων θανάτων

$$\text{var}(M) = \frac{\sum n_i P_i}{(\sum n_i P_i)^2} = \frac{1}{\sum n_i P_i} \quad (4.12)$$

Συμβολίζοντας τον αριθμητή και τον παρονομαστή του (4.9) με O και E αντίστοιχα (για παρατηρούμενους-observed και αναμενόμενους-expected θανάτους) ένα προσεγγιστικό τεστ σημαντικότητας θα μπορούσε να είναι το να θεωρήσουμε ότι το O ακολουθεί μια κατανομή *Poisson* με μέσο E . Αν το μέσο δεν είναι πολύ μικρό, η κανονική προσέγγιση

της *Poisson*, οδηγεί στην χρήση του $(O-E)/\sqrt{E}$ ως μία σταθεροποιημένη κανονική κατανομή, ή ισοδύναμα, $(O-E)^2/E$ ως μία $\chi^2_{(1)}$ διακύμανση. Αυτός είναι φυσικά, ο οικείος τύπος για την διακύμανση $\chi^2_{(1)}$.

Παράδειγμα 3:

Ο πίνακας 4.2 δείχνει κάποια εργασιακά δεδομένα θνησιμότητας, ένας τομέας στον οποίο χρησιμοποιείται παραδοσιακά το SMR. Ο ειδικός πληθυσμός είναι αυτός των αγροτών του 1951, ηλικίας 20 έως 65. Ο σταθερός πληθυσμός αποτελείται από όλους τους άντρες σε αυτές τις ηλικιακές ομάδες, είτε εργάζονται είτε έχουν συνταξιοδοτηθεί. Οι θάνατοι των αγροτών πάνω από μία περίοδο πέντε ετών χρησιμοποιούνται ώστε να βοηθήσουν να μειωθούν τα δειγματικά σφάλματα, και οι παρατηρούμενοι και αναμενόμενοι αριθμοί εκφράζονται με μία βάση πέντε ετών. Το SMR είναι

$$100M = \frac{(100)(7678)}{11005} = 69.8\%$$

και

$$\begin{aligned} \text{var}(SMR) &= 10^4 \text{var}(M) \\ &= \frac{(10^4)(7678)}{(11005)^2} \quad , \text{από (4.11)} \\ &= 0.634 \end{aligned}$$

και

$$SE(SMR) = 0.80\%$$

Το μικρό ποσό του σταθερού σφάλματος (SE) του SMR στο παράδειγμα 4.3 είναι τυπικό παράδειγμα των vital στατιστικών δεδομένων, και είναι ο λόγος για τον οποίο τα δειγματικά σφάλματα συχνά αγνοούνται σ' αυτόν τον τύπο επαγγελμάτων. Πράγματι, υπάρχουν προβλήματα στην ερμηνεία της εργατικής στατιστικής θνησιμότητας που συχνά επισκιάζουν τα δειγματικά σφάλματα. Για παράδειγμα, τα επαγγέλματα ίσως είναι λιγότερο αξιόπιστα κατά την καταγραφή τους στην απογραφή παρά όταν αναφέρονται ως αιτίες θανάτου και αυτό μπορεί να οδηγήσει σε μεροληψία για τους εκτιμώμενους δείκτες θνησιμότητας. Ακόμη κι αν τα δεδομένα είναι εξ' ολοκλήρου αξιόπιστα, δεν είναι ξεκάθαρο πότε ένα ιδιαίτερα υψηλό ή χαμηλό SMR για ένα συγκεκριμένο επάγγελμα αντανακλά το ρίσκο υγείας για το συγκεκριμένο επάγγελμα ή την τάση για επίλεκτες ομάδες ανθρώπων να εισχωρήσουν. Στο παράδειγμα 4.3, για παράδειγμα, το SMR των αγροτών ίσως είναι χαμηλό, γιατί η αγροτική καλλιέργεια είναι υγιές επάγγελμα, ή γιατί οι άνθρωποι με προβλήματα υγείας δεν ασχολούνται με τις αγροτικές δουλειές ή είναι πιο πιθανό να τις εγκαταλείψουν.

4.4 Έρευνες για την διερεύνηση συνδέσμων

Μια ερώτηση που τίθεται συχνά στις επιδημιολογικές έρευνες και αφορά την αιτιολογία της ασθένειας είναι αν κάποιες ενδείξεις για την ασθένεια συνδέονται με προσωπικά χαρακτηριστικά ή συνήθειες, με συγκεκριμένες περιβαλλοντολογικές συνθήκες κάτω από τις οποίες ένα άτομο ζει και εργάζεται, ή με συγκεκριμένα βιώματα του ατόμου. Παραδείγματα τέτοιων ζητημάτων είναι τα εξής:

1. Είναι ο κίνδυνος του θανάτου από καρκίνο στον πνεύμονα συνδεδεμένος με τον βαθμό του καπνίσματος, των πρόσφατων ή περασμένων χρόνων;
2. Είναι ο κίνδυνος θανάτου ενός παιδιού από λευχαιμία συνδεδεμένος με το εάν ή όχι η μητέρα δέχθηκε ακτινοβολία κατά την διάρκεια της εγκυμοσύνης;
3. Είναι ο κίνδυνος του να προβληθεί κάποιος από μία συγκεκριμένη ασθένεια αυξημένος για τα άτομα που έλαβαν θεραπεία ενός συγκεκριμένου φαρμάκου κατά τη διάρκεια μιας προηγούμενης ασθένειας;

Κάποιες φορές σε ζητήματα όπως τα παραπάνω μπορεί να δοθεί απάντηση από ελεγχόμενα πειράματα στα οποία ο πιθανός προσωπικός παράγοντας τίθεται στην ευχέρεια του ερευνητή αν θα τον διαχειριστεί ή θα τον παρακρατήσει. Στην περίπτωση 3, για παράδειγμα, είναι πιθανό για τον ερευνητή να χορηγήσει το υπό αμφισβήτηση φάρμακο σε κάποιους ασθενείς και όχι σε άλλους και να συγκρίνει τα αποτελέσματα. Σε τέτοιες περιπτώσεις τα ζητήματα ασχολούνται με αιτιολογικά συμπεράσματα. Είναι το φάρμακο μέρος αυτής της ασθένειας; Τις περισσότερες φορές, παρόλα αυτά, η εμπειρική προσέγγιση είναι εκτός ερώτησης. Ο ερευνητής επιθυμεί να παρατηρήσει αν υπάρχει σύνδεσμος μεταξύ ενός παράγοντα και μιας ασθένειας, και να αναλάβει τον κίνδυνο του εάν επιθυμεί να συμπεράνει έναν αιτιολογικό σύνδεσμο.

Αυτά τα ζητήματα διερευνώνται συνήθως με μελέτες παρά μέσω πειραμάτων. Στην περίπτωση αυτή συνήθως το πρωταρχικό ενδιαφέρον δεν στρέφεται προς τον ακριβή πληθυσμό προς διερεύνηση. Ένας λόγος είναι ότι στις επιδημιολογικές έρευνες τις περισσότερες φορές κάτι διοικητικά φαίνεται αδύνατον να μελετηθεί ένας εθνικός ή τοπικός πληθυσμός, ακόμα και πάνω σε μία δειγματική βάση. Ο ερευνητής ίσως έχει, παρόλα αυτά, ευκολία στο να μελετήσει μία εργασιακή ομάδα ή έναν πληθυσμό που γεωγραφικά σχετίζεται με ένα συγκεκριμένο ιατρικό κέντρο. Δεύτερον, αν και οι μέσες τιμές ή οι σχετικές συχνότητες των διαφορετικών μεταβλητών διαφέρουν κάπως από πληθυσμό σε πληθυσμό, το μέγεθος και ο έλεγχος των συνδέσμων μεταξύ των μεταβλητών είναι απίθανο να διαφέρουν πολύ, ας πούμε για παράδειγμα, μεταξύ διαφορετικών εργατικών ομάδων ή διαφορετικών γεωγραφικών πληθυσμών.

Υπάρχουν δύο βασικοί σχεδιασμοί για αιτιολογικές έρευνες:

η *case-control* (περίπτωση-έλεγχος) μελέτη, γνωστή και σαν *case-referent* (περίπτωση-αναφερόμενη) μελέτη και η μελέτη *cohort* (ακολουθία). Σε μία μελέτη *case-control* μία ομάδα ατόμων που έχει επηρεαστεί από την υπό μελέτη ασθένεια συγκρίνεται με μία υπό έλεγχο ομάδα ατόμων που δεν έχει επηρεαστεί από την ασθένεια αυτή. Η πληροφορία επιτυγχάνεται, συνήθως με έναν αναδρομικό τρόπο, από την συχνότητα σε κάθε ομάδα των ποικίλων περιβαλλοντολογικών και προσωπικών παραγόντων που ίσως συνδέονται με την ασθένεια. Αυτός ο τύπος έρευνας είναι πρακτικός στην μελέτη σπάνιων συνθηκών οι οποίες μπορεί να εμφανιστούν πολύ σπάνια σε ένα τυχαίο δείγμα πληθυσμού. Ξεκινώντας με μία ομάδα από επηρεασμένα άτομα, θεωρούμε ουσιαστικά ότι κάποιος

λαμβάνει ένα υψηλότερο δειγματικό ποσοστό όταν βρίσκεται στην ομάδα περιπτώσεων παρά στην ομάδα ελέγχου. Η μέθοδος είναι επίσης κατάλληλη όταν η κατηγοριοποίηση της ασθένειας είναι απλή, αλλά και σε περιπτώσεις όπου πρέπει να μελετηθούν πολλοί πιθανοί αιτιολογικοί παράγοντες. Ένα επιπλέον πλεονέκτημα της μεθόδου είναι εκείνο το οποίο, από τους μέσους της αναδρομικής έρευνας, μπορεί να επιτύχει συγκριτικά πολύ γρήγορα σχετικές πληροφορίες.

Σε μία μελέτη cohort ένας πληθυσμός ατόμων, επιλεγόμενος συνήθως με γεωγραφικά και επαγγελματικά κριτήρια παρά βασιζόμενος σε ιατρικούς παράγοντες, μελετάται είτε με ολοκληρωτική καταμέτρηση ή με ένα αντιπροσωπευτικό δείγμα. Ο πληθυσμός κατηγοριοποιείται βάση του παράγοντα ή των παραγόντων του άμεσου ενδιαφέροντος και ακολουθεί μια χρονολογική έτσι ώστε οι δείκτες των περιπτώσεων των διαφόρων ενδείξεων της ασθένειας να μπορούν να παρατηρηθούν και να συσχετιστούν με τις κατηγοριοποιήσεις λόγω αιτιολογικών παραγόντων. Ο αναμενόμενος τύπος της cohort μελέτης σημαίνει ότι θα επιμηκυνόταν κανονικά περισσότερο στον χρόνο σε σχέση με την μελέτη case-study και είναι πιθανό να είναι πιο περίπλοκα διαχειρίσιμο. Τα αντίστοιχα πλεονεκτήματα είναι ότι μπορούν να μελετηθούν ταυτόχρονα πολλές ιατρικές συνθήκες και αυτή η άμεση πληροφορία σχετικά με την υγεία κάθε ατόμου επιτυγχάνεται σε ένα χρονικό διάστημα.

Οι μελέτες case-control και cohort καλούνται συχνά *αναδρομικές* και *προοπτικές* μελέτες αντίστοιχα. Οι τελευταίοι αυτοί όροι είναι περισσότερο κατάλληλοι, αλλά η ονοματολογία μπορεί να γίνει παραπλανητική σε κάποιες περιπτώσεις αφού μία cohort μελέτη μπορεί να είναι εξ' ολοκλήρου βασισμένη σε αναδρομικό ιστορικό. Για παράδειγμα, αν είναι διαθέσιμο το ιατρικό ιστορικό των τελευταίων 30 ετών για τους εργαζόμενους σ' ένα συγκεκριμένο εργοστάσιο, μία προοπτική μελέτη ίσως σχετίζεται με τους εργαζόμενους που προσλήφθηκαν 30 χρόνια πριν και βασίζεται στο ιστορικό της υγείας τους κατά την διάρκεια των 30 αυτών ετών. Μια τέτοια μελέτη σε κάποιες περιπτώσεις καλείται *ιστορικά αναμενόμενη μελέτη (historical prospective study)*.

Ένα κεντρικό πρόβλημα σε μία case-control μελέτη είναι η μέθοδος επιλογής των ελέγχων. Ιδεατά, θα έπρεπε να είναι κατά μέσο όρο παρόμοια με όλες τις περιπτώσεις εκτός από την υπό μελέτη ιατρική συνθήκη και να συνδέονται με τους αιτιολογικούς παράγοντες. Οι περιπτώσεις θα επιλέγονται συχνά από ένα ή περισσότερα νοσοκομεία και τότε θα ικανοποιούνται τα χαρακτηριστικά του πληθυσμού που χρησιμοποιούν τα νοσοκομεία αυτά, όπως κοινωνικές και περιβαλλοντικές συνθήκες, όπως επίσης και εθνικά χαρακτηριστικά. Θα είναι επίσης επιθυμητό να επιλέγεται μία ομάδα ελέγχου από την ίδια περιοχή ή περιοχές, ίσως ακόμη κι από τα ίδια νοσοκομεία, αλλά που υποφέρει από αρκετά διαφορετικές ασθένειες ώστε να είναι απίθανο να προέρχονται από τους ίδιους αιτιολογικούς παράγοντες. Επιπλέον, οι συχνότητες με τις οποίες ανακαλύπτονται οι διάφοροι παράγοντες συνήθως ποικίλουν με την ηλικία και το φύλο. Οι συγκρίσεις ανάμεσα στις ομάδες case και control πρέπει, παρόλα αυτά, να λαμβάνουν υπόψη οποιεσδήποτε διαφορές ίσως υπάρχουν στις κατανομές ηλικίας και φύλου των δύο ομάδων. Τέτοιες τροποποιήσεις αποφεύγονται ευρέως με το να σχεδιάζεται κάθε επηρεασμένο άτομο να ζευγαρώνει με ένα άτομο από την ομάδα ελέγχου που επιλέχθηκε σκόπιμα στην ίδια ηλικία και στο ίδιο φύλο και που ικανοποιεί οποιοδήποτε άλλο δημογραφικό χαρακτηριστικό το οποίο θεωρείται ότι επηρεάζει την μελέτη.

Τα σχόλια που επισημάνθηκαν στην υποενότητα 4.2 για τα μη-δειγματικά σφάλματα είναι επίσης σχετικά με τις αιτιολογικές μελέτες. Τα μη-ανταποκρινόμενα

αποτελούν πάντοτε έναν πιθανό κίνδυνο και θα πρέπει να καταβάλλεται κάθε προσπάθεια να μειωθούν όσο είναι δυνατόν.

Παράδειγμα 4:

Οι Doll και Hill (1950) ανέφεραν τα αποτελέσματα από μία αναδρομική έρευνα σχετικά με την αιτιολογία του καρκίνου του πνεύμονα. Μία ομάδα από 709 ασθενείς με καρκίνωμα του πνεύμονα σε 20 νοσοκομεία συγκρίθηκε με μία ομάδα ελέγχου από 709 ασθενείς με καρκίνωμα του πνεύμονα και μία τρίτη ομάδα αποτελούμενη από 637 ασθενείς με καρκίνωμα του στομαχιού ή του εντέρου. Για κάθε ασθενή με καρκίνο του πνεύμονα επιλέχθηκε ένας ασθενής από την ομάδα ελέγχου από το ίδιο νοσοκομείο, του ίδιου φύλου και σε ηλικιακή απόκλιση ± 5 χρόνια. Σε κάθε ασθενή από κάθε ομάδα τέθηκε μία σειρά ίδιων ερωτήσεων από κάποιον υπάλληλο του ανθρώπινου δυναμικού.

Οι μόνες ουσιαστικές διαφορές ανάμεσα στις ομάδες control και case ήταν η αναφορά τους στην συνήθεια του καπνίσματος. Κάποια από τα συμπεράσματα φαίνονται αναλυτικά στον πίνακα 4.3. Η διαφορά στην αναλογία των μη-καπνιστών στις δύο ομάδες είναι ξεκάθαρα σημαντική. (Αν απαιτούνταν ένα τεστ για δεδομένα τέτοιου τύπου, μία κατάλληλη μέθοδος θα ήταν το τεστ για την διαφορά των δύο αναλογιών). Η ομάδα ασθενών με άλλου τύπου καρκίνο είχαν παρόμοιο ιστορικό καπνίσματος με αυτούς της ομάδας ελέγχου και διέφεραν εμφανώς από την ομάδα με καρκίνο του πνεύμονα. Οι συγκρίσεις που εσωκλείονται στην τρίτη ομάδα είναι περισσότερο περίπλοκες καθώς δεν μπορεί να ζευγαρωθεί κάθε ασθενής με κάποιο μέλος από την ομάδα ελέγχου ή την ομάδα με καρκίνο του πνεύμονα και επιπλέον υπάρχει μία διαφορά στην ηλικιακή κατανομή.

Πίνακας 4.3 : Η πρόσφατη κατανάλωση καπνού από ασθενείς με καρκίνωμα του πνεύμονα και οι ασθενείς υπό έλεγχο χωρίς καρκίνωμα του πνεύμονα (Doll και Hill, 1950)

	Daily consumption of cigarettes						Total
	Non smoker	1-	5-	15-	25-	50-	
<i>Male</i>							
Lung carcinoma	2	33	250	196	136	32	649
Control	27	55	293	190	71	13	649
<i>Female</i>							
Lung carcinoma	19	7	19	9	6	0	60
Control	32	12	10	6	0	0	60

Αυτή η δημοσίευση των Doll και Hill είναι ένα εξαιρετικό παράδειγμα του μελήματος που θα έπρεπε να παρέχεται ώστε να αποφεύγεται η μεροληψία εξαιτίας κάποιων απαραίτητων διαφορών ανάμεσα στις ομάδες case και control ή για να διαφοροποιηθούν κάποιες προδιαγραφές των δεδομένων. Αυτή η μελέτη, όπως και κάποιες άλλες παρόμοιες μελέτες, συμπεραίνουν μία ισχυρή συσχέτιση ανάμεσα στο κάπνισμα και τον κίνδυνο ενός ατόμου να προσβληθεί από καρκίνο του πνεύμονα. Σε αναδρομικές μελέτες τέτοιου τύπου, παρόλα αυτά, υπάρχει διαφωνία σχετικά με τις προϋποθέσεις με τις οποίες επιλέγεται η συγκεκριμένη ομάδα ελέγχου, για την οποία αποκτάται μικρή πληροφορία για τις συσχέτισεις που εσωκλείονται, ενώ δεν γνωρίζουμε τίποτα άλλο για την συσχέτιση καπνίσματος και ασθενειών πέρα από αυτές τις πληροφορίες που συλλέχθηκαν από την έρευνα. Οι Doll και Hill (1954, 1956, 1964) διεξήγαγαν μία cohort μελέτη στέλνοντας ερωτηματολόγια σε όλους τους 59600 γιατρούς του Ηνωμένου Βασιλείου τον Οκτώβρη του 1951. Δόθηκε ικανοποιητικό ποσοστό απαντήσεων της τάξης του 68.2% του πληθυσμού (34439 άντρες και 6194 γυναίκες). Κάποια από τα αποτελέσματα αυτά φαίνονται στον πίνακα 4.4. Οι ομάδες που ορίστηκαν σε διαφορετική κατηγορία καπνιστών έχουν και ηλικιακή διαφορά, και οι δείκτες θνησιμότητας που φαίνονται στον πίνακα έχουν προτυποποιηθεί με βάση την ηλικία. Το κάπνισμα φαίνεται πάλι να συνδέεται με μία οξεία αύξηση του δείκτη θνησιμότητας από καρκίνο του πνεύμονα, ενώ υπάρχει επιπλέον μία ισχυρή συσχέτιση με την χρόνια αποφρακτική πνευμονοπάθεια, και μία σχετικά πιο ανίσχυρη συσχέτιση με την αύξηση του δείκτη θνησιμότητας από ισχαιμικό καρδιακό νόσημα.

Η αναδρομική αυτή μελέτη παρέχει ισχυρές ενδείξεις ότι η συσχέτιση ανάμεσα στον καρκίνο του πνεύμονα και το κάπνισμα είναι θετική.

Σε μία μελέτη cohort στην οποία το περιστατικό μιας συγκεκριμένης ασθένειας είναι ιδιαίτερου ενδιαφέροντος, ίσως υιοθετείται μία προσέγγιση case-control για την ανάλυση των δεδομένων όλων των περιπτώσεων και μία ομάδα ελέγχου που επιλέγεται τυχαία και δεν εσωκλείει περιπτώσεις. Η προσέγγιση αυτή καλείται *συνθετική αναδρομική μελέτη (synthetic retrospective study, Mantel-1973)*. Συχνά οι έλεγχοι που επιλέγονται για κάθε περίπτωση προέρχονται από τυχαία δειγματοληψία για τις cohort που περιέχουν άτομα που δεν είναι περιστατικά την στιγμή που η συγκεκριμένη περίπτωση εμφανίζει την ασθένεια. Αυτή συνήθως αναφέρεται ως *μελέτη ελέγχου ένθετης υπόθεσης (nested case control study, Liddell-1977)*. Ένας παρεμφερής σχεδιασμός είναι και η *μελέτη περίπτωση-ακολουθία (case-control study)*, η οποία αποτελείται από ένα τυχαίο δείγμα ολόκληρης της ακολουθίας. Κάποια μέλη του δείγματος αυτού θα γίνουν περιπτώσεις και θα συμπληρωθούν από περιπτώσεις οι οποίες προκύπτουν στην ακολουθία που μένει. Οι σχεδιασμοί αυτοί είναι χρήσιμοι σε καταστάσεις που είναι απαγορευτική η μελέτη όλων των δεδομένων λόγω κόστους, ή όταν απαιτούνται ακριβή τεστ.

Το μέτρο του βαθμού της σύνδεσης ανάμεσα στον κίνδυνο της ασθένειας και την παρουσία ενός αιτιολογικού παράγοντα θα συζητηθεί με λεπτομέρεια σε επόμενη υποενότητα.

Πίνακας 4.4 : Προτυποποιημένοι ετήσιοι δείκτες θνησιμότητας ανάμεσα σε άντρες γιατρούς για τρεις αιτίες θανάτου .1951-91, που σχετίζονται με την συνήθεια του καπνίσματος. (Doll ,1994)

	Standardize death rate per 100000 men						
	Number of Deaths	non smokers	ex-smokers	Current smokers	Cigarette smokers (cigarettes per day)		
					1-14	15-24	25-
Lung cancer	893	14	58	209	105	208	355
Chronic Obstructive lung disease	542	10	57	127	86	112	225
Ischaemic heart disease	6438	572	678	802	892	892	1025

4.5 Σχετικός κίνδυνος(*Relative Risk*)

Στην προηγούμενη υποενότητα συζητήθηκαν οι μέθοδοι cohort και case control για την μελέτη της αιτιολογίας μιας ασθένειας. Σε τέτοιου τύπου μελέτες είναι συνηθισμένο να κάνουμε συγκρίσεις ανάμεσα σε ομάδες με διαφορετικά χαρακτηριστικά ,και στην συγκεκριμένη περίπτωση ανάμεσα σε μία ομάδα με άτομα που εκτίθενται σε κάποιον παράγοντα με μία ομάδα που δεν εκτίθεται σε αυτόν. Απαιτείται λοιπόν ένα μέτρο του αυξανόμενου ρίσκου (αν υπάρχει) της σύγκρισης μιας συγκεκριμένης ασθένειας στα εκτεθειμένα και μη εκτεθειμένα άτομα. Το μέτρο που συνήθως χρησιμοποιούμε είναι η αναλογία των περιστατικών στις δύο ομάδες που συγκρίθηκαν και είναι γνωστή ως *σχετικός κίνδυνος(relative risk- 'φ')*.

$$\varphi = I_E / I_{NE} \quad (4.13)$$

Όπου I_E και I_{NE} είναι οι δείκτες των περιστατικών στην εκτεθειμένη και μη εκτεθειμένη ομάδα αντίστοιχα. Το μέτρο αυτό αναφέρεται ως *αναλογία κινδύνου (risk ratio)*.

Σε μία μελέτη cohort ο σχετικός κίνδυνος μπορεί να εκτιμηθεί απευθείας αφού οι εκτιμήσεις είναι διαθέσιμες τόσο για το I_E όσο και για το I_{NE} . Σε μία case-control μελέτη όμως ο κίνδυνος δεν μπορεί να εκτιμηθεί απευθείας αφού δεν μπορεί να εκτιμηθεί ούτε το I_E ούτε το I_{NE} . Θα δείξουμε στην συνέχεια πως μπορούμε να επιτύχουμε μία χρήσιμη λύση στην περίπτωση αυτή.

Υποθέτουμε ότι σε έναν μεγάλο πληθυσμό κάθε άτομο έχει ταξινομηθεί ως θετικό ή αρνητικό σύμφωνα με κάποιον πιθανό αιτιολογικό παράγοντα, ή με θετικό ή αρνητικό σύμφωνα με κάποια κατάσταση ασθένειας. Ο παράγοντας ίσως βασίζεται σε κάποια πρόσφατη ταξινόμηση ή (περισσότερο συχνά στην αναδρομική μελέτη) στο ιστορικό του ατόμου. Η κατάσταση της ασθένειας ίσως αναφέρεται στην παρουσία ή απουσία μιας συγκεκριμένης κατηγορίας της ασθένειας μια συγκεκριμένη χρονική στιγμή, ή σε ένα συγκεκριμένο συμβάν (όπως η διάγνωση ή ο θάνατος) κατά την διάρκεια μιας προκαθορισμένης περιόδου- αυτό είναι, *εξάπλωση* ή *περιστατικό*, αντίστοιχα.

Οποιαδήποτε κατηγοριοποίηση του πληθυσμού γίνεται σε έναν πίνακα 2×2 , όπως ο ακόλουθος. Οι εισχωρήσεις στον πίνακα είναι αναλογίες του συνολικού πληθυσμού.

		<i>Ασθένεια</i>		
		+	-	
<i>Παράγοντας</i>	+	P_1	P_3	P_1+P_3
	-	P_2	P_4	P_2+P_4
		P_1+P_2	P_3+P_4	1

(I)

Αν οι αναλογίες ήταν γνωστές, οι συσχετίσεις ανάμεσα στον παράγοντα και την ασθένεια θα μπορούσαν να μετρηθούν από την αναλογία των κινδύνων του να είναι κάποιος θετικός αν έχει εκτεθεί στον παράγοντα ή αν δεν έχει εκτεθεί σ' αυτόν. Δηλαδή,

$$\begin{aligned}
 \text{Risk ratio} &= \frac{P_1}{(P_1 + P_2)} \div \frac{P_2}{(P_2 + P_4)} & (4.15) \\
 &= \frac{P_1(P_2 + P_4)}{P_2(P_1 + P_3)}
 \end{aligned}$$

Όταν οι περιπτώσεις γίνουν συμβάντα τότε η αναλογία κινδύνου(risk ratio) γίνεται σχετικός κίνδυνος (relative risk).

Τώρα σε κάποιες καταστάσεις (όχι σε όλες) στις οποίες έχουν γίνει αιτιολογικές μελέτες, το ποσοστό των ατόμων που ταξινομήθηκαν ως θετικά στην ασθένεια θα είναι πολύ μικρό. Αυτό συμβαίνει γιατί το P_1 είναι μικρό σε σχέση με το P_3 και το P_2 είναι μικρό σε σχέση με το P_4 . Σε μια τέτοια περίπτωση η (4.15) γράφεται ισοδύναμα,

$$\frac{P_1}{P_3} \div \frac{P_2}{P_4} = \frac{P_1 P_4}{P_2 P_3} (= \psi, \text{ας υποθέσουμε}) \quad (4.16)$$

Η αναλογία (4.16) καλείται *περιττή αναλογία (odds ratio)*, γιατί είναι αναλογία των P_1/P_3 προς P_2/P_4 . Αλλά κάποιες φορές θα την συναντήσουμε και ως *προσεγγιστικό σχετικό κίνδυνο (approximate relative risk)* ή απλά ως *σχετικό κίνδυνο (relative risk)*.

Οι περιττοί δείκτες (odds ratio (4.16)) θα μπορούσαν να εκτιμώνται από ένα τυχαίο δείγμα του πληθυσμού ή από μία απλή στρωματοποίηση από δύο επίπεδα του παράγοντα (όπως το να ξεκινήσει μία cohort μελέτη προτού γίνει εκτίμηση της ασθένειας). Θα μπορούσε να εκτιμηθεί επίσης μία απλή στρωματοποίηση δύο καταστάσεων της ασθένειας (πχ. από μία case-control μελέτη). Και είναι αυτός ο παράγοντας που κάνει την μελέτη ένα τόσο χρήσιμο μέτρο του σχετικού κινδύνου. Υποθέτουμε ότι διεξάγεται μία case-control μελέτη από ξεχωριστά επιλεγόμενα τυχαία δείγματα ασθενούντων και μη ασθενούντων ατόμων, και οι *συχνότητες* (όχι οι αναλογίες) είναι οι ακόλουθες,

		<i>Ασθένεια</i>		
		+	-	
		(περιπτώσεις)	(έλεγχοι)	
<i>Παράγοντας</i>	+	<i>a</i>	<i>c</i>	<i>a+c</i>
	-	<i>b</i>	<i>d</i>	<i>b+d</i>
		<i>a+b</i>	<i>c+d</i>	<i>n</i>

(II)

Συχνά, ο δειγματοληπτικός σχεδιασμός οδηγεί σε περιπτώσεις και ελέγχους ίδιου πλήθους. Δηλαδή το $a+b=c+d=n$. Τώρα, η αναλογία a/b μπορεί να θεωρηθεί μία εκτίμηση του c/d και το c/d παρόμοια μία εκτίμηση του P_3/P_4 . Ο παρατηρούμενος odds ratio είναι,

$$\psi = \frac{ad}{bc} \quad (4.18)$$

που είναι ο δείκτης του a/b προς c/d , και συνεπώς μπορεί να ληφθεί ως μία εκτίμηση του

$$\frac{P_1}{P_2} \div \frac{P_3}{P_4} = \frac{P_1 P_4}{P_2 P_3} (= \psi) \quad (4.19)$$

ο πληθυσμιακός odds ratio ή ο προσεγγιστικός relative risk (σχετικός κίνδυνος) καθορίζεται από τον τύπο (4.16)

Οι εικασίες ότι οι ομάδες case και control είναι τυχαία δείγματα από την ίδια σχετική πληθυσμιακή ομάδα είναι δύσκολο να ικανοποιήσει τις μελέτες case-control. Παρόλα αυτά, οι εκτιμήσεις των σχετικών κινδύνων προέρχονται από μελέτες case-control που συχνά συμφωνούν σε μεγάλο βαθμό με εκείνες που επιτυγχάνονται από τεκμηριωμένες cohort μελέτες. Στις αναδρομικές μελέτες, οι περιπτώσεις είναι συχνά συνδεδεμένες με άτομα της ομάδας ελέγχου για διαφορετικούς παράγοντες.

Παράδειγμα 5:

Ο πίνακας 5 συνοψίζει τα αποτελέσματα από 10 αναδρομικές έρευνες στις οποίες οι ασθενείς με καρκίνο του πνεύμονα και τα υπό έλεγχο άτομα ταξινομήθηκαν ως καπνιστές και μη καπνιστές. Στις περισσότερες ή σε όλες από αυτές τις μελέτες, οι περιπτώσεις και οι έλεγχοι θα ήταν ισοδύναμοι, αλλά τα πρωτότυπα δεδομένα συνήθως δεν περιέχουν τις επαρκείς λεπτομέρειες ώστε να καταστήσουν ικανό οι σχετικοί κίνδυνοι να υπολογιστούν από την σχέση

$$R = \frac{r}{s}$$

Τα δεδομένα συντάχθηκαν από τον Cornfield (1956) και σε αυτά έχει κάνει επίσης αναφορά και ο Gart (1962).

Ορίζοντας το w_i ως το αντίστροφο του $var(\ln \psi_i)$, ο σταθμισμένος μέσος όρος είναι

$$\frac{\sum w_i \ln \psi_i}{\sum w_i} = \frac{161.36}{105.4} = 1.531$$

και ο συσχετισμένος μέσος του ψ είναι $\exp(1.531) = 4.62$

Για το ετερογενές τεστ, το $\chi^2_{(9)}$ είναι

$$\sum w_i (\ln \psi_i)^2 - \frac{(\sum w_i \ln \psi_i)^2}{\sum w_i} = 253.678 - 247.061 = 6.62 \quad (P=0.68).$$

Δεν υπάρχει καμία ισχυρή ένδειξη της ετερογένειας ανάμεσα στις διαφορετικές εκτιμήσεις. Είναι, φυσικά, πιθανό ότι ο σχετικός κίνδυνος διαφέρει σε κάποιο βαθμό από μελέτη σε μελέτη, ειδικά όταν ο παράγοντας «καπνιστές» καλύπτει ένα μεγάλο εύρος δραστηριοτήτων. Παρόλα αυτά, η δειγματοληπτική διακύμανση των διαφορετικών εκτιμήσεων είναι προφανώς τόσο μεγάλη ώστε δεν είναι ικανή να αναδείξει την πραγματική διακύμανση. Αν θεωρήσουμε ότι όλες οι διακυμάνσεις οφείλονται στο δειγματοληπτικό σφάλμα, η απόκλιση από το σταθμισμένο μέσο του $\ln \psi_i$ μπορεί να επιτευχθεί από τον τύπο

$$\frac{1}{\sum w_i} = 0.00949$$

Προσεγγιστικά τα 95% διαστήματα εμπιστοσύνης για το $\ln \psi$ είναι

$$1.531 \pm (1.96)\sqrt{0.00949} = 1.340 \text{ και } 1.722$$

Τα αντίστοιχα όρια για το ψ που επιτυγχάνονται εκθετικά είναι

$$R_{MH} = \frac{302.840}{64.687} = 4.68$$

Η στατιστική που ελέγχει ότι αυτή η εκτιμήτρια διαφέρει από την μονάδα είναι

$$\chi^2_{MH} = (3793 - 3554.85)^2 / 193.17 = 293.60 \quad (P < 0.001)$$

Το τεστ που υπολογίζει τα διαστήματα εμπιστοσύνης από την σχέση

$$SE(\ln R_{MH}) = (\ln R_{MH}) / X_{MH}$$

δίνει

$$\begin{aligned} SE(\ln R_{MH}) &= \ln(4.68) / \sqrt{293.60} \\ &= 1.543 / 17.13 \\ &= 0.0901 \end{aligned}$$

και προσεγγιστικά τα 95% διαστήματα εμπιστοσύνης για το $\ln \psi$ είναι

$$1.531 \pm (1.96) \times 0.0901 = 1.366 \text{ και } 1.720$$

Τα αντίστοιχα όρια του ψ είναι τα εκθετικά 3.92 και 5.58. Η τιμή εκτίμησης και τα όρια εμπιστοσύνης είναι παρόμοια του Woolf.

Μία άλλη κατάσταση που προκύπτει επίσης πολύ συχνά είναι όταν ο παράγοντας, εκτός από το να έχει δύο μόνο επίπεδα, την παρουσία και την απουσία, έχει τρία ή περισσότερα επίπεδα τα οποία έχουν καθορισμένη σειρά. Για παράδειγμα εκτός από το να κατατάσσονται οι άνθρωποι σε καπνιστές και μη καπνιστές, θα μπορούσαν να διαχωριστούν οι καπνιστές περαιτέρω με βάση την ημερήσια ποσότητα τσιγάρων. Ο Mantel (1953) επέκτεινε την διαδικασία Mantel-Haenszel ώστε ένα δημιουργήσει νέο τεστ που θα δίνει συμπεράσματα με βάση την τάση των δεδομένων και το οποίο θα είναι ικανό συνδυάσει ένα αριθμό υποσυνόλων των δεδομένων.

Οι μέθοδοι που συζητήθηκαν παραπάνω είναι ισχυρές για τις συνηθισμένες καταστάσεις που προκύπτουν, όπου απαιτείται να αναλυθεί ο σύνδεσμος ανάμεσα σε μία ασθένεια και έναν απλό παράγοντα στον οποίο εκτίθεται το άτομο, λαμβάνοντας υπόψη και κάποιον άλλο παράγοντα, ή το πολύ κάποιους από αυτούς, και όπου η επίδραση των άλλων παραγόντων ίσως φαίνεται ξεκάθαρα από κατάλληλες υποδιαιρέσεις των δεδομένων σε στρώματα. Οι μέθοδοι, από την άλλη πλευρά, είναι ακατάλληλοι όταν μας ενδιαφέρουν αρκετοί παράγοντες στους οποίους εκτίθεται το άτομο και η χορήγηση κάθε παράγοντα πρέπει να γίνεται λαμβάνοντας υπόψη και τους άλλους. Επιπλέον, εκεί όπου αρκετοί παράγοντες θα πρέπει να αναπροσαρμοστούν, η στρωματοποίηση εσωκλείει ένα μεγάλο αριθμό στρωμάτων. Στην περίπτωση που κάθε στρώμα περιέχει ένα μόνο άτομο τότε οι συγκρίσεις εντός των στρωμάτων θα είναι αδύνατες. Στις περιπτώσεις αυτές χρησιμοποιούνται περισσότερο γενικές μέθοδοι και είναι απαραίτητο το μοντέλο λογιστικής παλινδρόμησης. Αυτό συμβαίνει εξαιτίας του συνδυασμού των δύο χαρακτηριστικών. Αρχικά, όπως παρατηρήσαμε και παραπάνω, οι odds ratio μπορούν να εκτιμηθούν από μία case-control μελέτη, και έπειτα, ο λογάριθμος του odds ratio είναι μία διαφορά των logits. Για μία case-control μελέτη υπολογίζονται όλοι οι συντελεστές της σχέσης

$$\ln\left(\frac{\mu}{1-\mu}\right) = \beta_0 + \beta_1\chi_1 + \beta_2\chi_2 + \dots + \beta_p\chi_p$$

εκτός από τον σταθερό όρο β_0 , ο οποίος τροποποιείται από τον ερευνητή ανάλογα με το μέγεθος των ομάδων case και control. Η εκτίμηση ενός συντελεστή παλινδρόμησης μετρά τον λογάριθμο του odds ratio για μία αλλαγή της μονάδας στην αντίστοιχη μεταβλητή, και έτσι οι εκθετικές των συντελεστών παλινδρόμησης δίνουν εκτιμώμενα odds ratio ή προσεγγιστικά τους δείκτες κινδύνου.

Παρατηρήσαμε παραπάνω ότι, για μία case-control μελέτη όπου επιλέγεται ένας έλεγχος που ταιριάζει σε κάθε περίπτωση, ίσως εφαρμόζεται η Mantel-Haenszel μέθοδος, και οδηγεί σε ένα τεστ McNemar. Όταν περιοριστεί το πλήθος των διαθέσιμων περιπτώσεων, ίσως αυξηθεί η πρόβλεψη, σε κάποιο βαθμό, επιλέγοντας περισσότερους του ενός ελέγχους σε κάθε περίπτωση. Η μέθοδος Mantel-Haenszel μπορεί να εφαρμοστεί αντιμετωπίζοντας κάθε σύνολο περιπτώσεων και τους ελέγχους τους σαν ένα υποσύνολο. Η ανάλυση μπορεί να απλοποιηθεί ομαδοποιώντας μαζί τα σύνολα που παρέχουν ίδιες πληροφορίες. Για παράδειγμα, με τρεις ελέγχους σε κάθε περίπτωση εκείνα τα σύνολα στα οποία η περίπτωση και οι δύο στους τρεις ελέγχους είναι θετικοί για την έκθεση στον παράγοντα παρέχουν όλα την ίδια πληροφορία.

Σε μία case-control μελέτη ταιριάσματος ίσως υπάρχουν συνδιακυμάνσεις καταγεγραμμένες σε κάθε άτομο και ίσως είναι απαραίτητο να λάβουμε υπόψη την επιρροή των συνδιακυμάνσεων στην πιθανότητα της ασθένειας. Οι διακυμάνσεις ίσως εσωκλείονται σε ένα μοντέλο λογιστικής παλινδρόμησης αλλά θα πρέπει να γίνεται και η περιγραφή του ταιριάσματος. Αυτό γίνεται δουλεύοντας υποθετικά εντός του κάθε συνόλου. Υποθέτουμε ότι το \mathbf{x} αντιπροσωπεύει το σύνολο των διακυμάνσεων, εσωκλείοντας τον εκτιθέμενο παράγοντα του βασικού μας ενδιαφέροντος, και χρησιμοποιώντας το 0 για την περίπτωση και από 1 έως c για τους c ελέγχους. Στη συνέχεια χρησιμοποιώντας το μοντέλο λογιστικής παλινδρόμησης και υποθέτοντας ότι η πιθανότητα της ασθένειας είναι μικρή, έτσι ώστε το logit να είναι προσεγγιστικά ίσο με τον λογάριθμο της πιθανότητας της ασθένειας, η πιθανότητα της ασθένειας κάθε μέλους του συνόλου είναι $\exp(\beta_0 + \beta^T \mathbf{x}_i)$ με $i=0,1,2,\dots,c$. Η παράμετρος β_0 έχει έναν δείκτη s για να υποδείξει ότι εξαρτάται από το συγκεκριμένο σύνολο. Τώρα κάθε ομάδα επιλέγεται ώστε να εσωκλείει ακριβώς μία περίπτωση ασθένειας. Συνεπώς η πιθανότητα του να είναι η παρατηρούμενη περίπτωση δίνεται από τον τύπο

$$\exp(\beta_{0s} + \beta^T \mathbf{x}_0) / \sum \exp(\beta_{0s} + \beta^T \mathbf{x}_i)$$

όπου το άθροισμα βασίζεται σε όλα τα μέλη του συνόλου, $i=0,1,\dots,c$.

Ο αριθμητής και ο παρονομαστής του παραπάνω κλάσματος περιέχει τον κοινό παράγοντα $\exp(\beta_{0s})$ κι έτσι η πιθανότητα εκφράζεται ως

$$\exp(\beta^T \mathbf{x}_0) / \sum \exp(\beta^T \mathbf{x}_i) \quad (4.20)$$

ένα συμπέρασμα που αρχικά δόθηκε από τον Thomas στο Liddell (1977). Μία έκφραση του τύπου (4.20) επιτυγχάνεται με τον τρόπο αυτό για κάθε περίπτωση και όλες μαζί συνδυάζονται για να συνθέσουν την πιθανοφάνεια. Τότε το β εκτιμάται από την μέθοδο της μέγιστης πιθανοφάνειας. Αν εργαστούμε υποθετικά εντός των συνόλων ο μεγάλος όγκος των παραμέτρων β_{0s} (ένα για κάθε περίπτωση) εξαλείφεται. Η περιγραφή της μεθόδου που αναπτύχθηκε παραπάνω ήταν αρκετά συνοπτική. Λεπτομερειακότερη περιγραφή της μεθόδου συναντάμε στις αναλύσεις που αναφέρουν οι Breslow και Day (1980) και ο Schlesselman (1982)

Το logrank test επομένως είναι παρόμοιο με το Mantel-Haenszel test. Συγκρίνοντας την σχέση (4.20) με τις σχέσεις

$$\psi_L = \frac{r}{(s+1)F_{0.025,2(s+1),2r}}$$

$$\psi_U = \frac{(r+1)F_{0.025,2(r+1),2s}}{s}$$

συμπεραίνουμε ότι η μέθοδος της λογιστικής παλινδρόμησης σε μία case-control μελέτη ταιριάσματος είναι παρόμοια με το αναλογικό μοντέλο κινδύνου για την ανάλυση επιβίωσης. Μία μελέτη case-control μπορεί να θεωρηθεί επίσης ένα δείγμα από μία

υποθετική cohort μελέτη. Πράγματι προς χάρη της εγκυρότητα θα έπρεπε να είναι έτσι. Δεν είναι, συνεπώς, σύμπτωση ότι οι μέθοδοι πρωταρχικά προέρχονται από την ανάλυση των cohort μελετών που εφαρμόζονται σε case-control μελέτες. Μια ανάλυση βασισμένη στον τύπο (4.20) μπορεί να πραγματοποιηθεί χρησιμοποιώντας το PROC PHREG ΠΡΟΓΡΑΜΜΑ στο SAS , με κάθε case-control σύνολο να καθορίζει ένα στρώμα.

4.6 Αποδοτέος κίνδυνος(Attributable Risk)

Όλοι οι υπαιτιγμοί ενός υπερβατικού ρίσκου δεν εξαρτώνται μόνο από το μέγεθος του σχετικού κινδύνου αλλά επιπλέον κι από την αναλογία του πληθυσμού που είναι θετικός ως προς έναν αιτιολογικό παράγοντα. Ένας μέτριος σχετικός κίνδυνος εφαρμοσμένος σε έναν μεγάλο πληθυσμό θα μπορούσε να παράγει περισσότερες περιπτώσεις της ασθένειας από ότι ένας υψηλός σχετικός κίνδυνος εφαρμοσμένος μόλις σε ένα μικρό ποσοστό του πληθυσμού. Ένα μέτρο του συνδέσμου , σύμφωνα με τον Levin (1953) , το οποίο λαμβάνει υπόψη την αναλογία του πληθυσμού σε κίνδυνο είναι ο *αποδοτέος κίνδυνος (attributable risk)*. Η ορολογία δεν είναι απόλυτα καθιερωμένη. Υπάρχουν διάφορα ονόματα για τον αποδοτέο κίνδυνο όπως *αιτιολογικό κλάσμα (aetiological fraction)* και *αποδοτέο κλάσμα (attributable fraction)*.

Ο πληθυσμιακός αποδοτέος κίνδυνος ορίζεται ως η αναλογία των περιπτώσεων του συνολικού πληθυσμού που μπορούν να αποδοθούν στον παράγοντα κινδύνου. Συνήθως υπολογίζεται στις περιπτώσεις που θεωρείται ζωτικής σημασίας να αναφέρουμε την αιτία από έναν παρατηρούμενο συσχετισμό. Τότε ερμηνεύεται ως η αναλογία των περιπτώσεων του πληθυσμού που οφείλονται στον παράγοντα, κι επομένως σαν ένα μέτρο της σημαντικότητας της εξάλειψης του παράγοντα ως μέρος της στρατηγικής πρόληψης της ασθένειας. Οι καταστάσεις στις οποίες η ερμηνεία αυτή είναι δικαιολογημένη είναι πιθανόν λίγες και υπάρχουν συγκεκριμένες δυσκολίες όταν έχουμε να διαχειριστούμε περισσότερες από μία αιτίες.

Υποθέτουμε ότι το I_P είναι η συχνότητα των ασθενούντων στον πληθυσμό και τα I_E και I_{NE} είναι οι συχνότητες των εκτιθέμενων και μη εκτιθέμενων, αντίστοιχα. Συνεπώς το πλεόνασμα που οφείλεται στον παράγοντα είναι $I_E - I_{NE}$ και διαιρούμενο από την συχνότητα των ασθενούντων του πληθυσμού δίνει τον πληθυσμιακό αποδοτέο κίνδυνο (population attribute risk)

$$\lambda_p = \frac{I_P - I_{NE}}{I_P} \quad (4.21)$$

Τώρα αν υποθέσουμε ότι ένα ποσοστό θ_E του πληθυσμού εκτίθεται στον παράγοντα ,έχουμε

$$\begin{aligned} I_P &= \theta_E I_E + (1 - \theta_E) I_{NE} \\ &= \theta_E \varphi I_{NE} + (1 - \theta_E) I_{NE} \end{aligned} \quad (4.22)$$

$$= I_{NE} [1 + \theta_E (\varphi - 1)]$$

όπου φ είναι ο σχετικός κίνδυνος και η δεύτερη γραμμή επιτυγχάνεται χρησιμοποιώντας τον τύπο (4.13). Αντικαθιστώντας τον παραπάνω τύπο στον (4.21), έχουμε

$$\lambda_p = \frac{\theta_E (\varphi - 1)}{1 + \theta_E (\varphi - 1)} \quad (4.23)$$

και ο αποδοτέος κίνδυνος μπορεί να εκτιμηθεί τώρα από τον σχετικό κίνδυνο και το ποσοστό του εκτιθέμενου πληθυσμού. Συχνά η αναλογία του αποδοτέου κινδύνου που πολλαπλασιάζεται με το 100 δίνει ένα αποτέλεσμα ποσοστιαίων όρων, ένα ποσοστό που αναφέρεται ως *ποσοστιαίος πληθυσμιακός αποδοτέος κίνδυνος (population attributable risk per cent)*.

Χρησιμοποιώντας τον πίνακα (I) προκύπτει μια εναλλακτική έκφραση. Αν η πιθανότητα της ασθένειας σε αυτούς που είναι θετικοί στον παράγοντα ήταν ίδια με το ποσοστό αυτών που ήταν αρνητικοί, τότε το ποσοστό του πληθυσμού που είναι θετικός και στον παράγοντα και στην ασθένεια θα ήταν

$$(P_1 + P_4) \times \frac{P_2}{P_2 + P_4}$$

Αφαιρώντας αυτό από τον τύπο της πραγματικής αναλογίας, P_1 , προκύπτει το πλεόνασμα της αναλογίας που σχετίζεται με τον παράγοντα, και, διαιρώντας την με την αναλογία των ασθενούντων, ο πληθυσμιακός αποδοτέος κίνδυνος (population attribute risk) δίνεται από

$$\begin{aligned} \lambda_p &= \left[P_1 - \frac{P_2 (P_1 + P_3)}{P_2 + P_4} \right] \div (P_1 + P_2) \\ &= \frac{P_1}{P_1 + P_2} \left[1 - \frac{P_2 (P_1 + P_3)}{P_1 (P_2 + P_4)} \right] \quad (4.24) \\ &= \frac{P_1}{P_1 + P_2} \left[1 - \frac{I_{NE}}{I_E} \right] \\ &= \theta_1 \frac{\varphi - 1}{\varphi} \end{aligned}$$

όπου το θ_1 είναι το ποσοστό των εκτιθέμενων περιπτώσεων στον παράγοντα.

Από την σχέση (4.23) συνεπάγεται ότι ο πληθυσμιακός αποδοτέος κίνδυνος μπορεί να εκτιμηθεί από οποιαδήποτε μελέτη η οποία παρέχει εκτιμήτριες τόσο για τον σχετικό κίνδυνο όσο και για το ποσοστό του εκτιθέμενου πληθυσμού στον παράγοντα. Άρα ίσως εκτιμάται τελικά από μια case-control μελέτη, αφού μπορεί να εκτιμηθεί

προσεγγιστικά ο σχετικός κίνδυνος και το ποσοστό του εκτιθέμενου πληθυσμού μπορεί να εκτιμηθεί από τους ελέγχους. Εναλλακτικά θα μπορούσε να χρησιμοποιηθεί η (4.24), αφού το θ_I μπορεί να εκτιμηθεί σε μια case-control μελέτη. Ξεκάθαρα ο πληθυσμιακός αποδοτέος κίνδυνος εκτιμάται από μία μελέτη cohort ενός τυχαίου δείγματος του συνολικού πληθυσμού, αλλά μόνο από στρωματοποιημένο δείγμα δύο επιπέδων έκθεσης εφόσον βέβαια το ποσοστό του πληθυσμού που εκτίθεται είναι γνωστό.

Ο αποδοτέος κίνδυνος ανάμεσα στα εκτιθέμενα άτομα καθορίζεται από την αναλογία των εκτιθέμενων περιπτώσεων στον παράγοντα. Δηλαδή

$$\lambda_E = \frac{I_E - I_{NE}}{I_E} = \frac{\varphi - 1}{\varphi} \quad (4.25)$$

Ο παραπάνω τύπος λαμβάνει υπόψη όρους του παράγοντα που αυξάνουν τον κίνδυνο, για παράδειγμα ο σχετικός κίνδυνος είναι μεγαλύτερος της μονάδας. Για έναν παράγοντα ο οποίος είναι προστατευτικός, όπως όταν ο σχετικός κίνδυνος είναι μικρότερος της μονάδας, ένας λίγο διαφορετικός τύπος είναι απαραίτητος. Αυτό συζητήθηκε από τον Kleinbaum (1982) αν και η απλούστερη προσέγγιση είναι να αντιστρέψουμε την κατεύθυνση του σχετικού κινδύνου. Για παράδειγμα «η άσκηση προλαμβάνει τις καρδιοπάθειες» είναι ισοδύναμη με την αντίστροφη πρόταση «η έλλειψη άσκησης είναι ένας παράγοντας ρίσκου»

Θεωρούμε τώρα την δειγματοληπτική διακύμανση μιας εκτιμήτριας του πληθυσμιακού αποδοτέου ρίσκου από μία μελέτη case-control, και δίνουμε ένα παράδειγμα. Υποθέτουμε ότι τα δεδομένα του πίνακα (II) προέρχονται από μία case-control μελέτη. Αφού ο λόγος $a/(a+b)$ εκτιμάει τον λόγο $P_1/(P_1+P_2)$ ή θ_I , και ο ad/bc εκτιμάει το ψ , το οποίο είναι περίπου το φ , η αντικατάσταση στον τύπο (4.24) δίνει μία εκτίμηση του λ_P που είναι η ακόλουθη

$$\begin{aligned} \lambda_P &= \frac{a}{a+b} \left(1 - \frac{bc}{ad} \right) \\ &= \frac{ad - bc}{d(a+b)} \end{aligned} \quad (4.26)$$

Ακριβώς η ίδια έκφραση επιτυγχάνεται αντικαθιστώντας τον λόγο $c/(c+d)$ στο θ_E στον τύπο (4.23). Είναι πρακτικό να δουλέψουμε με το $\ln(1 - \lambda_P)$ αφού όπως γνωρίζουμε από τον Walter (1975), η μεταβλητή αυτή ακολουθεί περίπου κανονική κατανομή.

$$\begin{aligned} 1 - \lambda_P &= \frac{bd + bc}{d(a+b)} \\ &= \frac{b}{a+b} \div \frac{d}{c+d} \end{aligned}$$

$$= \frac{1 - \theta_1}{1 - \theta_2}$$

και

$$\ln(1 - \lambda_P) = \ln(1 - \theta_1) - \ln(1 - \theta_2) ,$$

όπου το θ_1 είναι ένας εκτιμητής του $P_1/(P_1+P_2)$ και το θ_2 είναι ένας εκτιμητής του $P_3/(P_3+P_4)$. Οι θ_1 και θ_2 είναι ανεξάρτητοι των εκτιμητριών των αναλογιών , συνεπώς η διακύμανση θα δίνεται από τον τύπο

$$\text{var} \left[(1 - \lambda_P) \right] = \frac{a}{b(a+b)} + \frac{c}{d(c+d)}$$

4.7 Μέθοδος κατά ηλικία (Subject-years method)

Όπως ήδη έχουμε αναφέρει στην §4.4, μία ευρέως γνωστή μέθοδος είναι η μελέτη cohort, κατά την οποία μία ομάδα στρωματοποιείται με βάση την έκθεσή της σε κάποιον παράγοντα, με συνέπεια ο χρόνος επιβίωσης και η κατάσταση διαβίωσης κάθε ατόμου να καθορίζονται από την χρονική στιγμή διεξαγωγής της έρευνας. Μια περιγραφή των μεθόδων σχεδιασμού και της εφαρμογής της μελέτης cohort δόθηκε από τον Liddell (1988). Ίσως είναι πιθανό να χρησιμοποιηθεί το ιστορικό κάθε ατόμου για να καθοριστεί η έκθεσή του στο παρελθόν, μια διαδικασία που αναφέρεται ως *historical prospective cohort study* (ιστορικά αναμενόμενη μελέτη cohort), και χρησιμοποιείται κυρίως για έρευνα της υγείας των εργαζομένων ανάλογα με τις συνθήκες εργασίας. Τέτοιες μελέτες απαιτούν συνήθως έρευνα χρονικής διάρκειας 20 ετών. Σκοπός είναι να συγκριθεί η εμπειρική θνησιμότητα στις υποομάδες, όπως η υψηλή και η χαμηλή έκθεση, με σκοπό να εξακριβωθεί αν η έκθεση στον παράγοντα συμβάλλει στην αύξηση της θνησιμότητας. Καθώς τέτοιου τύπου έρευνες καλύπτουν μία μεγάλη χρονική περίοδο, τα άτομα θα μεγαλώνουν και ο κίνδυνος θνησιμότητας θα αλλάζει κατά την διάρκεια της μελέτης. Από την άλλη, ίσως υπάρχουν επίσης και περίοδοι που επιδρούν στον δείκτη θνησιμότητας. Τόσο η ηλικία όσο και οι περίοδοι αυτοί χρειάζεται να ληφθούν υπόψη στη μελέτη. Μία προσέγγιση είναι η μέθοδος κατά ηλικία (*subject-years method* ή *person-years method*) να αναφέρεται ως *τροποποιημένη προσέγγιση life table (modified life-table approach)*, μέθοδος την οποία χρησιμοποίησε νωρίτερα ο Doll (1952). Στην προσέγγιση αυτή ο αριθμός των θανάτων που αφορά κάθε ομάδα, ή κάποια υποομάδα, συνυπολογίζει και το πλήθος των θανάτων που θα αναμένονταν αν κάθε άτομο υφίσταται τον ίδιο δείκτη θανάτου με αυτόν του πληθυσμού του οποίου είναι μέρος η ομάδα στην οποία εντάσσεται το άτομο.

Η αναμενόμενη θνησιμότητα υπολογίζεται χρησιμοποιώντας δημοσιοποιημένους εθνικούς ή τοπικούς δείκτες θανάτου. Επιπλέον πρέπει να λαμβάνεται υπόψη η ηλικία κάθε ατόμου, τόσο κατά την στιγμή που εισήλθε στην ομάδα όσο και κατά την διάρκεια διεξαγωγής της έρευνας. Επίσης αφού ο ειδικός ηλικιακός δείκτης θανάτου εξαρτάται από την χρονική περίοδο που προκύπτει ο κίνδυνος, πρέπει να υπολογιστεί η ομάδα κάθε υποκειμένου χωριστά. Οι επίσημοι δείκτες θανάτου δημοσιοποιούνται συνήθως κάθε 5 χρόνια και ταξινομούνται σε ορθογώνιους πίνακες αποτελούμενους από κελιά (ανά στήλη και γραμμή). Για παράδειγμα ένα τέτοιο κελί είναι η ομάδα που αντιστοιχεί σε ηλικία 45-49 κατά την χρονική περίοδο 1976-80. Κάθε άτομο περνά από διαφορετικά κελιά κατά την διάρκεια της περιόδου παρακολούθησης και βιώνει κίνδυνο θανάτου ανάλογο με τον ηλικιακό κίνδυνο και τον δείκτη θανάτου του κελιού στο οποίο βρίσκεται. Ο κίνδυνος συσσωρεύεται όσο ένα άτομο είναι σε κίνδυνο κατά τη διάρκεια της μελέτης δηλαδή, μέχρι την ημερομηνία θανάτου ή για όσους είναι εν ζωή όταν η έρευνα τερματίζεται. Ο συσσωρευμένος κίνδυνος (accumulated risk) είναι ο ίδιος με τον αθροιστικό κίνδυνο (cumulative hazard), αναφορά του οποίου έχει γίνει στο 2^ο κεφάλαιο. Το αναμενόμενο πλήθος θανάτων επιτυγχάνεται με το να προσθέσουμε τους δείκτες θανάτου για όλα τα άτομα, και είναι υπολογιστικά πρακτικό το να προσθέσουμε τα χρόνια σε κίνδυνο κάθε κελιού για όλα τα άτομα πριν πολλαπλασιαστούν με τους αντίστοιχους δείκτες θανάτου.

Η μέθοδος μπορεί να εφαρμοστεί για το σύνολο των θανάτων κι επιπλέον για θανάτους που οφείλονται σε συγκεκριμένες αιτίες. Τον πίνακα ταξινόμησης ατόμων σε κίνδυνο βάση της ηλικίας, τον χρησιμοποιούμε και σε μελέτες για τις αιτίες, για τις οποίες όμως, οι πίνακες δεικτών θανάτου που χρησιμοποιούνται είναι διαφορετικοί. Κάτι τέτοιο συμβαίνει γιατί όταν θεωρούμε μια συγκεκριμένη αιτία θανάτου, το άτομο που πεθαίνει από οποιαδήποτε άλλη αιτία, θεωρείται για την έρευνα αποκομμένο εν ζωή άτομο. Για οποιαδήποτε αιτία θανάτου, θεωρούμε ότι ο παρατηρούμενος αριθμός είναι μία μεταβλητή που ακολουθεί κατανομή Poisson με προσδοκία ισοδύναμη με αυτήν του αναμενόμενου πλήθους. Η μέθοδος είναι παρόμοια με την έμμεση τυποποίηση των δεικτών θανάτου (§4.3) και η αναλογία των παρατηρούμενων προς τους αναμενόμενους θανάτους αναφέρεται συχνά ως τυποποιημένη αναλογία θνησιμότητας (SMR-Standardized Mortality Ratio).

Παράδειγμα 6:

Παρακολουθείται μία ομάδα 512 ατόμων που εργάζονται σε ένα εργοστάσιο με αμιάντο (Newhouse, 1985). Στην έρευνα περιλαμβάνονται όλοι οι άντρες που ξεκίνησαν να εργάζονται στην συγκεκριμένη δουλειά μετά το 1933 και αξιολογείται η θνησιμότητα μετά από περίοδο 10 ετών από την στιγμή της πρόσληψης κάθε εργαζομένου μέχρι το τέλος του 1980. Στον πίνακα 6 φαίνονται τα άτομα σε κίνδυνο κατά ηλικία και οι θανατικοί δείκτες για τον καρκίνο του πνεύμονα στην Αγγλία σε ηλικιακές ομάδες 5 ετών και για χρονικές περιόδους 10 ετών (στην πρωτότυπη έκφραση χρησιμοποιήθηκαν χρονικές περιόδοι 10 ετών).

Πίνακας 6: Άτομα σε κίνδυνο κατά ηλικία (*y*) εργαζομένων εκτιθέμενων σε αμίαντο και δείκτες θανάτου (*d*) για τον καρκίνο του πνεύμονα ανά 100000 άντρες στην Αγγλία και την Ουαλία.

Ηλικιακή Ομάδα	Περίοδος							
	1941-50		1951-60		1961-70		1971-80	
	<i>y</i>	<i>d</i>	<i>y</i>	<i>d</i>	<i>y</i>	<i>d</i>	<i>y</i>	<i>d</i>
25-29	82	2	62	1	13	1	3	1
30-34	148	3	273	4	156	3	43	2
35-39	74	9	446	10	435	8	141	6
40-44	41	21	395	25	677	22	290	16
45-49	33	46	229	58	749	54	485	46
50-54	23	78	172	124	590	119	642	106
55-59	14	112	158	216	399	226	621	201
60-64	11	137	109	294	288	370	479	346
65-69	4	137	78	343	185	508	273	530
70-74	0	107	47	325	124	562	151	651
75+	0	86	86	16	58	518	72	756

Το αναμενόμενο πλήθος θανάτων εξαιτίας καρκίνου του πνεύμονα είναι

$$(82 \times 2 + 62 \times 1 + 13 \times 1 + \dots + 72 \times 756) \times 10^{-5} = 13.8$$

Ο παρατηρούμενος αριθμός ήταν 67, έτσι το SMR είναι $67/13.8=4.9$. Ένα τεστ με την ψευδή υπόθεση να υποστηρίζει ότι ο εμπειρικός δείκτης θανάτου των εργαζομένων που εκτίθενται σε αμίαντο είναι

$$z = (67 - 13.8) / \sqrt{13.8} = 14.3 \quad (P < 0.001)$$

και προσεγγιστικά το 95% όρια εμπιστοσύνης για το SMR είναι 3.8 και 6.2.

Η μέθοδος συνήθως εφαρμόζεται για να συγκρίνει το SMR μεταξύ διαφορετικών υποομάδων, ή πιο γενικά, για να λάβει υπόψη συνδιακυμάνσεις καταγεγραμμένες για κάθε άτομο, εκφράζοντας το SMR σαν ένα μοντέλο παλινδρόμησης αναλογικού κινδύνου (proportional hazard regression model-Berry, 1983). Κάτι τέτοιο γίνεται με την χρήση ενός γενικευμένου γραμμικού μοντέλου. Για κάθε υποομάδα i , αν το m_i είναι ο αθροιστικός κίνδυνος (cumulative hazard) από την αναφορά του πληθυσμού και το γ_i , ο πολλαπλασιαστής του αναλογικού κινδύνου, τότε το μ_i , που είναι το αναμενόμενο πλήθος θανάτων, δίνεται από το

$$\ln \mu_i = \ln m_i + \ln \gamma_i$$

Αν το γ_i μοντελοποιείται με όρους από ένα σύνολο συνδιακύμανσεων

$$\ln \gamma_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}$$

τότε

$$\ln \mu_i = \ln m_i + \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \quad (4.27)$$

Αυτό το μοντέλο είναι παρόμοιο με το γενικευμένο γραμμικό μοντέλο της μεταβλητής Poisson αλλά περιέχει τον επιπρόσθετο όρο $\ln m_i$, στον οποίο κάποιες φορές κάνουμε αναφορά με τον όρο *offset* (αντιστάθμιση), και ο οποίος εξασφαλίζει ότι η ηλικία και η χρονική περίοδος είναι προσαρμοσμένα.

Ένα μειονέκτημα της παραπάνω προσέγγισης είναι ότι εμπεριέχει ασάφεια σχετικά με τους αναλογικούς κινδύνους (proportional hazards) ανάμεσα στον υπό μελέτη πληθυσμό και τον εξωτερικά καταγεγραμμένο πληθυσμό για όλα τα ηλικιακά στρώματα και τα στρώματα των χρονικών περιόδων, πράγμα το οποίο σημαίνει, ότι οι αναφερόμενοι δείκτες θανάτου εφαρμόζονται στον υπό μελέτη πληθυσμό τουλάχιστον για μία σταθερή αναλογία. Η απλούστερη εναλλακτική λύση που δεν εξαρτάται από την παραπάνω ασάφεια είναι να δουλέψουμε ολοκληρωτικά πάνω στην βάση δεδομένων χωρίς να κάνουμε οποιαδήποτε αναφορά στον εξωτερικό πληθυσμό. Οι δείκτες θανάτου υπολογίζονται για κάθε κελί ηλικίας-χρονικής περιόδου από τα δεδομένα ολόκληρης της ομάδας. Αυτοί οι ενδιάμεσοι δείκτες χρησιμοποιούνται τότε για να υπολογίσουν το αναμενόμενο πλήθος θανάτων για τις υποομάδες που καθορίζονται από τις συνδιακυμάνσεις, κι έτσι τα SMRs που παράγονται για κάθε υποομάδα βασίζονται στους παρατηρούμενους δείκτες θανάτου του συνολικού πληθυσμού σε κάθε κελί, αντί των εξωτερικών δεικτών θανάτου (Breslow & Day, 1987). Αν υπάρχουν περισσότερες από δύο υποομάδες, η σύγκριση των ενδιάμεσων SMRs εξαρτάται ακόμα από τους αναλογικούς κινδύνους των υποομάδων για κάθε ηλικία-χρονική περίοδο, αλλά δεν είναι πλέον απαραίτητες οι ασάφειες αναλογικών κινδύνων ανάμεσα στον υπό μελέτη πληθυσμό και σε έναν αναφορικό πληθυσμό. Η προσέγγιση αυτή είναι γνωστό ότι είναι συντηρητική και η βελτιωμένη έκθεση της προσέγγισης αυτής δόθηκε από τους Mantel-Haenszel (Breslow, 1984). Η μέθοδος αυτή αναφέρεται ως *interval standardization* (ενδιάμεση τυποποίηση).

Μια δεύτερη προσέγγιση είναι να χρησιμοποιήσουμε απευθείας την προτυποποίηση με μία ενδιάμεση υποομάδα ως σταθερά. Μια κατάλληλη επιλογή για την ενδιάμεση αναφορά ίσως είναι μία μη εκτιθέμενη ομάδα ή η λιγότερο εκτιθέμενη υποομάδα στην μελέτη. Η μέθοδος αυτή παράγει τυποποιημένους δείκτες αναλογίας (SRRs-Standardized rate ratios) οι οποίοι είναι δείκτες ενός άμεσα τυποποιημένου δείκτη για κάθε υποομάδα προς τον δείκτη της σταθερής ομάδας αναφοράς. Η μέθοδος αποφεύγει τα πιθανά προβλήματα που συνδέονται με τις συγκρίσεις των SMRs σε περισσότερες από δύο ομάδες, αλλά τα SRRs έχουν λιγότερη ακρίβεια ως προς την εκτίμηση από τα SMRs αν οι υποομάδες περιέχουν κελιά με λίγους θανάτους.

Η πιο εκτενής προσέγγιση είναι τελικά αυτή της μοντελοποίησης Poisson. Για τις υποομάδες i και τα στρώματα ηλικίας-χρονικής περιόδου j , αν n_{ij} , είναι το πλήθος των ατόμων κατά ηλικία και γ_{ij} ο δείκτης θανάτου τότε, το μ_{ij} , που είναι ο αναμενόμενος αριθμός θανάτων δίνεται από την σχέση

$$\ln \mu_{ij} = \ln m_{ij} + \ln \gamma_{ij}$$

Αν το γ_{ij} μοντελοποιηθεί υπό τους όρους των στρωμάτων ηλικίας-χρονικής περιόδου και ένα σύνολο συνδιακυμάνσεων τότε προκύπτουν οι παρακάτω τύποι

$$\ln \gamma_{ij} = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \dots + \beta_p x_{ijp}$$

και

$$\ln \mu_{ij} = \ln m_{ij} + \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \dots + \beta_p x_{ijp} \quad (4.28)$$

Αυτό είναι ένα γενικευμένο γραμμικό μοντέλο μιας μεταβλητής Poisson, με επιπρόσθετο όρο $\ln m_{ij}$.

Σε πολλές περιπτώσεις η εξωτερική μέθοδος (4.27) και η εσωτερική μέθοδος (4.28) θα δώσουν παρόμοια συμπεράσματα στην επίδραση των συνμεταβλητών. Το τελευταίο έχει το πλεονέκτημα ότι δεν υπάρχει ασάφεια και ότι οι δείκτες θανάτου σε όλα τα στρώματα ηλικίας-χρονικών περιόδων ακολουθούν ένα μοντέλο αναλογικού κινδύνου με εκτίμηση του πληθυσμού αναφοράς, αλλά με την ακρίβεια των συγκρίσεων να είναι ελαφρώς κατώτερη της εξωτερικής μεθόδου. Ένας συνδυασμός των δύο προσεγγίσεων, χρησιμοποιώντας τη μέθοδο (4.27) για την συνολική ομάδα και τις βασικές υποομάδες και την μέθοδο (4.28) για την μοντελοποίηση παλινδρόμησης, έχει το πλεονέκτημα να εκτιμάει τις επιδράσεις των συνδιακυμάνσεων και επίσης να εκτιμάει το πώς η θνησιμότητα σε ολόκληρη την ομάδα συγκρίνεται με εκείνη του πληθυσμού του οποίου αποτελεί μέρος. Ένα παράδειγμα δόθηκε από τον Checkoway (1993).

4.8 Ανάλυση ηλικίας-περιόδου-ομάδας (Age-Period-Cohort analysis)

Στο κεφάλαιο 2 (§2.2) συζητήθηκε το life-table και καταγράφηκαν δύο τύποι, το current life-table και το cohort life-table. Ο πρώτος τύπος βασίζεται σε μία συγκεκριμένη χρονική περίοδο (πχ. 1930-32) για άτομα διαφορετικών ηλικιών, ενώ ο δεύτερος τύπος βασίζεται σε δείκτες θανάτου ατόμων που γεννήθηκαν μία συγκεκριμένη χρονική περίοδο (πχ. 1929-1933). Υπάρχουν τρεις μεταβλητές που σχετίζονται με τον χρόνο και ίσως επηρεάζουν τους δείκτες θανάτου

1.Ηλικία (age)- οι δείκτες θανάτου ξεκάθαρα σχετίζονται με την ηλικία.

2.Ημερολογιακή περίοδος(calendar period)-ο ειδικός ηλικιακός δείκτης θανάτου ίσως αλλάζει με την πάροδο του χρόνου.

3.Ομάδα (cohort)- μία ομάδα που γεννήθηκε σε μία συγκεκριμένη χρονική στιγμή (δηλαδή μια ομάδα cohort) , ίσως έχει διαφορετικό ηλικιακό δείκτη θανάτου από όσους γεννήθηκαν σε μία άλλη χρονική στιγμή.

Οι επιδράσεις των τριών μεταβλητών αναφέρονται σαν *age*(ηλικιακές),*period*(περιοδικές) και *cohort*(ομαδικές) επιδράσεις, αντίστοιχα. Αν και όλα τα παραπάνω αναφέρονται σε δείκτες θανάτου, τα συμπεράσματα μπορεί να αφορούν οποιαδήποτε μεταβλητή σχετιζόμενη με την υγεία, όπως οι δείκτες ασθένειας, ή η αρτηριακή πίεση ή η αναπνευστική λειτουργία. Οι επιδράσεις *period* θα μπορούσαν να προκύψουν αν μια αλλαγή λάβει χώρα σε συγκεκριμένη χρονική στιγμή και εφαρμοστεί για όλες τις ομάδες. Για παράδειγμα αν υιοθετούσαμε μια νέα θεραπεία για να βελτιώσει την πρόγνωση για αυτούς που προσβάλλονται από μια συγκεκριμένη νόσο με υψηλό δείκτη θνησιμότητας , τότε αυτή θα εμφανιζόταν ως μία *period* επίδραση για τον δείκτη θανάτου εξαιτίας της ασθένειας αυτής. Οι επιδράσεις *cohort* μερικές φορές αναφέρονται ως *γενικευμένες* επιδράσεις. Οι επιδράσεις αυτές συνδέονται με γεγονότα που επηρεάζουν τους ανθρώπους που γεννήθηκαν μία συγκεκριμένη χρονική στιγμή. Για παράδειγμα, το ύψος του ενήλικα επηρεάζεται από την διατροφή στα πρώτα χρόνια της ζωής του. Έτσι όσοι γεννήθηκαν σε περιόδους κατοχής θα τρέφονταν με λιγότερη ποσότητα σε σχέση με όσους γεννήθηκαν σε περιόδους ειρήνης. Επίσης, η αποδοχή του καπνίσματος τείνει να έχει μία γενικευμένη επίδραση και αυξάνει την επίδραση *cohort* στον καρκίνο του πνεύμονα σε όποιες άλλες ασθένειες συνδέονται με το κάπνισμα.

Η ανάλυση ενός συμπεράσματος σε σχέση με όλες αυτές τις μεταβλητές καλείται *age-period-cohort analysis* (ανάλυση κατά ηλικία-περίοδο-ημερολογιακή ομάδα). Τέτοιες αναλύσεις είναι χρήσιμες στον καθορισμό τάσεων που χρησιμοποιούνται κυρίως για πρόβλεψη , και για αξιολόγηση των επιδράσεων των μέτρων που υιοθετήθηκαν για να βελτιώσουν την υγεία ή για να βοηθήσουν κάποιον να αλλάξει συνήθειες καταστροφικές για την υγεία του. Για παράδειγμα, τίθεται το ερώτημα: Υπάρχει ελάττωση των ασθενών με καρκίνο του πνεύμονα που ίσως σχετίζεται με την καμπάνια υγείας που έχει ως στόχο να πείσει τους καπνιστές να σταματήσουν το τσιγάρο και σε ποιο βαθμό αντιστοιχεί η μείωση των ασθενών με καρκίνο του πνεύμονα με την μείωση του καπνίσματος;

Υποθέτουμε ότι υπάρχουν δεδομένα για ένα αποτέλεσμα, που αφορούν άτομα σε διαφορετικές ομάδες *cohort* ,σε διαφορετικές χρονικές περιόδους και για διαφορετικό σφάλμα ηλικιών. Για ένα άτομο συμβολίζουμε την ημερομηνία γέννησης με μία μεταβλητή *cohort*, *c* , τον χρόνο παρατήρησης από την μεταβλητή *period* , *p* , και την ηλικία με μία μεταβλητή *age* , *a* . Τότε οι τρεις μεταβλητές αυτές συνδέονται με την σχέση

$$a=p-c \quad (4.29)$$

Τώρα υποθέτουμε, για ένα συμπέρασμα y , ότι ένα γραμμικό μοντέλο εφαρμόστηκε στον τύπο

$$E(y) = \beta_0 + \beta_a a + \beta_p p + \beta_c c \quad (4.30)$$

Τότε, όταν εφαρμόσουμε το μοντέλο αυτό στα δεδομένα θα υπάρχει ένα πρόβλημα εξαιτίας της συγγραμμικότητας. Στην πραγματικότητα, καθώς η (4.29) αντιπροσωπεύει ακριβώς την συγγραμμικότητα δεν θα υπάρχει μόνο ένας τύπος παλινδρόμησης. Αν η μία λύση είναι

$$Y_1 = b_0 + b_a a + b_p p + b_c c \quad (4.31)$$

τότε

$$Y_2 = b_0 + b'_a a + b'_p p + b'_c c \quad (4.32)$$

Ισοδύναμα, για οποιαδήποτε σταθερά u , έχουμε

$$\begin{aligned} b'_a &= b_a + u \\ b'_p &= b_p + u \\ b'_c &= b_c + u \end{aligned} \quad (4.33)$$

Αφού για οποιοδήποτε δεδομένο η διαφορά των δύο τιμών Y_1, Y_2 είναι

$$Y_2 - Y_1 = u(a - p + c)$$

θα είναι μηδέν εξαιτίας της ταυτότητας (4.29). Για παράδειγμα οι τιμές για οποιανδήποτε θεμιτή *age-period-cohort* θα είναι ίδιες. Δεν υπάρχει παρόλα αυτά, μοναδική καλύτερα εφαρμοσμένη γραμμή παλινδρόμησης. Αυτό είναι γνωστό ως πρόβλημα αναγνωρισιμότητας (*identifiability problem*).

Πρακτικά η (4.30) δεν εφαρμόζεται ακριβώς εξαιτίας της έλλειψης ακρίβειας με την οποία γίνεται η καταγραφή των μεταβλητών. Αν κάθε μεταβλητή εφαρμόζεται σαν ένας ακέραιος αριθμός, τότε κάποιος που έχει γεννηθεί το 1937 και καταγράφηκε το 2002 θα είναι 62 ή 63 ετών, σύμφωνα με το αν η μελέτη ξεκίνησε πριν ή μετά την ημερομηνία γέννησης του. Μερικές φορές, συγκεκριμένα σε μελέτες των δεικτών θανάτου, οι μελέτες καταγράφονται σε ομάδες εύρους 5 ετών και η ομάδα cohort 1935-39 θα αλληλεπικαλύψει τις ηλικιακές ομάδες 55-59 και 60-64 κατά την διάρκεια της περιόδου 1995-99. Το πρόβλημα αναγνωρισιμότητας, παρόλα αυτά, παραμένει γιατί η συγγραμμικότητα παρουσιάζει έλλειψη ακρίβειας.

Με μία πρώτη ματιά, ίσως εμφανίζεται ότι, αν δεν υπάρχει μοναδική παλινδρόμηση, δεν υπάρχουν και μοναδικές λύσεις για οποιοδήποτε ποσότητες οι οποίες εκτιμώνται με την παλινδρόμηση. Ευτυχώς, δεν είμαστε σε αυτή την περίπτωση, και υπάρχει ένα μεγάλο εύρος καταστάσεων που η εφαρμοσμένη παλινδρόμηση, παρά την έλλειψη μοναδικότητας δίνει μοναδική λύση. Η απλούστερη μορφή όλων των

παραπάνω είναι μία πρόβλεψη που χρησιμοποιεί τις εφαρμοσμένες τάσεις. Υποθέτουμε ότι τα δεδομένα είναι διαθέσιμα σε ένα έτος που συμβολίζεται με $p_0 = p$ και χρειάζεται να κάνουμε πρόβλεψη για το επόμενο έτος $p = p_0 + 1$. Όσοι το έτος αυτό έχουν ηλικία a θα ανήκουν στην ομάδα που ορίζεται από το $c = p_0 + 1 - a$. Τότε χρησιμοποιώντας τον τύπο (4.32) καταλήγουμε στην σχέση

$$\begin{aligned} Y &= b_0 + b'_a a + b'_p (p_0 + 1) + b'_c (p_0 + 1 - a) \\ &= b_0 + (b'_a - b'_c) a + (b'_p - b'_c) p_0 + (b'_p - b'_c) \end{aligned}$$

Στην δεύτερη γραμμή όλοι οι όροι της παρένθεσης προέκυψαν από τον τύπο (4.33) και είναι ανεξάρτητοι του u . Όπως προκύπτει, η προβλεπόμενη τιμή δεν βασίζεται στην αυθαίρετη σταθερά u , και είναι η μοναδική εκτιμήτρια.

Το πρόβλημα αναγνωρισιμότητας έχει επανεμφανιστεί εξαιτίας της (4.29). Αν η εφαρμοσμένη παλινδρόμηση περιέχει μη γραμμικούς όρους για την ηλικία, την περίοδο ή την ομάδα cohort τότε οι όροι αυτοί δεν θα είναι ακριβώς συγγραμμικοί.

Υποθέτουμε ότι εισάγουμε κάποια αλλαγή στο έτος p_0 , και διατυπώνεται η θέση ότι αυτό έχει προκαλέσει άμεση επίπτωση σε όλες τις ηλικίες και όλα τα έτη που ακολουθούν. Αυτό σημαίνει ότι,

$$E(y) = \beta_0 + \beta_a a + \beta_p p + \beta_c c \text{ για } p \leq p_0$$

$$E(y) = \beta_0 + \beta_a a + \beta_p p + \beta_c c + \beta_e \text{ για } p > p_0$$

όπου το β_e είναι η επίδραση εξαιτίας της αλλαγής. Ύστερα, οι δύο αυτές εξισώσεις μπορεί να συνδυαστούν από τον καθορισμό της ψεύτικης μεταβλητής, e , που παίρνει την τιμή 0 για $p \leq p_0$ και την τιμή 1 για $p > p_0$. Δηλαδή προκύπτει ο τύπος

$$E(y) = \beta_0 + \beta_a a + \beta_p p + \beta_c c + \beta_e e$$

Η μεταβλητή e δεν είναι συγγραμμική με οποιονδήποτε συνδυασμό των a , p και c , έτσι το β_e είναι εκτιμήσιμο.

Η παραπάνω συζήτηση είναι σκόπιμα απλοποιημένη με στόχο να επεξηγήσει το πρόβλημα της αναγνωρισιμότητας αλλά σε κάποιες περιπτώσεις με τον τρόπο αυτό δεν γίνεται να βγάλει συμπεράσματα χρησιμοποιώντας κατά τον βέλτιστο δυνατό τρόπο τα δεδομένα. Πρακτικά, θα ήταν σπάνιο να εφαρμόσουμε ένα απλό γραμμικό μοντέλο τύπου (4.30). Αν το αποτέλεσμα ήταν ένας δείκτης θανάτου, τότε η σχέση με την ηλικία θα απείχε πολύ από την συγγραμμικότητα και η ανάλυση ίσως να γινόταν με την βοήθεια του $\ln y$, ή οι ηλικιακές ομάδες ίσως εκπροσωπούσαν από ένα είδος ψευδομεταβλητών. Επιπλέον θα ήταν μάλλον μη αναμενόμενο ότι η επίδραση οποιασδήποτε αλλαγής θα ήταν τόσο απλή όσο αυτή που θεωρήσαμε.

4.9 Το μέτρο Κάπα της συμφωνίας(Kappa measure of agreement)

Όταν μια κατηγορική μεταβλητή είναι δύσκολο να καταγραφεί αντικειμενικά , είναι κοινή πρακτική να χρησιμοποιούνται περισσότεροι από δύο εκτιμητές για να αξιολογηθεί η μεταβλητή αυτή, και χρησιμοποιείται ο μέσος ή η διάμεσος ως εκτιμητές για την βασική ανάλυση δεδομένων. Είναι επομένως λογικό να επιβεβαιώσουμε τον υπολογισμό οποιουδήποτε είδους μέσης τιμής που οι εκτιμητές χρησιμοποίησαν σε παρόμοιες καταστάσεις , αφού σε αντίθετη περίπτωση δεν θα ήταν ξεκάθαρο τι αντιπροσωπεύουν οι μέσες τιμές. Παραδείγματα είναι η ταξινόμηση μιας ασθένειας σε κλίμακα με δύο ή περισσότερες κατηγορίες βασισμένες στην εκτίμηση των ενδείξεων και των συμπτωμάτων που εμφανίζονται , και η ανάγνωση των ακτινογραφιών θώρακος.

Θεωρούμε αρχικά τη περίπτωση όπου υπάρχουν δύο κατηγορίες , παρουσία και απουσία μια συνθήκης , και υποθέτουμε ότι δύο εκτιμητές αξιολογούν n ανεξάρτητα άτομα. Τότε τα αποτελέσματα μπορούν να εμφανιστούν σε έναν 2×2 πίνακα όπως ο ακόλουθος:

Εκτιμητής 1	Εκτιμητής 2		
	Παρουσία	Απουσία	
Παρουσία	a	b	$a+b$
Απουσία	c	d	$c+d$
	$a+c$	$b+d$	n

Ένας φυσικός τρόπος να αξιολογήσουμε την συμφωνία μεταξύ των εκτιμητών θα ήταν να δηλώσουμε ότι οι εκτιμητές συμφωνούν στο άθροισμα των ατόμων $a+b$ και να εκφράσουμε το άθροισμα αυτό λαμβάνοντας υπόψη το πλήθος των ασθενών, δηλαδή $(a+b)/n$. Στα παρακάτω παραδείγματα γίνεται φανερό το μειονέκτημα της απλής αυτής προσέγγισης .

A	Εκτιμητής 2			B	Εκτιμητής 2		
	Εκτιμητής 1	Παρουσία	Απουσία		Εκτιμητής 1	Παρουσία	Απουσία
Παρουσία	5	5	10	Παρουσία	35	5	40
Απουσία	5	85	90	Απουσία	5	55	60
	10	90	100		40	60	100

Και στα δύο αυτά παραδείγματα οι εκτιμητές συμφωνούν στο 90% των ατόμων, αλλά το μειονέκτημα εμφανίζεται περισσότερο στο παράδειγμα A σε σχέση με το παράδειγμα B αφού, στο πρώτο υπάρχει συμφωνία ότι η συνθήκη εκπροσωπείται μόνο για 5 από τα 15 άτομα τα οποία αξιολογήθηκαν τουλάχιστον από τον ένα εκτιμητή για το αν έχουν ή όχι τη συνθήκη, ενώ για το τελευταίο παράδειγμα συμφωνήθηκε η συνθήκη να εκπροσωπείται για 35 από τα 45 άτομα που αξιολογήθηκαν για το αν έχουν την συνθήκη από τουλάχιστον έναν εκτιμητή.

Η δυσκολία παρουσιάζεται γιατί δεν λάβαμε στις συμφωνίες καθόλου υπόψη ότι θα μπορούσε να προκύψει κάποια αλλαγή. Αν δεν υπάρχει πραγματική συμφωνία μεταξύ των εκτιμητών, τότε οι αναμενόμενες συχνότητες στα κελιά του πίνακα 2x2 θα μπορούσαν να καθοριστούν στην τύχη. Για το παράδειγμα A, θα αναμενόταν ότι θα μπορούσαν να είναι 81 άτομα που αξιολογήθηκαν αρνητικά και από τις δύο εκτιμητρίες, και ένα άτομο που αξιολογήθηκε θετικά και από τις δύο εκτιμητρίες, όπου αυτοί οι αναμενόμενοι αριθμοί υπολογίστηκαν ακριβώς από τις ίδιες μεθόδους όπως αυτές που χρησιμοποιήθηκαν για να υπολογίσουν τις αναμενόμενες συχνότητες σε ένα χ^2 τεστ. Η παρατηρούμενη συμφωνία για τα 90 άτομα δεν φαίνεται τόσο εντυπωσιακή όταν συγκρίνεται με μία μη αναμενόμενη συμφωνία για 82 ασθενείς. Στο παράδειγμα B, ο αναμενόμενος αριθμός των συμφωνιών είναι μόνο 52 και ο παρατηρούμενος αριθμός είναι πολύ μεγαλύτερος από αυτόν.

Η παραπάνω λογική είναι η βασική λογική του μέτρου $kappa(k)$ της συμφωνίας που εισήχθηκε πρώτη φορά από τον Cohen (1960). Η μέθοδος είναι ως εξής. Αρχικά καθορίζουμε το I_0 να είναι το παρατηρούμενο μέτρο της συμφωνίας, και το I_e να είναι το αντίστοιχο αναμενόμενο μέτρο.

Τότε

$$I_0 = \frac{a+d}{n}$$

$$I_e = \frac{(a+d)(a+b) + (b+d)(c+d)}{n^2}$$

Δεύτερον το κ καθορίζεται ως η διαφορά μεταξύ παρατηρούμενων και αναμενόμενων συμφωνιών, που εκφράζονται σαν ένα κλάσμα της μέγιστης διαφοράς. Αφού η μέγιστη τιμή του I_o είναι 1, αυτό δίνει

$$\kappa = \frac{I_o - I_e}{1 - I_e} \quad (4.38)$$

Για τα παραδείγματα Α και Β οι τιμές του κ είναι 0.44 και 0.79, αντίστοιχα με την μέγιστη τιμή του κ να είναι 1, όταν αντιπροσωπεύει μια τέλεια συμφωνία, και να είναι 0 στην περίπτωση που η συμφωνία είναι τυχαία. Αν και οι αρνητικές τιμές είναι μαθηματικά πιθανές, αντιπροσωπεύοντας την αντίθετη άποψη, κάτι τέτοιο είναι απίθανο να συμβεί στην πράξη.

Παράδειγμα 7:

Σε μία μελέτη του Bergen (1992), η όψιμη δυσκινησία (νευρολογική πάθηση που αντανακλάται σε ακούσιες κινήσεις του προσώπου και της γνάθου) αξιολογήθηκε από δύο εκτιμητές βαθμολογώντας επτά είδη σε μία κλίμακα 5 μονάδων, κωδικοποιημένες από το 0 έως το 4. Ένας καθορισμός της όψιμης δυσκινησίας θα είναι είτε όταν δύο βαθμοί είναι τουλάχιστον 2 είτε όταν ένας είναι τουλάχιστον 3. Για μία σειρά αξιολογήσεων υπήρχαν 168 άτομα και η συμφωνία ήταν η εξής.

Εκτιμητής 2	Εκτιμητής 1		
	Παρουσία	Απουσία	
Παρουσία	123	10	133
Απουσία	6	29	35
	129	39	168

Τότε προκύπτουν,

$$I_o = \frac{152}{168} = 0.905$$

$$I_e = \frac{129 \times 133 + 39 \times 35}{168^2} = 0.656$$

$$k = \frac{0.905 - 0.656}{1 - 0.656} = 0.72$$

Η συμφωνία μεταξύ των δύο εκτιμητών είναι καλή , κι έτσι οι δύο αξιολογήσεις συνδυάζονται και καταλήγουν στην εμφάνιση της όψιμης δυσκινησίας για τους ασθενείς που έχουν διαγνωστεί θετικοί στην ασθένεια και από τους δύο εκτιμητές.

Έμμεση μέθοδος

Όταν υπάρχουν περισσότερες από δύο κατηγορίες και οι κατηγορίες διατάσσονται σε σειρά, μία διαφορά ανάμεσα στους εκτιμητές της μία κατηγορίας είναι μικρότερο πρόβλημα από μία διαφορά ανάμεσα σε δύο κατηγορίες, και μία διαφορά τριών κατηγοριών θα μπορούσε να δημιουργήσει ακόμα μεγαλύτερη διαφωνία κ.ο.κ. Είναι ξεκάθαρα επιθυμητό ότι θέλουμε να ενσωματώσουμε την λογική αυτή σε ένα μέτρο συμφωνίας κάτι που επιτυγχάνεται χρησιμοποιώντας το σταθμισμένο κάπα μέτρο-*weighted kappa* (Cohen, 1968). Το μέτρο αυτό υπολογίζεται συμβολίζοντας τον συντελεστή βαρύτητας, w_i , για τα άτομα των οποίων οι εκτιμητές από i κατηγορίες διαφέρουν. Όταν οι εκτιμητές συμφωνούν, ο συντελεστής βαρύτητας είναι 1, δηλαδή $w_0=1$. Αν υπάρχουν k κατηγορίες τότε η μέγιστη διαφωνία είναι $k-1$ κατηγορίες και ο αντίστοιχος συντελεστής είναι ίσος με 0, δηλαδή $w_{k-1}=0$. Το πιο κοινό σύνολο συντελεστών για όλες τις ενδιάμεσες τιμές που είναι εξίσου κατανεμημένες δίνεται από τον τύπο

$$w_i = 1 - \frac{i}{k-1} \quad (4.39)$$

Επιπλέον ο παρατηρούμενος δείκτης συμφωνίας ορίζεται ως

$$I_0 = \frac{\sum w_i r_i}{n}$$

Όπου το r_i είναι ένα πλήθος ατόμων για τα οποία οι εκτιμητές από i -κατηγορίες διαφέρουν. Η αναμενόμενη τιμή υπολογίζεται παρόμοια, χρησιμοποιώντας τις αναμενόμενες συχνότητες, και ο σταθμισμένος *kappa* k εκτιμάται από τον τύπο (4.38)

BIBΛΙΟΓΡΑΦΙΑ

- [1] Abdelbasit K.M and Plackett, R.L. (1983). Experimental Design for binary data. *J. Am. Stat. Ass.* 78, 90-98
- [2] Altman D.G. (1991) *Practical Statistics for Medical Research*. Chapman and Hall, London
- [3] Armitage P. (1999) Data and safety monitoring in the Concorde and Alpha trials. *Cont. Clin. Trials*, 207-230
- [4] Armitage P. , Berry G. and Matthews J.N.S (2002) *Statistical Methods in Medical research*, Blackwell, USA
- [5] Bailey N.T.J (1975) *The Mathematical Theory of Infectious Diseases and its Applications*, 2nd edn. Griffin, London.
- [6] Barnett V. and Lewis T. (1994) *Outliers in Statistical Data*, 3rd edn. Wiley, Chichester.
- [7] Becker N.G. (1989) *The Analysis of Infectious Disease Data*. Chapman and Hall, London.
- [8] Benichou J. (1998) Attributable Risk. In *Encyclopedia of Biostatistics* , eds. P. Armitage and T.Colton pp.216-229. Wiley, Chichester.
- [9] Breslow N (1990) Biostatistics and Bayes. *Stat. Sci.*, 269-298
- [10] Collett D. (1994) *Modelling Survival Data in Medical Research*. Chapman and Hall, London.
- [11] Cox D.R (1958) Regression Model and life-tables. *J.R Stat. Soc.*, 187-220
- [12] Dobson A.J. (1990) *An Introduction to Generalized Linear Models*. Chapman and Hall, London.
- [13] Edgington E.S. (1987) *Randomization Tests*. Marcel Dekker, New York
- [14] Etzioni R.D. and Weiss N.S. (1998) Analysis of case-control studies of screening: impact of misspecifying the duration of detectable preclinical pathological changes. *Am. J. Epidemiol.*, 292-297
- [15] Fleming T.R., DeGruetola V. and DeMets D.L. (1998) Surrogate endpoints. In *Encyclopedia of Biostatistics*, eds P. Armitage and

T. Colton, pp.4425-4431. Wiley, Chichester.

- [16] Gehan E. (1965) A generalized Wilcoxon test for comparing arbitrarily singly censored samples. *Biometrika* **52**, 203-223.
- [17] Geigy Scientific Tables (1982) Vol. 2: *Introduction to Statistics, Statistical Tables, Mathematical Formulae*, 8th edn. Giba-Geisy, Basle.
- [18] Gilks W.R., Richardson S. and Spiegelhalter D.J. (1996) *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.
- [19] Hennekens C.H., Buring J.E. and Mayrent S.L (eds) (1987) *Epidemiology in Medicine*. Little , Brown, Boston.
- [20] Hsu J.C (1996) *Multiple Comparisons: Theory and Methods*. Chapman and Hall, London.
- [21] Kleinbaum D.G. (1996) *Survival Analysis: A Self-Learning Text*. Springer-Verlag, New York.
- [22] Pauker S.G. and Kassirer J.P (1992) Decision Analysis. In *Medical Uses of Statistics*, 2nd edn. eds J.C. Bailar III and F. Mosteller, pp.158-179. NEJM Books, Boston.
- [23] Pocock S. J. (1983) *Clinical Trials: A Practical Approach*. Wiley, Chichester.
- [24] Ratkowsky D.A., Evans M.A. and Alldredge J.R. (1993) *Cross Over Experiments: Design, Analysis and Application*. Dekker, New York.
- [25] Senn S. (1993) *Cross-Over Trials in Clinical Research*. Wiley, Chichester.
- [26] Τριχόπουλος Δ. και Παγώνας Δ. (2002). *Γενική και Κλινική Επιδημιολογία*. Παρισιάνου, Αθήνα
- [27] Yates f. (1981) *Sampling Methods for Censuses and Surveys* , 4th edn. Griffin, London.