

Εθνικό Μετσόβιο Πολυτεχνείο



Μεταπτυχιακή Διπλωματική Εργασία

Προσεγγίσεις στην Ανάλυση Επαναλαμβανόμενων

Γεγονότων

Ρούφης Ευστάθιος

Επιτροπή Καθηγητών

- 1. Χ. Κουκουβίνος, Καθηγητής*
- 2. Χ. Καρώνη (Επιβλέπουσα), Αναπλ. Καθηγήτρια*
- 3. Ι. Βόντα, Επίκ. Καθηγήτρια*

*Τομέας Μαθηματικών
Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών
Αθήνα, 2012*

Πριν αρχίσει η παρουσίαση της διπλωματικής αυτής εργασίας, θα ήθελα να ευχαριστήσω την Δρα. Χρυσή Καρώνη, Αναπληρώτρια Καθηγήτρια του Τομέα Μαθηματικών της Σχολής Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών του Εθνικού Μετσόβιου Πολυτεχνείου, που μου επέτρεψε να ασχοληθώ με το συγκεκριμένο θέμα. Την ευχαριστώ θερμά για τις πολύτιμες παρατηρήσεις, συμβουλές αλλά και την πολύ σημαντική βοήθεια που μου προσέφερε ώστε να μπορέσει να ολοκληρωθεί η παρούσα διπλωματική εργασία στο σύντομο διάστημα που είχαμε στην διάθεση μας.

Επίσης, θα ήθελα να ευχαριστήσω τους συμφοιτητές και την οικογενειά μου για την αμέριστη συμπαράστασή τους καθ' όλη τη διάρκεια της φοιτησής μου στο Εθνικό Μετσόβιο Πολυτεχνείο.

Αθήνα 2012

Ρούφης Ευσταθιος

Περιεχόμενα

Περίληψη 7

Abstract 9

Κεφάλαιο 1: Επαναλαμβανόμενα Γεγονότα 11

1.1 Εισαγωγή..... 11

1.2 Μερικές μέθοδοι που βασίζονται σε απαριθμητά δεδομένα..... 18

1.2.1 Poisson μοντέλα και μικτά Poisson μοντέλα..... 18

1.2.2 Εύρωστες μέθοδοι για ποσοστιαίες και μέσες συναρτήσεις..... 21

1.2.3 Ανάλυση ενδιάμεσων χρόνων..... 24

Κεφάλαιο 2: Μοντέλα Επαναλαμβανόμενων

Γεγονότων και άλλες χρήσιμες έννοιες 27

2.1 Βασικές συνιστώσες μοντέλου 27

2.1.1 Διαστήματα Διακινδύνευσης 29

2.1.2 Βασική Συνάρτηση Διακινδύνευσης 31

2.1.3 Σύνολο που είναι σε κίνδυνο 32

2.1.4 Διαχείριση της Συσχέτισης των Γεγονότων 34

2.2 Μοντέλο Andersen-Gill..... 35

2.3 Μοντέλο PWP 39

2.4 Μοντέλο WLW..... 43

2.5 Μοντέλο των Pepe και Cai..... 47

2.6 Επαναλαμβανόμενα Γεγονότα: Χαρακτηριστικά και Μοντέλα..... 48

2.7 Το υπό συνθήκη μοντέλο ευπάθειας 51

<u>Κεφάλαιο 3: Στατιστική Ανάλυση.....</u>	<u>53</u>
3.1 Καρκίνος της ουροδόχου κύστης	53
3.2 rhDNase σε ασθενείς με κυστική ίνωση	65
3.3 rIFN-g σε ασθενείς με χρόνια κοκκιοματώδη ασθένεια.....	77
<u>Συμπεράσματα.....</u>	<u>91</u>
<u>Βιβλιογραφία.....</u>	<u>93</u>

Περίληψη

Σε αρκετές μελέτες το γεγονός που βιώνει ένα άτομο μπορεί να εμφανίζεται περισσότερες από μια φορές στο άτομο που μελετάται. Τέτοια γεγονότα ονομάζονται επαναλαμβανόμενα γεγονότα. Δεδομένα επαναλαμβανόμενων γεγονότων γενικά εμφανίζονται στους κλάδους της βιοϊατρικής, αλλά και στον έλεγχο ποιότητας των υλικών. Συνήθως η μελέτη μας έχει έναν καθορισμένο χρόνο λήξης. Η δομή των δεδομένων των επαναλαμβανόμενων γεγονότων αποτελεί μια ειδική περίπτωση των πολυμεταβλητών δεδομένων επιβίωσης, όπου οι χρόνοι αποτυχίας για ένα άτομο είναι διατεταγμένοι. Ως εκ τούτου, τα επαναλαμβανόμενα γεγονότα συχνά αναλύονται χρησιμοποιώντας μεθόδους της πολυμεταβλητής ανάλυσης επιβίωσης. Αυτές οι μέθοδοι που θα συζητηθούν και θα αναλυθούν στην παρούσα διπλωματική εργασία είναι η μέθοδος των Andersen και Gill, η μέθοδος των Wei, Lin και Weissfeld και η μέθοδος των Prentice, Williams και Peterson.

Στην παρούσα λοιπόν διπλωματική εργασία θα αναλυθούν οι συγκεκριμένες μέθοδοι και θα δοθεί και μια έμφαση στο ρόλο που παίζουν τα μοντέλα ευπάθειας στα επαναλαμβανόμενα γεγονότα. Στο τελευταίο Κεφάλαιο θα χρησιμοποιηθούν τρία σύνολα δεδομένων όπου με τη βοήθεια της R αλλά και των τριών μεθόδων ανάλυσης, θα γίνει η ανάλυση τους και θα συζητηθούν τα αποτελέσματα που θα προκύψουν. Τα τρία σύνολα δεδομένων που θα χρησιμοποιηθούν είναι "ο καρκίνος της ουροδόχου κύστης", "rhdNase σε ασθενείς με κυστική ίνωση" και "rIFN-g σε ασθενείς με χρόνια κοκκιωματώδη ασθένεια".

Abstract

In many research settings, the event of interest can be experienced more than once per subject. Such outcomes have been termed recurrent events. Recurrent event data are often encountered in biomedicine and quality control. Typically, the study will have a fixed termination date. The data structure for recurrent events represents a special case of multivariate survival data, where the failure times for a subject are ordered. As such, recurrent event data have often been analyzed using methods of multivariate survival analysis. The methods that will be discussed and analyzed in this dissertation are the method of Andersen and Gill, the method of Wei, Lin and Weissfeld and the method of Prentice, Williams and Peterson.

In this dissertation we will analyze these methods and we will emphasize the role that frailty models play on recurrent events. In the last Chapter we will use three data sets which will be analyzed with the help of R and the three methods and their results will be discussed. The three data sets which will be used are "bladder cancer", "rhdNase in patients with cystic fibrosis " and "rIFN-g in patients with chronic granulomatous disease."

Κεφάλαιο 1: Επαναλαμβανόμενα Γεγονότα

1.1 Εισαγωγή

Πολλές ιατρικές μελέτες περιλαμβάνουν καταστάσεις στις οποίες τα αντικείμενα προς μελέτη βιώνουν επαναλαμβανόμενα γεγονότα. Για παράδειγμα, οι ασθενείς που κινδυνεύουν να βιώσουν κάποιο αγγειοεγκεφαλικό επεισόδιο μπορεί να βιώσουν κάποιο παροδικό ισχαιμικό επεισόδιο, οι ασθενείς με καρκίνο μπορεί να βιώσουν επαναλαμβανόμενους επιφανειακούς όγκους και τέλος ασθενείς με ψυχικές διαταραχές μπορεί να βιώσουν επαναλαμβανόμενες φορές τη νοσηλεία τους σε κάποιο νοσοκομείο. Γενικώς στόχος είναι να αναλύσουμε αυτά τα επαναλαμβανόμενα γεγονότα και ταυτόχρονα να καταλάβουμε και να περιγράψουμε τις διαδικασίες ανάλυσης των γεγονότων, την από άτομο σε άτομο διαφορετικότητα που προκύπτει και τέλος τη σχέση των προσαρμοσμένων συμμεταβλητών, των τρόπων θεραπείας που ακολουθούμε κάθε φορά και των χρόνο-μεταβαλλόμενων παραγόντων των διάφορων γεγονότων που προκύπτουν.

Δύο θεμελιώδεις τρόποι για να περιγράψουμε και να μοντελοποιήσουμε αυτά τα γεγονότα είναι μέσω απαριθμητών δεδομένων ή μέσω αριθμών γεγονότων σε ένα συγκεκριμένο χρονικό διάστημα και μέσω διαστημάτων ή χρόνων αναμονής ανάμεσα σε επιτυχημένα γεγονότα. Το ποια διαδικασία θα χρησιμοποιηθεί εξαρτάται από τα δεδομένα που έχουμε να εξετάσουμε. Τα απαριθμητά δεδομένα είναι πιο χρήσιμα σε τακτικά γεγονότα που συμβαίνουν στα άτομα που εξετάζονται, όταν τα ποσοστά των γεγονότων των πληθυσμών ή των συνόλων των ατόμων που εξετάζουμε είναι άξια ενδιαφέροντος και όταν τα προσαρμοσμένα κόστη ή οι συνέπειες είναι συνδεδεμένες με τα γεγονότα. Από την άλλη διαστήματα μεταξύ των γεγονότων είναι συχνά χρήσιμα σε μη-τακτικά γεγονότα, σε διαδικασίες στις οποίες συμβαίνει κάποιου τύπου ανανέωση μετά από ένα γεγονός που έχει συμβεί ήδη και τέλος όταν είναι άξιες ενδιαφέροντος οι προβλέψεις για τον

ακριβή χρόνο του επόμενου γεγονότος. Σε πολλές μελέτες και τα απαριθμητά δεδομένα αλλά και τα διαστήματα είναι αρκετά σχετικά μεταξύ τους.

Ένας πολύ γενικός και μαθηματικός τρόπος για να μοντελοποιήσουμε αυτά τα επαναλαμβανόμενα γεγονότα είναι μέσω των συναρτήσεων διακινδύνευσης. Στην ουσία ο κίνδυνος στο χρόνο t περιγράφει την πιθανότητα να εμφανιστεί ένα νέο γεγονός στο χρόνο t δεδομένου ότι γνωρίζουμε το ιστορικό των γεγονότων του ατόμου που εξετάζουμε. Αυτή η προσέγγιση είναι πολύ σημαντική όταν οι μεταβαλλόμενες συνθήκες στην κατάσταση ενός ατόμου επηρεάζουν την εμφάνιση κάποιων γεγονότων αλλά και όταν το ιστορικό ενός ατόμου από προηγούμενα γεγονότα μας βοηθάει να προβλέψουμε μελλοντικά γεγονότα.

Τα επαναλαμβανόμενα γεγονότα μπορεί να έχουν άλλα διακριτά χαρακτηριστικά που επηρεάζουν τον τρόπο ανάλυσης τους. Μια θεώρηση είναι αν υπάρχουν περίοδοι ή χρονικές διάρκειες συνδεδεμένες με τα γεγονότα. Για παράδειγμα σε μια μόλυνση ένα άτομο μπορεί να λάβει μια θεραπεία που να απομακρύνει τον κίνδυνο μιας νέας μόλυνσης. Επίσης ένας ασθενής που είναι ψυχικά άρρωστος θα νοσηλευτεί ένα συγκεκριμένο χρονικό διάστημα στο νοσοκομείο. Μια δεύτερη θεώρηση, είναι αν υπάρχουν συνέπειες όπως το κόστος που είναι συνδεδεμένα με τα γεγονότα και αν πρέπει να συμπεριληφθούν στην ανάλυση. Μια τρίτη θεώρηση, μπορεί να είναι η ανάγκη να διακρίνουμε διαφορετικούς τύπους γεγονότων ούτως ώστε να αναλύσουμε πιο πολύπλοκα γεγονότα. Τέλος το χρονικό πλαίσιο ενός ατόμου που εξετάζεται για κάποιο επαναλαμβανόμενο γεγονός μπορεί να εξαρτάται από κάποιο τερματικό γεγονός όπως ο θάνατος, ο οποίος αποκλείει την εμφάνιση κάποιου άλλου γεγονότος.

Παράδειγμα 1: Πνευμονικές επιδεινώσεις σε άτομα με κυστική ίνωση

Τα άτομα με κυστική ίνωση κινδυνεύουν με επαναλαμβανόμενες μολύνσεις του πνεύμονα και αυτό εξαιτίας της βλέννας που υπάρχει στους πνεύμονες. Ο Fuchs (1994) περιέγραψε μια κλινική μελέτη με σκοπό να εξερευνήσει την αποτελεσματικότητα της καθημερινής χορήγησης μιας ανασχεδιασμένης φόρμουλας του ανθρώπινου ενζύμου rhDNase 1 το οποίο προστατεύει από πνευμονικές εξάρσεις. Στο πείραμα τα άτομα τα οποία δεν έχουν πνευμονικές επιδεινώσεις διαλέχτηκαν τυχαία ώστε κάποιοι από αυτούς

να λάβουν τη νέα θεραπεία (η οποία ονομάζεται rhDNase) και κάποιοι άλλοι να λάβουν ένα πλασίμπο φάρμακο. Και οι δύο ομάδες παρατηρούνταν για 169 μέρες. Οι μέρες στις οποίες ένα περιστατικό πνευμονικής επιδείνωσης εμφανιζόταν, σημειώνονταν για κάθε ασθενή ξεχωριστά. Όταν ένα περιστατικό ξεκινούσε τότε αμέσως δίνονταν στο αντικείμενο προς μελέτη αντιβιοτικά και όταν αυτή η πνευμονική επιδείνωση εξαφανιζόταν τότε το αντικείμενο μας κινδύνευε με μια νέα πλέον πνευμονική επιδείνωση. Ένα περιστατικό πνευμονικής επιδείνωσης έχει μια συγκεκριμένη διάρκεια η οποία δεν είναι πρωταρχικού ενδιαφέροντος όμως χρησιμοποιείται για να μειώσουμε το συνολικό χρόνο στον οποίο ένα άτομο κινδυνεύει με νέα μόλυνση. Πολλά αντικείμενα δεν είχαν πνευμονικές επιδεινώσεις κατά τη διάρκεια της μελέτης. Συγκεκριμένα 185 από τους 324 από την ομάδα των πλασίμπο και 217 από τους 321 από την ομάδα των rhDNase δεν είχαν τέτοιες πνευμονικές επιδεινώσεις. Επίσης ο βαθμός βίαια εκπνεόμενου αέρα μετρήθηκε για κάθε άτομο τυχαία.

Παρόλο που οι πνευμονικές επιδεινώσεις μπορεί να συμβαίνουν συχνά για ένα αντικείμενο μακροπρόθεσμα, οι αριθμοί των γεγονότων της μελέτης είναι κάπως μικροί. Θα μπορούσε κανείς να βασίσει τις συγκρίσεις των δύο θεραπειών στο αριθμό των πνευμονικών επιδεινώσεων, στο χρόνο αναμονής μέχρι την πρώτη πνευμονική επιδείνωση και μετά στα διαστήματα μεταξύ των πετυχημένων πνευμονικών επιδεινώσεων ή στις συναρτήσεις διακινδύνευσης των γεγονότων. Θα αναλύσουμε αυτό το παράδειγμα παρακάτω.

Παράδειγμα 2: Επεισόδια δύσπνοιας και βήχα σε ασθενείς που έχουν βρογχικό άσθμα.

Ο Sears (1990) περιέγραψε μια τυχαίοποιημένη μελέτη όπου οι ομάδες των ασθενών που παίρνουν κανονικό φάρμακο και των ασθενών που παίρνουν ένα πλασίμπο φάρμακο διασταυρώνονται. Σε αυτή την μελέτη εξετάζονται τα αποτελέσματα μιας εισπνεόμενης βρογχοδιασταλτικής θεραπείας όπου χρησιμοποιείται η ουσία φενοτερόλη σε άτομα με σταθερό άσθμα. Στη μελέτη, 64 άτομα ακολουθήθηκαν για περίπου ένα χρόνο και εισέπνεαν κανονικά είτε την φενοτερόλη είτε το πλασίμπο υπό τη μορφή ξηρής σκόνης μέσω ενός συστήματος χορήγησης. 30 άτομα λάμβαναν το πλασίμπο για τις πρώτες 167 ημέρες και τη φενοτερόλη από την μέρα 168 έως την μέρα 343, και 34 άτομα λάμβαναν την ίδια θεραπεία αλλά με αντίθετη διαδικασία. Τα άτομα

κατέγραφαν κάθε μέρα τα συμπτώματα τους, τα ποσοστά μέγιστης εκπνευστικής ροής, τα φάρμακα που έπαιρναν και επέστρεφαν κάθε 4 εβδομάδες για εξετάσεις στην κλινική.

Τα συμπτώματα που είναι άξια ενδιαφέροντος είναι η καθημερινή δύσπνοια και ο βήχας, και κύριος στόχος είναι η εκτίμηση των ποσοστών αυτών των συμπτωμάτων σε σχέση με τη θεραπεία που ακολουθήθηκε καθώς και σε σχέση με άλλες συμμεταβλητές. Σε αυτή την περίπτωση είναι σημαντικό να δώσουμε έμφαση στη συχνότητα των συμπτωμάτων, ή στις μετρήσεις των δύο γεγονότων (επεισόδια βήχα και επεισόδια δύσπνοιας). Η διάρκεια των επεισοδίων δεν είναι πρωταρχικού ενδιαφέροντος παρόλα αυτά είναι σημαντικό να προσδιοριστεί τι προκαλεί αυτά τα μεμονωμένα επεισόδια. Μια απλοϊκή προσέγγιση είναι να προσδιορίσουμε τα γεγονότα που είναι συνδεδεμένα με τις μέρες στις οποίες κάνουν την εμφάνιση τους αυτά τα επεισόδια βήχα ή δύσπνοιας.

Παράδειγμα 3: Σκελετικές επιπλοκές σε ασθενείς που έχουν μετάσταση στα οστά.

Οι ασθενείς με καρκίνο του μαστού συχνά τους παρουσιάζουν μετάσταση του καρκίνου στα οστά. Η κατανομή των σημείων όπου συμβαίνει αυτή η μετάσταση ποικίλλει σημαντικά και εξαρτάται από τον αρχικό τύπο του όγκου αλλά και από το μέρος του σώματος στο οποίο δρα ο όγκος. Η θνησιμότητα που οφείλεται από αυτές τις βλάβες των οστών εξαρτάται από την τοποθεσία και από το μέγεθος της μετάστασης. Όλοι αυτοί οι ασθενείς κινδυνεύουν από παθολογικά κατάγματα, από μη παθολογικά κατάγματα, από πόνο των οστών και από καταστροφή των οστών τόσο σοβαρή που κάνει την ανάγκη εγχείρησης τόσο σημαντική ώστε να αποφευχθούν αυτές οι συνέπειες. Ο Hortobagyi (1996) παρουσίασε μια πλασίμπο-ελεγχόμενη μελέτη 380 ασθενών με καρκίνο του μαστού και με μετάσταση στα οστά τους οι οποίοι λαμβάνουν είτε μια θεραπεία από διφωσφονικά είτε κάποιο πλασίμπο φάρμακο. Σκοπός της μελέτης είναι να αξιολογηθεί η αποτελεσματικότητα των διφωσφονικών στη μείωση των επιπτώσεων ενός σύνθετου επαναλαμβανόμενου σκελετικού γεγονότος, όπως τα μη-σπονδυλικά κατάγματα, τα σπονδυλικά κατάγματα, η ανάγκη χρήσης ακτινοβολίας για τη θεραπεία του πόνου των οστών και η ανάγκη χειρουργείου. Τα σκελετικά γεγονότα έπρεπε να καταγράφονται στις τακτικές επισκέψεις που γίνονταν στο

νοσοκομείο κατά την διάρκεια ενός χρόνου από την ημερομηνία τυχαιοποίησης των ασθενών. Μακροπρόθεσμα δεδομένα επιβίωσης συλλέχθηκαν πέρα από αυτό το διάστημα.

Για τους ασθενείς οι οποίοι υποφέρουν από εκτεταμένο πρωταρχικό όγκο, υπάρχει ελλιπής πρόγνωση για τη μετάσταση του όγκου στα κόκαλα. Ο εκτεταμένος πρωταρχικός όγκος επίσης, αυξάνει τον κίνδυνο θανάτου. Ως αποτέλεσμα υπάρχει σχέση μεταξύ των διαδικασιών των σκελετικών γεγονότων και την επιβίωση, η οποία αποκλείει τη χρήση κάποιας γνωστής ανάλυσης επαναλαμβανόμενων γεγονότων. Συγκεκριμένα αυτά τα άτομα που βιώνουν αυτά τα σκελετικά γεγονότα στο υψηλότερο ποσοστό, τυπικά παρατηρούνται για μικρότερα διαστήματα εξαιτίας της θνησιμότητας.

Τα μοντέλα για επαναλαμβανόμενα γεγονότα μπορούν εύκολα να παρουσιαστούν μέσω των συναρτήσεων έντασής τους. Ας υποθέσουμε ότι $N(t)$ (είναι η από δεξιά συνεχής αριθμητική διαδικασία η οποία) καταγράφει τον αριθμό των γεγονότων που έχει βιώσει το αντικείμενο που μελετάμε, στο διάστημα $[0, t]$. Αν $H(t) = \{N(s): 0 \leq s < t\}$ αντιπροσωπεύει το ιστορικό του αντικειμένου μέχρι το χρόνο t , τότε η συνάρτηση έντασής του δίνεται από τον παρακάτω τύπο:

$$\lambda(t|H(t)) = \lim_{\Delta t \rightarrow 0} \frac{\Pr\{N(t + \Delta t^-) - N(t^-) = 1|H(t)\}}{\Delta t} \quad (1.1.1)$$

Αν υποθέσουμε ότι η πιθανότητα περισσότερα από ένα γεγονότα στο διάστημα

$$[t, t + \Delta t) \text{ να είναι } o(\Delta t),$$

$$\text{τότε } \Pr\{N(t + \Delta t^-) - N(t^-) = 1|H(t)\} = \lambda(t|H(t))\Delta t + o(\Delta t)$$

$$\text{και } \Pr\{N(t + \Delta t^-) - N(t^-) = 0|H(t)\} = 1 - \lambda(t|H(t))\Delta t + o(\Delta t).$$

Οι μέθοδοι μέγιστης πιθανοφάνειας μπορούν να χρησιμοποιηθούν για να προσαρμόσουμε και να εκτιμήσουμε μοντέλα που ορίζονται σε σχέση με τη συνάρτηση έντασης και αυτό διότι η πιθανότητα δεδομένου του ιστορικού ενός γεγονότος ενός ατόμου μπορεί να ληφθεί μέσω της (1.1.1). Για παράδειγμα η πυκνότητα πιθανότητας για n γεγονότα που συμβαίνουν στους χρόνους

$t_1 < \dots < t_n$, μιας διαδικασίας που παρατηρείται για ένα συγκεκριμένο διάστημα (τ_0, τ) είναι:

$$\prod_{j=1}^n \lambda(t_j | H(t_j)) \cdot \exp\left\{-\int_{\tau_0}^{\tau} \lambda(u | H(u)) du\right\} \quad (1.1.2)$$

Η πιθανοφάνεια στη συνάρτηση (1.1.2) έχει ισχύ επίσης όταν τα τ_0 και τ είναι τυχαία, υπό την προϋπόθεση βέβαια ότι έχουν προσδιορισθεί βάση μιας διαδικασίας η οποία εξαρτάται μόνο από το προηγούμενο ιστορικό των γεγονότων ή από τις συμμεταβλητές.

Ας υποθέσουμε ότι η $x(s)$ αντιπροσωπεύει ένα διάνυσμα χρόνο-εξαρτημένων συμμεταβλητών στο χρόνο s και το ενδιαφέρον μας έγκειται στο να εξετάσουμε το αποτέλεσμα που επιφέρουν στη διαδικασία των επαναλαμβανόμενων γεγονότων. Η ερμηνεία της συνάρτησης έντασης της (1.1.1) μπορεί να επεκταθεί στο να ενσωματωθεί η εξάρτηση των συμμεταβλητών. Αυτό γίνεται ενσωματώνοντας τις πληροφορίες που μας δίνουν οι συμμεταβλητές μέχρι το χρονικό σημείο t συμπεριλαμβανομένου βεβαίως και του χρονικού σημείου t , π.χ.

$$H(s) = \{N(s), 0 \leq s < t; x(s), 0 \leq s < t\}.$$

Στην πράξη συνήθως υποθέτουμε ότι οι συμμεταβλητές επηρεάζουν τη συνάρτηση έντασης μέσω ενός πολυμεταβλητού μοντέλου που είναι της μορφής:

$$\lambda(t | H(t), x(t)) = \lambda_0(t; H(t))g(x(t); \beta) \quad (1.1.3)$$

όπου $\lambda_0(t; H(t))$ είναι μια βασική συνάρτηση έντασης. Χρησιμοποιείται $g(x(t); \beta) = \exp\{x'(t)\beta\}$ σχεδόν χωρίς εξαίρεση, αλλά δεν αποτελεί μέρος του ορισμού του μοντέλου.

Ο ορισμός (1.1.1) είναι πολύ γενικός και μπορεί να φιλοξενήσει κάθε πιθανή εξάρτηση της έντασης από το ιστορικό της διαδικασίας $H(t)$. Στην πράξη οι ειδικές μορφές της $\lambda(t | H(t))$ πρέπει να καθορίζονται, ώστε να διευκολύνεται η ανάλυση. Δύο πολύ απλοί τύποι διαδικασιών είναι συνδεδεμένοι με την ανάλυση που εστιάζει είτε στην μέτρηση των γεγονότων είτε στα κενά διαστήματα μεταξύ των γεγονότων που εξετάζουμε. Η **ανέλιξη**

Poisson είναι διαδικασία για την οποία η (1.1.1) εξαρτάται από το t αλλά όχι από το ιστορικό των προηγούμενων γεγονότων μέχρι το t . Μια τέτοια ανελίξη μπορεί να γραφτεί ως:

$$\lambda(t|H(t)) = \rho(t) \quad (1.1.4)$$

Σε αυτή την περίπτωση ο αριθμός των γεγονότων $N(s, t)$ στο διάστημα $(s, t]$ ακολουθεί κατανομή Poisson με:

$$E[N(s, t)] = \int_s^t \rho(u) du = \mu(s, t) \quad (1.1.5)$$

και ο αριθμός των γεγονότων που πραγματοποιούνται στα ξένα μεταξύ τους διαστήματα χρόνων, είναι ανεξάρτητες τυχαίες μεταβλητές. Τα **ομοιογενή ως προς τον χρόνο μοντέλα Poisson** είναι αυτά για τα οποία ισχύει $\lambda(t|H(t)) = \rho(t)$ και $E[N(s, t)] = (t - s)\rho$. Οι **ανανεώσιμες ανελίξεις** είναι διαδικασίες για τις οποίες η (1.1.1) εξαρτάται μόνο από τον χρόνο που έχει περάσει από το πιο πρόσφατο γεγονός (υποθέτοντας βεβαίως ότι έχει συμβεί ένα γεγονός). Αν T_j είναι ο χρόνος του j -στου γεγονότος, τότε:

$$\lambda(t|H(t)) = h(t - T_{N(t^-)}) \quad (1.1.6)$$

Σε αυτή την διαδικασία τα κενά διαστήματα μεταξύ διαδοχικών γεγονότων είναι ανεξάρτητα και ισόνομα κατανομημένα με την συνάρτηση διακινδύνευσης $h(s)$.

Σε αρκετές περιπτώσεις ούτε οι Poisson αλλά ούτε και οι ανανεώσιμες ανελίξεις είναι κατάλληλες. Μια προσέγγιση για αυτά τα προβλήματα είναι να ψάξουμε έναν ορισμό της συνάρτησης έντασης ο οποίος να είναι ικανοποιητικός. Άλλες χρήσιμες προσεγγίσεις, είναι να θεωρήσουμε οικογένειες μοντέλων τα οποία γενικεύουν τις ανελίξεις Poisson και τις ανανεώσιμες ανελίξεις. Για παράδειγμα, ένα μοντέλο που επιτρέπει οι μετρήσεις ενός ατόμου να σχετίζονται μεταξύ τους είναι η οικογένεια αρνητικών διονυμικών διαδικασιών που περιγράφονται παρακάτω.

Στους ενδιάμεσους χρόνους, οι γενικεύσεις της ανανεώσιμης ανελίξης πρέπει να επιτρέπει στα διαδοχικά διαστήματα χρόνων S_1, S_2, \dots να είναι εξαρτημένα. Μια προσέγγιση είναι να υποθέσουμε ότι για μια συγκεκριμένη τιμή του k οι χρόνοι S_1, \dots, S_k έχουν μια συγκεκριμένη μορφή μιας

πολυμεταβλητής κατανομής. Οι Aalen, Husebye (1991) και Hougaard (2000) πρότειναν τη χρήση τυχαίων επιδράσεων με σκοπό να προσδιορίσουν αυτά τα μοντέλα. Μια δεύτερη προσέγγιση είναι να θεωρήσουμε μια ακολουθία υπό συνθήκη μοντέλων $f(s_1), f(s_2|s_1), f(s_3|s_1, s_2), \dots$ μέσα στα οποία η ανανεώσιμη ανέλιξη είναι μια ειδική περίπτωση.

Οι Poisson και οι ανανεώσιμες ανελίξεις είναι κανονικές υπό την έννοια ότι είναι τα πιο απλά μοντέλα για την μέτρηση γεγονότων και για τα ενδιάμεσα διαστήματα, αντίστοιχα. Στην επόμενη παράγραφο εξετάζουμε λεπτομερώς τα μοντέλα Poisson και σε άλλες παρόμοιες μεθόδους με σκοπό την ανάλυση των μετρήσεων.

1.2 Μερικές μέθοδοι που βασίζονται σε απεριθμητά δεδομένα

1.2.1 Poisson μοντέλα και μικτά Poisson μοντέλα

Αφού η συνάρτηση έντασης ενός μοντέλου Poisson δεν εξαρτάται από το ιστορικό της διαδικασίας, μπορούμε να την εκφράσουμε ως $\lambda(t|H(t))dt = \rho(t)dt = E[dN(t)]$ και μπορεί να ερμηνευθεί και ως μια περιθώρια και ως μια δεσμευμένη πιθανότητα. Έτσι θεωρούμε, το $dN(t) = N(t) - N(t^-)$ ως τον αριθμό των γεγονότων σε ένα ολοένα και πιο μικρό διάστημα με δεξιό τελικό σημείο t . Όταν εμφανίζονται στο πρόβλημα μας συμμεταβλητές, τότε εκφράζουμε τα **μοντέλα παλινδρόμησης Poisson** ως $\rho(t; x(t)) = \rho_0 \exp\{x'(t)\beta\}$. Στη συνέχεια θα περιορίσουμε την ανάλυση μας στις συμμεταβλητές που είναι ανεξάρτητες του χρόνου, και το $x_i(t) = x_i$ δηλώνει τα διανύσματα των συμμεταβλητών των αντικειμένων $i = 1, \dots, m$, και επίσης ισχύει ότι $\rho_i(t) = \rho_0 \exp(x_i'\beta)$. Οι επεκτάσεις στην περίπτωση των χρονικά μεταβαλλόμενων συμμεταβλητών είναι σχετικά απλές.

Υποθέτουμε ότι $\int_s^t \rho_0(u)du = \mu_0(s, t)$ και $\int_s^t \rho_i(u)du = \mu_i(s, t)$. Αν $\tau_0 = 0$ τότε $\int_0^t \rho_0(u)du = \mu_0(0, t)$ και $\mu_i(t) = \mu_0 \exp(x_i' \beta)$, $i = 1, 2, \dots, m$. Αν $N_i(\tau_i) = n_i$ και t_{i1}, \dots, t_{in} δηλώνουν τον παρατηρούμενο αριθμό των γεγονότων και τους παρατηρούμενους χρόνους των γεγονότων για το αντικείμενο i στην παρατηρούμενη περίοδο $[0, \tau_i]$, τότε η συνάρτηση πιθανοφάνειας για τον αντικείμενο i παίρνει τη μορφή:

$$L_i = \left\{ \prod_{j=1}^{n_i} \rho(t_{ij}; x_i) \right\} \exp \left\{ - \int_0^{\tau_i} \rho(u; x_i) du \right\}$$

$$= \left\{ \prod_{j=1}^{n_i} \frac{\rho_0(t_{ij})}{\mu_0(\tau_i)} \right\} \{ \mu_0(\tau_i) \exp(x_i' \beta)^{n_i} \exp(-\mu_0(\tau_i) \exp(x_i' \beta)) \} \quad (1.2.1)$$

Ο πρώτος όρος στην παραπάνω σχέση προκύπτει από τη δεσμευμένη κατανομή των χρόνων των γεγονότων δεδομένου του συνολικού αριθμού των γεγονότων, και ο δεύτερος όρος είναι η πιθανοφάνεια που βασίζεται μόνο στον αριθμό των γεγονότων. Ο Lawless (1987) ισχυρίζεται ότι η ανάλυση που βασίζεται στις μετρήσεις των γεγονότων είναι απόλυτα επαρκής κάτω από ένα ομοιογενές μοντέλο Poisson και επίσης για ποιο γενικές συναρτήσεις έντασης υπό την προϋπόθεση ότι $\tau_i = \tau$, $i = 1, 2, \dots, m$. Οι Dean και Balshaw (1997) τονίζουν ότι όταν αγνοούμε τον πρώτο όρο στην (1.2.2) τότε υπάρχει μια μικρή απώλεια στην αποτελεσματικότητα της εκτίμησης του β . Οι συναρτήσεις score που εξάγονται από τις πιθανοφάνειες των ανεξίτητων Poisson μπορούν επίσης να χρησιμοποιηθούν σαν εκτιμητρίες συναρτήσεων για γενικές σημειακές ανεξίτητες και ως βάση για την εύρωστη συμπερασματολογία.

Το μοντέλο Poisson εμπεριέχει τον περιορισμό $Var(N_i(\tau_i)) = E\{N_i(\tau_i)\} = \mu_i(\tau_i)$, ο οποίος περιορίζει σημαντικά την χρησιμότητά του. Συχνά, για παράδειγμα, ο ρυθμός εμφάνισης των γεγονότων ποικίλλει μεταξύ των αντικειμένων, σε μεγαλύτερο βαθμό από αυτό που μπορεί να εξηγηθεί με τη βοήθεια των συμμεταβλητών. Σε τέτοιες περιπτώσεις, είναι σύνηθες να υιοθετούμε ένα μοντέλο Poisson τυχαίων επιδράσεων στο οποίο η συνάρτηση έντασης για το αντικείμενο μελέτης i παίρνει τη μορφή $u_i \rho(t; x_i)$, όπου το u_i είναι ανεξάρτητες θετικές τυχαίες μεταβλητές που προέρχονται από την κατανομή Γάμμα $G(u; \varphi)$ με $E(u) = 1$

και $Var(u) = \varphi$. Η συνάρτηση $\rho(t; x)$ τώρα μεταφράζεται ως συνάρτηση έντασης στον πληθυσμό αντικειμένων με διάνυσμα συμμεταβλητής x , αφού $E\{dN_i(t)|x\} = \rho(t; x)dt$. Όταν το u_i ακολουθεί την κατανομή Γάμμα τότε η περιθώρια κατανομή του n_i είναι η αρνητική διωνυμική και η πιθανοφάνεια για το αντικείμενο i είναι:

$$L_i \left\{ \prod_{j=1}^{n_i} \frac{\rho_0(t_{ij})}{\mu_0(\tau_i)} \right\} \frac{\Gamma(n_i + \varphi^{-1})}{\Gamma(\varphi^{-1})} \frac{(\varphi \mu_0(\tau_i) \exp(x'_i \beta))^{n_i}}{(1 + \varphi \mu_0(\tau_i) \exp(x'_i \beta))^{n_i + \varphi^{-1}}} \quad (1.2.2)$$

Όπως και στην περίπτωση του μοντέλου Poisson, αν ισχύει $\tau_i = \tau$, $i = 1, 2, \dots, m$, ή όταν χρησιμοποιείται ένα ομοιογενές ως προς το χρόνο μοντέλο, τότε η ανάλυση που βασίζεται μόνο στον αριθμό των γεγονότων είναι επαρκής.

Στα μοντέλα τυχαίων επιδράσεων, ισχύει

$$Var(N_i(\tau_i)) = \mu_i(\tau_i) + \mu_i^2(\tau_i)\varphi$$

δηλαδή υπάρχουν επιπλέον Poisson διακυμάνσεις. Επιπλέον αν $t_1 < t_2$, τότε:

$$Cov(N_i(s_1, t_1), N_i(s_2, t_2)) = \varphi \mu_i(s_1, t_1) \mu_i(s_1, t_1) \quad (1.2.3)$$

που σημαίνει ότι η υπόθεση ανεξάρτητων μετρήσεων ξένων διαστημάτων δεν ισχύει πλέον. Η παράμετρος φ αντανακλά τόσο το βαθμό των έξτρα Poisson διακυμάνσεων όσο και το βαθμό σύνδεσης μεταξύ των μετρήσεων των ξένων διαστημάτων. Οι Thall και Vail (1995) πρότειναν ένα πιο γενικό μοντέλο στο οποίο:

$$Var(N_i(s, t)) = \mu_i(s, t) + \mu_i^2(s, t)\varphi_1, \quad \varphi_1 > 0 \quad (1.2.4)$$

$$Cov(N_i(s_1, t_1), N_i(s_2, t_2)) = \mu_i(s_1, t_1) \mu_i(s_2, t_2) \varphi_2, \quad \varphi_2 > 0 \quad t_1 < s_2 \quad (1.2.5)$$

επιτρέποντας έτσι έξτρα-Poisson διακυμάνσεις αλλά ανεξάρτητες μετρήσεις σε ξένα χρονικά διαστήματα καθώς επίσης και Poisson διακυμάνσεις αλλά μη ανεξάρτητες προσαυξήσεις σε ξένα μεταξύ τους χρονικά διαστήματα. Όμως η συμπερασματολογία με βάση την πιθανοφάνεια είναι δύσκολη.

1.2.2 Εύρωστες μέθοδοι για συναρτήσεις ρυθμού και μέσες συναρτήσεις

Τα περιθώρια μοντέλα που εξαρτώνται λιγότερο από συγκεκριμένες κατανομές, απέκτησαν πρόσφατα ενδιαφέρον όσον αφορά την χρήση των εύρωστων μεθόδων για τον υπολογισμό συναρτήσεων ρυθμού και μέσων συναρτήσεων που βασίζονται στις εξισώσεις Poisson με εύρωστες εκτιμήσεις. Θεωρείστε το περιθώριο μοντέλο $E\{dN_i(t)|x_i\} = \rho_0(t) \exp\{x_i'\beta\} dt$ και καθόλου υποθέσεις που αφορούν κάποια κατανομή πέρα από την ανεξαρτησία του τ_i και $\{N_i(s), s > 0\}, i = 1, \dots, m$. Τέτοιες υποθέσεις είναι ελκυστικές, δεδομένου είναι ένα περιθώριο στοιχείο τις διαδικασίας. Ακόμα και όταν ένα μοντέλο Poisson δεν είναι έγκυρο, η βασική συνάρτηση ρυθμού και η παράμετρος β είναι ερμηνεύσιμες. Είναι σχετικά απλό να δείξουμε ότι αν θ είναι δείκτης της $\rho_0(t; \theta)$, τότε οι εξισώσεις score για το (θ, β) που προκύπτουν από την παραγοντοποίηση του λογαρίθμου του $L = \prod_{i=1}^m L_i$, όπου L_i δίνεται από την (1.2.1) είναι οι εξής:

$$U_\theta(\theta, \beta) = \sum_{i=1}^m \int_0^\tau \delta_i(s) \{dN_i(s) - \rho_0(s; \theta) ds \exp\{x_i'\beta\}\} \frac{\partial \log \rho_0(s; \theta)}{\partial \theta} = 0 \quad (1.2.6)$$

$$U_\beta(\theta, \beta) = \sum_{i=1}^m \int_0^\tau \delta_i(s) \{dN_i(s) - \rho_0(s; \theta) ds \exp\{x_i'\beta\}\} x_i = 0 \quad (1.2.7)$$

όπου $\delta_i(s) = I(s < \tau_i)$, και τ_i είναι η μέγιστη διάρκεια παρακολούθησης. Υπό την προϋπόθεση ότι $E\{dN_i(t)|x_i\} = \rho_0(s) \exp\{x_i'\beta\} dt$ και η κατανομή του τ_i είναι ανεξάρτητη της διαδικασίας των γεγονότων, τότε η λύση των συναρτήσεων (1.2.6) και (1.2.7) είναι συνεπείς εκτιμήτριες του (θ, β) . Οι εκτιμήσεις της εύρωστης διασποράς από τον εκτιμητήρια τύπου Sandwich (1991) μπορούν να ληφθούν από $(\hat{\theta}, \hat{\beta})$ όπως περιγράφεται και από τους Lawless και Nadeau (1995). Οι εκτιμήτριες τύπου Sandwich είναι ένα σύνηθες εργαλείο που χρησιμοποιείται για την εκτίμηση της διασποράς των εκτιμητριών των παραμέτρων.

Αν επιθυμούμε να χρησιμοποιήσουμε ένα ημιπαραμετρικό μοντέλο, τότε μπορούμε να αφήσουμε τη μορφή του $\rho_0(s)$ μη προσδιορισμένη και να αντικαταστήσουμε την (1.2.6) με:

$$\sum_{i=1}^m \delta_i(s) \{dN_i(s) - \rho_0(s) ds \exp\{x_i' \beta\}\} = 0 \quad s \geq 0 \quad (1.2.8)$$

Λύνοντας την (1.2.8) για $\rho_0(s) ds$ συναρτήσει του β έχουμε:

$$\tilde{\rho}_0(s; \beta) ds = \frac{\sum_{i=1}^m \delta_i(s) dN_i(s)}{\sum_{i=1}^m \delta_i(s) \exp\{x_i' \beta\}} \quad (1.2.9)$$

η οποία μπορεί να αντικατασταθεί στην (1.2.7) και να δώσει ένα $p \times 1$ σύστημα εξισώσεων του β . Αυτό το σύστημα μπορεί να λυθεί ώστε να ληφθεί το $\hat{\beta}$, το οποίο αποδεικνύεται ότι είναι εκτιμήτρια μερικής πιθανοφάνειας του μοντέλου του Cox και είναι διαθέσιμο από οποιοδήποτε λογισμικό που σχετίζεται με την ανάλυση επιβίωσης. Και το $\hat{\beta}$ αλλά και εύρωστο τυπικό σφάλμα του $\hat{\beta}$ είναι διαθέσιμα μέσω της συνάρτησης `coxph` της R, χρησιμοποιώντας την επιλογή `cluster` για τα άτομα. Η ημιπαραμετρική εκτιμήτρια που αντιστοιχεί στον σωρευτικό αναμενόμενο αριθμό γεγονότων στο διάστημα $(0, \tau]$ για το αντικείμενο i , είναι:

$$\hat{E}[N_i(t)] = \int_0^t \hat{\rho}_0(s) \exp\{x_i' \hat{\beta}\} ds \quad (1.2.10)$$

όπου $\hat{\rho}_0(s) = \tilde{\rho}_0(s; \hat{\beta})$. Αυτό είναι παρόμοιο με τη γενικευμένη εκτιμήτρια Nelson-Aalen από την ανάλυση επιβίωσης.

Σε τυχαίοποιημένες κλινικές μελέτες, το ενδιαφέρον συχνά βρίσκεται στις συγκρίσεις δύο ομάδων για την αξιολόγηση των αποτελεσμάτων των πειραματικών θεραπευτικών επεμβάσεων. Αν δεν υπάρχουν συμμεταβλητές τότε η εκτιμήτρια συνάρτηση που αντιστοιχεί στην (1.2.8) με $x_i = 0$, δίνει ένα μη παραμετρικό εκτιμητή της συνηθισμένης ποσοστιαίας συνάρτησης ρυθμού $\rho(s)$ και ο εκτιμητής Nelson-Aalen (1.2.10) εκτιμάει την αντίστοιχη μέση συνάρτηση $E\{N_i(t)\}$. Αν τα $\rho_1(s)$ και $\rho_2(s)$ δηλώνουν την συνάρτηση ρυθμού στο χρονικό διάστημα s για τους ασθενείς στις ομάδες 1 και 2 αντίστοιχα, τότε η μηδενική υπόθεση της μη διαφοράς μεταξύ των σκελών της θεραπείας δίνεται από $H_0: \rho_1(s) = \rho_2(s), s > 0$. Ας υποθέσουμε

ότι $\delta_{ki}(s) = 1$ αν το i -στο αντικείμενο μελέτης της ομάδας k είναι σε κίνδυνο εμφάνισης κάποιου γεγονότος στο χρονικό διάστημα s , και $\delta_{ki}(s) = 0$ σε διαφορετική περίπτωση. Επιπλέον, ας υποθέσουμε ότι οι τελείες δηλώνουν άθροιση υπό τον αντίστοιχο δείκτη, δηλαδή $\delta_k(s)$ είναι ο συνολικός αριθμός αντικειμένων που είναι σε κίνδυνο στο χρονικό διάστημα s της ομάδας k , και $\delta_{..}(s) = \delta_{.1}(s) + \delta_{.2}(s)$. Ο Cook κ.α (1996) περιέγραψαν μια κλάση ψευδό-στατιστικών συναρτήσεων score της μορφής:

$$U(\tau) = \int_0^\tau w(s)\{\hat{\rho}_1(s) - \hat{\rho}_2(s)\}ds$$

όπου $w(s) = \delta_{.1}(s)\delta_{.2}(s)a(s)/\delta_{..}(s)$, και $a(s)$ είναι μια αναμενόμενη συνάρτηση βάρους. Χρησιμοποιούν εκτιμητές διασποράς (Lawless, 1995) για να κατασκευάσουν στατιστικούς ελέγχους, με σκοπό την ανίχνευση διαφορών τύπων απομάκρυνσης από την H_0 , μέσω της κατάλληλης επιλογής κατάλληλων συναρτήσεων βάρους $a(s)$, $s > 0$. Οι Cook και Lawless (1996) επέκτειναν το έργο αυτό με την ανάπτυξη στρατηγικών διαδοχικών δοκιμών μη επίδρασης της θεραπείας βασισμένες σε δαπανηρές συναρτήσεις σφάλματος.

Συχνά εμφανίζονται διάφοροι τύποι γεγονότων και το ενδιαφέρον μας εστιάζεται σε μια από κοινού ανάλυση που ξεχωρίζει μεταξύ των τύπων των γεγονότων και στα αποτελέσματα των συμμεταβλητών που σχετίζονται με αυτά τα γεγονότα. Παραδείγματα προκύπτουν σε διάφορες δοκιμές οστεοπόρωσης όπου είναι σημαντικός ο διαχωρισμός μεταξύ της συχνότητας εμφάνισης καταγμάτων σπονδυλικής στήλης καθώς και της μη εμφάνισής τους, σε διάφορες έρευνες αναπνευστικών παθήσεων όπου το ενδιαφέρον επικεντρώνεται στην από κοινού ανάλυση επεισοδίων βήχα και δύσπνοιας, και τέλος στην οικονομία του συστήματος υγείας όπου είναι σημαντικό να περιγραφούν οι περιθώριες πλευρές του τρόπου χρήσης διαφόρων υπηρεσιών υγείας. Ας υποθέσουμε ότι το $N_i(t) = (N_{i1}(t), N_{i2}(t))'$ δηλώνει ένα διμεταβλητό διάνυμα μετρήσεων όπου $N_{ij}(t)$ δηλώνει τον αριθμό των γεγονότων τύπου j στο διάστημα $(0, t]$, $j=1,2$. Τα περιθώρια μοντέλα της μορφής:

$$E\{dN_{ij}(s)|x_{ij}\} = \rho_{0j}(s)g(x_{ij}(s))_j(\beta_j)ds, \quad j = 1,2 \quad (1.2.11)$$

μπορούν να εκτιμηθούν αρκεί να βασιστούν σε μια πολυμεταβλητή επέκταση των (9) και (10) (Dean και Balshaw 1997). Τα αποτελέσματα των συμμεταβλητών μπορούν ελάχιστα να ερμηνευθούν.

1.2.3 Ανάλυση ενδιάμεσων χρόνων

Ας υποθέσουμε ότι το $T_{i1} < T_{i2} < \dots$ αναπαριστά τους χρόνους στους οποίους το αντικείμενο i έχει κάποιο γεγονός και $S_{ij} = T_{ij} - T_{i,j-1}$ ($j = 1, 2, \dots$) με $T_{i0} = 0$ να δηλώνει την διαφορά μεταξύ επιτυχημένων γεγονότων. Για λόγους απλότητας, υποθέτουμε ότι οι συμμεταβλητές είναι καθορισμένες για τα διαστήματα μεταξύ των γεγονότων και το x_{ij} να δηλώνει έναν δείκτη συμμεταβλητών για το S_{ij} .

Οι ανανεώσιμες ανελίξεις έχουν μια ισχυρή ιδιότητα. Αυτή είναι δεδομένου του x_{ij} , τα S_{i1}, S_{i2}, \dots να είναι ανεξάρτητα. Συνήθως χρειάζονται πιο γενικά μοντέλα και αυτά μπορούν να διατυπωθούν μέσω των υπό συνθήκη κατανομών:

$$F_j(s | x_{ij}, s_i^{(j-1)}) = \Pr(S_{ij} \leq s | x_{ij}, s_i^{(j-1)}) \quad (1.2.12)$$

όπου $s_i^{(j-1)} = (s_{i1}, \dots, s_{i,j-1})'$ είναι οι παρατηρούμενοι ενδιάμεσοι χρόνοι που συμβαίνουν στο αντικείμενο $i-1$ και το $s_i^{(0)}$ είναι κενό. Η συνάρτηση πιθανοφάνειας στα παρατηρούμενα δεδομένα στο διάστημα $\tau_{i0} \leq t \leq \tau_i$ είναι εύκολη στο να υπολογιστεί, υπό την προϋπόθεση ότι το τ_{i0} και το τ_i είναι χρόνοι διακοπής (Aalen και Husebye 1991). Αυτό ουσιαστικά σημαίνει ότι το τ_{i0} και το τ_i εξαρτώνται από το ιστορικό της διαδικασίας που ακολουθείται και από τις προηγούμενες συμμεταβλητές, αλλά όχι από διάφορες μελλοντικές τιμές. Η μορφή της συνάρτησης πιθανοφάνειας για το αντικείμενο i , είναι μια ακολουθία από χρόνους αποτυχίας $S_{i1}, S_{i2}, \dots, S_{ik_i}$, με το S_{i1} να είναι πιθανόν από αριστερά κολοβή παρατήρηση (αν $t_{i0} < \tau_{i0}$) και το S_{ik_i} να είναι πιθανόν από δεξιά αποκομμένο (αν $t_{i,k_i-1} < \tau_i < t_{i,k_i}$). Αν το S_{i1} είναι από αριστερά κολοβό και το t_{i0} είναι άγνωστο τότε τα πράγματα είναι πιο πολύπλοκα.

Τα μοντέλα αναλογικής διακινδύνευσης είναι αρκετά χρήσιμα. Ένα κάπως πιο γενικό μοντέλο για την (1.2.12) είναι αυτό για το οποίο η συνάρτηση διακινδύνευσης για S_{ij} δεδομένου των x_{ij} και $s_i^{(j-1)}$ είναι της μορφής:

$$h_{ij}(s) = h_{0j}(s) \exp(z'_{ij}\beta_j) \quad j = 1, 2, \dots \quad (1.2.13)$$

όπου το z_{ij} δηλώνει ένα δείκτη που περιέχει συναρτήσεις των x_{ij} και $s_i^{(j-1)}$, και για λόγους ευκολίας θα γράφουμε $h_{ij}(s)$ αντί για $h_j(s|x_{ij}, s_i^{(j-1)})$. Αν το z_{ij} δεν εξαρτάται από το $s_i^{(j-1)}$, τότε το μοντέλο μας είναι μια μαρκοβιανή ανανεώσιμη ανέλιξη με υπό προϋποθέσεις ανεξάρτητα S_{ij} . Τα μοντέλα όπου το $z'_{ij}\beta_j$ παίρνει την παρακάτω μορφή

$$z'_{ij}\beta_j = x'_{ij}\gamma_j + s_i^{(j-1)'}\omega_j \quad (1.2.14)$$

μπορούν να χρησιμοποιηθούν για να ενσωματώσουν ή για να ελέγξουμε την εξάρτηση ανάμεσα στους ενδιάμεσους χρόνους. Σε ορισμένες περιπτώσεις οι παράμετροι παλινδρόμησης β_j και η βασική συνάρτηση διακινδύνευσης $h_{0j}(s)$ στη σχέση (1.2.13) μπορεί να έχουν κοινές τιμές για $j=1, 2, \dots$.

Μερικές φορές είναι επιθυμητό να θεωρήσουμε την περιθώρια κατανομή για τους ενδιάμεσους χρόνους S_{ij} , υπό την προϋπόθεση όμως μόνο για τις συμμεταβλητές x_{ij} και όχι για τους προηγούμενους ενδιάμεσους χρόνους. Αν τα S_{ij} είναι υπό συνθήκη ανεξάρτητα (π.χ να δίνονται οι συμμεταβλητές) τότε αυτό είναι απλό μοντέλο. Στα μοντέλα όμως στα οποία υπάρχει εξάρτηση τότε αυτά δίνουν πιο πολύπλοκες ερμηνείες για τις κατανομές των S_{ij} δεδομένου των x_{ij} για $j=1, 2, \dots$. Το περιθώριο αποτέλεσμα των x_{ij} δεν είναι εύκολο να ερμηνευθεί σε τέτοια μοντέλα. Μια θεώρηση μπορεί να είναι "η προσέγγιση εργάσιμης ανεξαρτησίας" στην οποία τα ξεχωριστά PH μοντέλα για τα S_{ij} δεδομένου των x_{ij} είναι προσδιορισμένα, αλλά όχι από κοινού μοντέλα. Παρόλα αυτά, η ξεχωριστή ανάλυση για τα $j=1, 2, \dots$ είναι έγκυρη όταν το παρατηρούμενο διάστημα (τ_{i0}, τ_i) αντιστοιχεί σε έναν προσδιορισμένο από πριν αριθμό γεγονότων. Η ξεχωριστή περιθώρια ανάλυση είναι μη έγκυρη πιο γενικά, επειδή η εξάρτηση ανάμεσα στα S_{ij} δημιουργεί μη ανεξάρτητη αποκοπή για τα S_{i2}, S_{i3}, \dots . Για παράδειγμα, αν το

παρατηρούμενο διάστημα για το αντικείμενο i είναι το $(0, \tau_i)$, τότε ο χρόνος αποκοπής για το S_{i2} είναι $\tau_i - \min(S_{i1}, \tau_i)$. Αυτό δεν είναι ανεξάρτητο του S_{i2} , και έτσι η τυπική ανάλυση για το S_{i2} από μόνη της είναι μη έγκυρη. Αυτό το θέμα συζητήθηκε από τους Viser (1996), Wang και Wells (1998), Lin κ.α. (2000), οι οποίοι παρείχαν εκτιμήσεις για τις περιθώριες συναρτήσεις επιβίωσης των S_{i2}, S_{i3}, \dots χωρίς συμμεταβλητές.

Ουσιαστικά, δεν μπορούμε να τα καταφέρουμε μόνο με τις περιθώριες κατανομές για τα S_{ij} , εκτός και αν αυτά είναι ανεξάρτητα και επομένως τότε είναι σημαντικό να εξετάσουμε την από κοινού κατανομή. Αυτό επί της ουσίας πραγματοποιείται από τα μοντέλα που παρουσιάζονται στις (1.2.12) και (1.2.13), όμως για πιο εύκολη ερμηνεία των περιθώριων κατανομών προτιμούμε να εξετάσουμε πολυμεταβλητά μοντέλα (Hougaard, 2000) στα οποία, για ένα δεδομένο J , οι ενδιάμεσοι χρόνοι S_{i1}, \dots, S_{ij} έχουν μια κατανομή που έχει μια πιο απλή μορφή. Τα λογαριθμοκανονικά μοντέλα $(\log S_{i1}, \dots, \log S_{ij})$ τα οποία ακολουθούν μια πολυμεταβλητή Κανονική κατανομή εμφανίζονται αρκετά συχνά. Τα χαρακτηριστικά των τυχαίων επιπτώσεων μπορούν επίσης να εξετασθούν. Για παράδειγμα, μπορούμε να εξετάσουμε ένα μοντέλο που εξαρτάται από συμμεταβλητές, από την τυχαία επίδραση u_i , και τα S_{ij} είναι ανεξάρτητα για $j=1,2,\dots$ με:

$$h_{ij}(t|u_i) = u_i h_{0j}(t) \exp(x'_{ij} \beta_j) \quad (1.2.15)$$

Οι Follman και Goldberg (1988), Aalen και Husebye (1991), Lawless και Fong (1999), και άλλοι εξέτασαν αυτά τα μοντέλα. Μερικά μοντέλα (1.2.15) μπορούν να εκτελεστούν χρησιμοποιώντας το μοντέλο του Cox με τη βοήθεια της R.

Κεφάλαιο 2: Μοντέλα

Επαναλαμβανόμενων Γεγονότων και

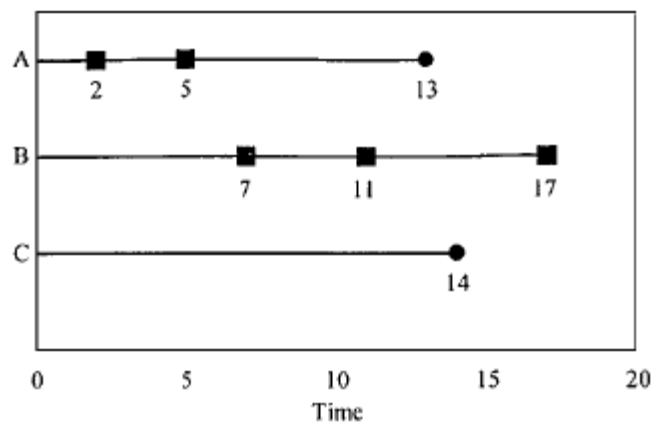
άλλες χρήσιμες έννοιες

2.1 Βασικές συνιστώσες μοντέλου

Οι τέσσερις συνιστώσες ενός βασικού μοντέλου επαναλαμβανόμενων γεγονότων του Cox είναι οι εξής: ορισμός του διαστήματος διακινδύνευσης , ορισμός του σύνολο κινδύνων, επιλογή μιας κοινής συνάρτησης διακινδύνευσης σε σχέση με το προσδιορισμένο γεγονός και τέλος η διαχείριση της συσχέτισης του αντικειμένου. Αυτές οι έννοιες παρουσιάζονται με ένα υποθετικό παράδειγμα (Σχήμα 2.1) όπου βλέπουμε τις εμπειρίες τριών ατόμων που έχουν βιώσει επαναλαμβανόμενα γεγονότα.

Τα μοντέλα διορθωμένης διασποράς που θα παρουσιαστούν σε άλλη παράγραφο αναπτύχθηκαν για λογαριασμό της παρατηρήτης, ή τουλάχιστον της ανυπολόγιστης ετερογένειας χρησιμοποιώντας εύρωστα τυπικά σφάλματα και μερικές φορές στρωματοποίηση. Οι διακυμάνσεις μέσα στην οικογένεια των διορθωμένων μοντέλων διασποράς βασίζονται σε διαφορετικούς ορισμούς του συνόλου κινδύνων, δηλαδή το πόσα από τα άτομα βρίσκονται σε κίνδυνο για ένα οποιοδήποτε γεγονός k , και πότε το επιτρέπουν για βασικές συναρτήσεις διακινδύνευσης. Στο μοντέλο των Andersen-Gill (Andersen και Gill, 1982) όλες οι μονάδες βρίσκονται σε κίνδυνο για κάθε γεγονός σε όλες τις περιόδους και μοιράζονται μια βασική κοινή συνάρτηση ρυθμού. Αντίθετα τα υπό συνθήκη μοντέλα στρωματοποιούν τα δεδομένα ανά γεγονός, έτσι ώστε η βασική συνάρτηση διακινδύνευσης να επιτρέπεται να ποικίλλει ανάλογα με το κάθε γεγονός. Τα υπό συνθήκη μοντέλα που θα παρουσιαστούν σε άλλη παράγραφο εκτιμώνται σε χρόνο που έχει παρέλθει γενικά ή σε ενδιάμεσο χρόνο, και οι μονάδες έχουν σχεδιαστεί έτσι ώστε να βρίσκονται σε κίνδυνο τα αντικείμενα μελέτης να βιώσουν το k -στό γεγονός αφού φυσικά έχουν βιώσει το $k-1$ γεγονός (Prentice, Williams και Peterson 1981, Gail, Santner και Brown 1981). Η εκτίμηση του χρόνου που έχει παρέλθει, παράγει το κίνδυνο εμφάνισης ενός γεγονότος από την έναρξη της μελέτης, ενώ οι ενδιάμεσοι

χρόνοι δίνουν τον κίνδυνο από το τελευταίο γεγονός. Για παράδειγμα, στο χρόνο που έχει παρέλθει μία παρατήρηση με γεγονότα στους μήνες 4, 10 και 14 θα έχουν χρόνους έναρξης και διακοπής 0-4, 0-10, 0-14 αντίστοιχα. Στους ενδιάμεσους χρόνους η παρατήρηση θα έχει χρόνους έναρξης και διακοπής 0-4, 0-6, 0-4. Η επιλογή των ενδιάμεσων χρόνων σε σχέση με το χρόνο που έχει παρέλθει εξαρτάται από το πρόβλημα που έχουμε να μελετήσουμε. Η χρησιμοποίηση του χρόνου που έχει παρέλθει προϋποθέτει ότι υπάρχουν ουσιώδεις λόγοι για τους οποίους το ρολόι να κάνει επανεκκίνηση μετά από κάθε γεγονός. Ένα τέτοιο μοντέλο χρησιμοποιείται για να καθορίσει το αποτέλεσμα των μεταβλητών στο k-στο γεγονός δεδομένου του χρονικού διαστήματος από την προηγούμενη περίπτωση. Αντίθετα τα μοντέλα του χρόνου που έχει παρέλθει, αξιολογούν το αποτέλεσμα των μεταβλητών στην περίπτωση του k-στου γεγονότος από την έναρξη της μελέτης.

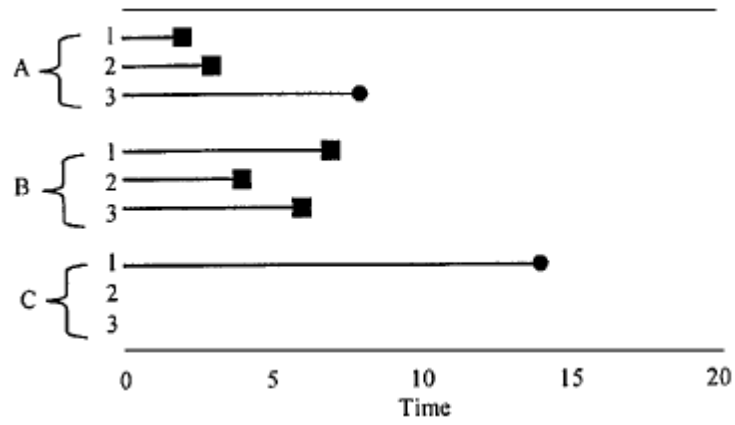


Σχήμα 2.1. Ένα υποθετικό παράδειγμα τριών αντικειμένων που βιώνουν επαναλαμβανόμενα γεγονότα

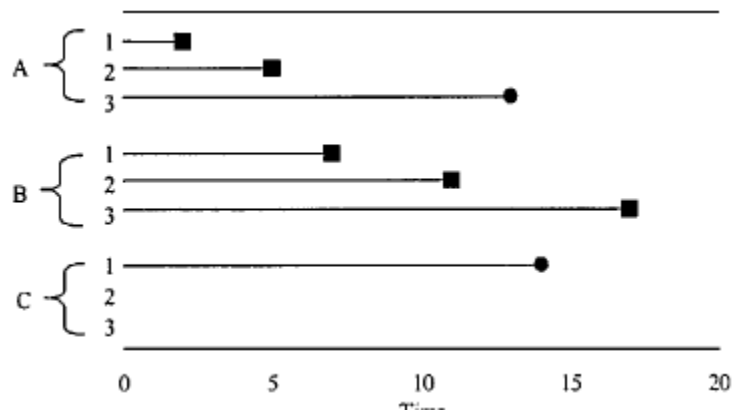
2.1.1 Διαστήματα Διακινδύνευσης

Τα διαστήματα διακινδύνευσης, ορίζονται όταν ένα αντικείμενο είναι σε κίνδυνο να βιώσει ένα γεγονός σε μια δεδομένη χρονική κλίμακα. Τρεις συνθέσεις είναι διαθέσιμες. Αυτές είναι: οι ενδιάμεσοι χρόνοι (gap times), ολικοί χρόνοι (total times) και ο τύπος της διαδικασίας απαριθμητών δεδομένων (counting process formulation). Στο Σχήμα 2.2 παρουσιάζονται τα τρία είδη διαστημάτων διακινδύνευσης των αντικειμένων του Σχήματος 2.1. Ενδιάμεσος χρόνος είναι χρόνος από το προηγούμενο γεγονός, δηλαδή, γίνεται επανεκκίνηση του ρολογιού μετά από κάθε γεγονός. Για παράδειγμα ο ενδιάμεσος χρόνος του αντικειμένου A (Σχήμα 2.2(α)) είναι σε κίνδυνο για το πρώτο γεγονός στο διάστημα $(0,2]$, και για το δεύτερο και τρίτο γεγονός κατά τη διάρκεια των διαστημάτων $(0,3]$ και $(0,8]$ αντίστοιχα. Ολικός χρόνος είναι ο χρόνος από ένα επιλεγμένο σημείο, συνήθως ο χρόνος από την έναρξη της θεραπείας. Για τον ολικό χρόνο, το αντικείμενο A (Σχήμα 2.2(β)), είναι σε κίνδυνο για το πρώτο, δεύτερο και τρίτο γεγονός κατά τη διάρκεια των διαστημάτων $(0, 2]$, $(0, 5]$ και $(0, 13]$, αντίστοιχα.

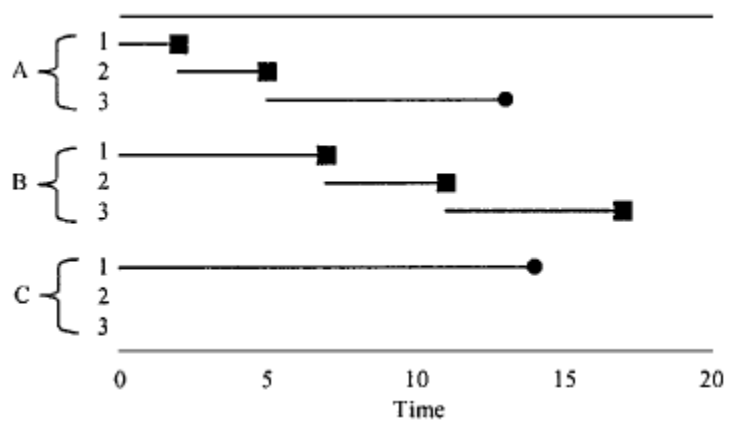
α) ενδιάμεσοι χρόνοι



β) ολικοί χρόνοι



γ) διαδικασία απαριθμητών δεδομένων



Σχήμα 2.2. Απεικόνιση των διατοπώσεων των διαστημάτων διακινδύνευσης με βάση το Σχήμα 1

Η διαδικασία απαριθμητών δεδομένων χρησιμοποιεί την ίδια χρονική κλίμακα όπως ο ολικός χρόνος, αλλά αναγνωρίζει ότι ένα αντικείμενο μπορεί να έχει μια καθυστερημένη έναρξη ή μια περίοδο αποκοπής, πριν το αντικείμενο βρεθεί σε κίνδυνο να βιώσει ένα γεγονός, με τα επαναλαμβανόμενα γεγονότα να θεωρείται ότι βρίσκονται σε κίνδυνο να βιώσουν κάποιο k -στο γεγονός αφού έχουν βιώσει το $(k-1)$ γεγονός.

Στη διαδικασία απαριθμητών δεδομένων το αντικείμενο A (Σχήμα 2.2(γ)) είναι σε κίνδυνο για το πρώτο γεγονός στο διάστημα $(0,2]$, και για το δεύτερο και τρίτο γεγονός κατά την διάρκεια των διαστημάτων $(2,5]$ και $(5,13]$ αντίστοιχα. Και στους ενδιάμεσους χρόνους αλλά και στη διαδικασία απαριθμητών δεδομένων, το αντικείμενο είναι σε κίνδυνο να βιώσει κάποιο γεγονός για το ίδιο χρονικό διάστημα. Το διάστημα διακινδύνευσης για το πρώτο γεγονός είναι ίδιο και για τους τρεις ορισμούς διαστημάτων διακινδύνευσης.

Το διάστημα διακινδύνευσης καθορίζει αν ένα μοντέλο είναι είτε περιθώριο είτε υπό συνθήκη. Οι ενδιάμεσοι χρόνοι και η διαδικασία καταμέτρησης είναι υπό συνθήκη αφού ένα αντικείμενο μελέτης δεν μπορεί να είναι σε κίνδυνο μέχρι να τελειώσει ένα προηγούμενο γεγονός, δηλαδή η πιθανότητα ένα αντικείμενο να είναι σε κίνδυνο εξαρτάται από τα προηγούμενα γεγονότα. Ο ολικός χρόνος είναι περιθώριος, αφού ένα αντικείμενο είναι σε κίνδυνο από την έναρξη της θεραπείας και δεν εξαρτάται από τα προηγούμενα γεγονότα.

2.1.2 Βασική Συνάρτηση Διακινδύνευσης

Υπάρχουν δύο επιλογές για τη βασική συνάρτηση διακινδύνευσης για τα μοντέλα επαναλαμβανόμενων γεγονότων: τα κοινά και τα προσδιορισμένα γεγονότα. Το μοντέλο με μια κοινή βασική συνάρτηση διακινδύνευσης έχει τον ίδιο βασικό κίνδυνο για όλα τα γεγονότα. Μια βασική συνάρτηση διακινδύνευσης προσδιορισμένων γεγονότων είναι μια στρωματοποιημένη βασική συνάρτηση διακινδύνευσης, η οποία επιτρέπει στη βασική συνάρτηση διακινδύνευσης να είναι διαφορετική για κάθε k -στο γεγονός. Η

στρωματοποίηση ανά γεγονός είναι σημαντική για να προσαρμόσουμε ένα ξεχωριστό μοντέλο για κάθε k -στο γεγονός.

2.1.3 Σύνολο που είναι σε κίνδυνο

Το k -στο σύνολο διακινδύνευσης περιέχει τα άτομα που βρίσκονται σε κίνδυνο μέχρι την εκδήλωση του k -στου γεγονότος. Υπάρχουν τρία πιθανά σύνολα διακινδύνευσης : τα απεριόριστα, τα περιορισμένα και τα ημιπεριορισμένα. Ο ορισμός του συνόλου που βρίσκεται σε κίνδυνο προϋποθέτει την επιλογή της βασικής συνάρτησης διακινδύνευσης. Το σύνολο που βρίσκεται σε κίνδυνο μιας δεδομένης χρονικής στιγμής εξαρτάται από τα άτομα που περιλαμβάνονται στο σύνολο που βρίσκεται σε κίνδυνο και όταν αυτά τα άτομα βρίσκονται βεβαίως σε κίνδυνο, δηλαδή, το διάστημα διακινδύνευσης .

Όταν το σύνολο που βρίσκεται σε κίνδυνο είναι απεριόριστο, όλα τα διαστήματα διακινδύνευσης των αντικειμένων που μελετάμε, μπορεί να συμβάλουν στο σύνολο που βρίσκεται σε κίνδυνο για οποιοδήποτε δοσμένο γεγονός, ανεξαρτήτου του αριθμού των γεγονότων που έχει βιώσει κάθε αντικείμενο. Ως εκ τούτου, ο χρόνος εμφάνισης του δεύτερου γεγονότος ενός ατόμου μπορεί να συμβάλει στο σύνολο που βρίσκεται σε κίνδυνο του πρώτου γεγονότος ενός άλλου ατόμου. Για παράδειγμα, σκεφτείτε το δεύτερο γεγονός του B. Στη διατύπωση των ενδιάμεσων χρόνων, το σύνολο που βρίσκεται σε κίνδυνο για το δεύτερο γεγονός του B στο χρονικό σημείο 4 (Σχήμα 2(α)) περιλαμβάνει πληροφορίες από το τρίτο γεγονός του A, από το πρώτο, δεύτερο και τρίτο γεγονός του B και από το πρώτο γεγονός του C. Στη διατύπωση των ολικών χρόνων, η συνεισφορά του συνόλου που βρίσκεται σε κίνδυνο που αντιστοιχεί στο δεύτερο γεγονός του ατόμου B στο χρονικό σημείο 11 (Σχήμα 2(β)) περιλαμβάνει πληροφορίες από το τρίτο γεγονός του A, το δεύτερο και τρίτο γεγονός του B και τον ενιαίο χρόνο του C. Στη διαδικασία καταμέτρησης για το δεύτερο χρόνο εμφάνισης γεγονότος του ατόμου B (Σχήμα 2(γ)) το σύνολο διακινδύνευσης περιλαμβάνει τη συνεισφορά του τρίτου γεγονότος του ατόμου A, του δεύτερου γεγονότος του B και του πρώτου γεγονότος του C.

Ένα απεριόριστο σύνολο που βρίσκεται σε κίνδυνο έχει μια συνηθισμένη βασική συνάρτηση διακινδύνευσης για όλα τα γεγονότα.

Με ένα περιορισμένο σύνολο που βρίσκεται σε κίνδυνο, η συνεισφορά του συνόλου που βρίσκεται σε κίνδυνο του k -στου γεγονότος περιορίζεται στο να περιλαμβάνει μόνο το διάστημα διακινδύνευσης του k -στου γεγονότος αυτών των ατόμων που έχουν βιώσει $(k-1)$ γεγονότα. Για παράδειγμα, μόνο τα άτομα που έχουν ήδη βιώσει τρία γεγονότα θα πρέπει να θεωρούνται ότι βρίσκονται σε κίνδυνο για να βιώσουν το τέταρτο γεγονός. Και στους ενδιάμεσους χρόνους αλλά και στους ολικούς χρόνους, η διατύπωση του συνόλου κινδύνων για το δεύτερο γεγονός του A στο χρονικό σημείο 3 (Σχήματα 2(α) και 2(β)) περιλαμβάνουν πληροφορίες για το δεύτερο γεγονός του A και του B. Για τη διατύπωση της διαδικασίας καταμέτρησης, οι συνεισφορές του σύνολο διακινδύνευσης που συμβάλουν στο δεύτερο γεγονός του ατόμου A στο χρονικό σημείο 5 (σχήμα 2(γ)) περιλαμβάνει μόνο το δεύτερο γεγονός του A. Ένα περιορισμένο σύνολο διακινδύνευσης περιλαμβάνει δηλαδή συγκεκριμένες βασικές συναρτήσεις διακινδύνευσης των γεγονότων.

Τα ημιπεριορισμένα σύνολα που βρίσκονται σε κίνδυνο περιλαμβάνουν προσδιορισμένες βασικές συναρτήσεις διακινδύνευσης των γεγονότων, αλλά επιτρέπουν στα άτομα που έχουν βιώσει λιγότερα από $(k-1)$ γεγονότα να είναι σε κίνδυνο να βιώσουν το k -στο γεγονός μέσω της δημιουργίας ψεύτικων διαστημάτων διακινδύνευσης. Έτσι, ένα αντικείμενο που είχε κανένα ή ένα γεγονός μπορεί να θεωρηθεί ότι είναι σε κίνδυνο να βιώσει ένα τέταρτο γεγονός. Ωστόσο, ένα ημιπεριορισμένο σύνολο που βρίσκεται σε κίνδυνο δεν επιτρέπει στην πληροφορία του διαστήματος διακινδύνευσης του k -στου γεγονότος να συμβάλει στον διακινδύνευσης ενός προγενέστερου γεγονότος. Αυτό το σύνολο διακινδύνευσης ισχύει μόνο για τις διατυπώσεις των ολικών χρόνων και της διαδικασίας καταμέτρησης που περιλαμβάνουν προσδιορισμένες βασικές συναρτήσεις διακινδύνευσης γεγονότων. Για παράδειγμα ο ολικός χρόνος του δεύτερου γεγονότος του B και το ψεύτικο διάστημα διακινδύνευσης του ατόμου C, συμβάλει στο σύνολο διακινδύνευσης για το δεύτερο γεγονός του A (Σχήμα 2,2(β)). Ωστόσο, το πρώτο διάστημα διακινδύνευσης για το άτομο B, δεν μπορεί να συμβάλει στο σύνολο διακινδύνευσης αυτού του γεγονότος εξαιτίας της προσδιορισμένης βασικής συνάρτησης διακινδύνευσης αυτού του γεγονότος.

2.1.4 Διαχείριση της Συσχέτισης των Γεγονότων

Τρεις προσεγγίσεις έχουν προταθεί για να υπολογίσουμε τη συσχέτιση μεταξύ των γεγονότων. Αυτές είναι η περιθώρια, η υπό συνθήκη και των τυχαίων επιδράσεων. Η υπό συνθήκη προσέγγιση θεωρεί ότι τα τωρινά γεγονότα δεν επηρεάζονται από προγενέστερα γεγονότα που συνέβησαν στο αντικείμενο μελέτης μας. Αυτή η υπόθεση μπορεί να "χαλαρώσει" μέσω μιας μεθόδου χαλάρωσης. Αυτό επιτυγχάνεται με την εισαγωγή χρόνο-εξαρτημένων συμμεταβλητών στο μοντέλο, όπως ο αριθμός των εκ των προτέρων επαναλήψεων, οι οποίες μπορεί να αποτυπώνουν την εξάρτηση μεταξύ των επαναλαμβανόμενων χρόνων. Η περιθώρια προσέγγιση υποθέτει ότι τα γεγονότα ενός ατόμου είναι ανεξάρτητα, και εκτιμάει ότι μια εύρωστη διασπορά χρησιμοποιεί μια εκτιμήτρια "σάντουιτς". Οι μη προσαρμοσμένες εκτιμήσεις διασποράς, ονομάζονται αφελείς εκτιμήσεις. Η προσέγγιση τυχαίων επιδράσεων εισάγει μια τυχαία συμμεταβλητή στο μοντέλο που προκαλεί την εξάρτηση μεταξύ των χρόνων των επαναλαμβανόμενων γεγονότων.

Από τις τέσσερις συνιστώσες δύο παίζουν πολύ σημαντικό ρόλο στην κατασκευή ενός μοντέλου. Αυτές είναι το διάστημα διακινδύνευσης και το σύνολο που βρίσκεται σε κίνδυνο. Αυτές επίσης προσδιορίζουν την ερμηνεία του μοντέλου που προκύπτει. Ο προσδιορισμός του συνόλου που βρίσκεται σε κίνδυνο που θα χρησιμοποιηθεί εξαρτάται από μια βασική συνάρτηση που βρίσκεται σε κίνδυνο. Αυτό σημαίνει ότι ένα μη προσδιορισμένο σύνολο που βρίσκεται σε κίνδυνο είναι αναγκαία συνδεδεμένο με μια συνηθισμένη συνάρτηση που βρίσκεται σε κίνδυνο, ενώ ένα ημιπροσδιορισμένο και ένα προσδιορισμένο σύνολο που βρίσκεται σε κίνδυνο είναι συνδεδεμένα με μια συγκεκριμένη συνάρτηση που βρίσκεται σε κίνδυνο.

Τα μοντέλα που θα παρουσιαστούν παρακάτω είναι πέντε γνωστά μοντέλα βασισμένα στο μοντέλο του Cox. Αυτά είναι: AG, PWP-GT, PWP-CP, LWA και WLW. Λεπτομέρειες αυτών των μοντέλων που βασίζονται στο μοντέλο του Cox σε σχέση με τις συνιστώσες που τα προσδιορίζουν, αλλά και τις μερικές πιθανοφάνειες του κάθε μοντέλου, παρουσιάζονται παρακάτω.

2.2 Μοντέλο Andersen-Gill

Το πιο δημοφιλές μοντέλο παλινδρόμησης για την εκτίμηση των συµµεταβλητών της πιθανότητας επιβίωσης, είναι το μοντέλο αναλογικής διακινδύνευσης του Cox (Cox, 1972), στο οποίο υποθέτουμε ότι η συνάρτηση διακινδύνευσης για το i –στο αντικείμενο μελέτης στο χρόνο t δίνεται από τον παρακάτω τύπο:

$$\lambda_i(t) = \lambda_0(t)e^{\beta_0^T Z_i(t)}, \quad (2.2.1)$$

όπου $\lambda_0(t)$ είναι μια απροσδιόριστη μη αρνητική συνάρτηση η οποία αντιπροσωπεύει μια βασική συνάρτηση διακινδύνευσης και το β_0 είναι ένα $p \times 1$ διάνυσμα αγνώστων παραμέτρων. Η εκτίμηση της παραπάνω συνάρτησης επιτυγχάνεται μέσω της συνάρτησης μερικής πιθανοφάνειας (Cox, 1975).

Το μοντέλο των Andersen-Gill είναι μια γενίκευση του μοντέλου αναλογικής διακινδύνευσης του Cox και σχετίζει τη διαδικασία εμφάνισης των επαναλαμβανόμενων γεγονότων μέσω των αντίστοιχων συµµεταβλητών και όλα αυτά μέσω του παρακάτω τύπου:

$$\lambda_{ik}(t) = \lambda_0(t)e^{\beta_0^T Z_i(t)} \quad (2.2.2),$$

για k επαναλήψεις, $k = 1, \dots, K$, όπου $K < \infty$. Ο δείκτης του συνόλου που βρίσκεται σε κίνδυνο για το AG μοντέλο δίνεται από τη σχέση $Y_{ik}(t) = I(X_{i,k-1} < t \leq X_{i,k})$. Παρόλο που τα αντικείμενα μελέτης κατά τη διάρκεια του πειράματος μπορεί να βιώνουν πολλά γεγονότα κανένα αντικείμενο δε βιώνει παραπάνω από ένα γεγονός σε ένα συγκεκριμένο χρονικό διάστημα. Αυτό σημαίνει ότι $dN_i(s) \in \{0,1\}$. Η εκτίμηση των παραμέτρων επιτυγχάνεται μέσω της μερικής πιθανοφάνειας (Cox, 1975). Η εκτιμήτρια του β_0 , είναι $\widehat{\beta}_n^A$, και τον βρίσκουμε από την επαναληπτική επίλυση της εξίσωσης $U_n^A(\beta) = 0_{p \times 1}$, δηλαδή:

$$U_n^A(\beta) = \sum_{i=1}^n \int_0^\tau \{Z_i(s) - E(s; \beta)\} dN_i(s) \quad (2.2.3),$$

όπου $E(s; \beta) = S^{(1)}(s; \beta)/S^{(0)}(s; \beta)$,

$$S^{(r)}(s; \beta) = n^{-1} \sum_{i=1}^n \sum_{k=1}^K Y_{ik}(s) Z_i(s)^{\times r} \times e^{\beta^T Z_i(s)}$$

και, για το διάνυσμα z ισχύει, $z^{\times 0} = 1$, $z^{\times 1} = z$ και $z^{\times 2} = zz^T$. Ο εκτιμητής Breslow-Aalen (Breslow, 1974) για την σωρευτική συνάρτηση διακινδύνευσης, $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$ δίνεται από $\widehat{\Lambda}_0(t; \widehat{\beta}_n^A) = n^{-1} \int_0^t dN.(s)/S^{(0)}(s; \widehat{\beta}_n^A)$, όπου $dN.(s) = \sum_{i=1}^n dN_i(s)$.

Κάτω από συγκεκριμένες συνθήκες κανονικότητας, δηλαδή όταν $n \rightarrow \infty$ $\widehat{\beta}_n^A \xrightarrow{P} \beta_0$, και $n^{\frac{1}{2}}(\widehat{\beta}_n^A - \beta_0) \xrightarrow{D} N_p(0_{p \times 1}, A(\beta_0)^{-1})$, όπου $A_n(\beta) = -\partial U_n^A(\beta)/\partial \beta^T \xrightarrow{P} A(\beta)$, και ένα συνεπή εκτιμητή του πίνακα συνδιασποράς που είναι $\widehat{A}_n(\widehat{\beta}_n^A)$, έχουμε

$$\widehat{A}_n(\beta) = n^{-1} \sum_{i=1}^n \int_0^\tau \left\{ \frac{S^{(2)}(s; \beta)}{S^{(0)}(s; \beta)} - E(s; \beta)^{\times 2} \right\} dN_i(s) \quad (2.2.4).$$

Αν επεκτείνουμε τα αποτελέσματα των Huber (1967) και White (1982) για τα κακώς προσδιορισμένα μοντέλα τότε ο πίνακας εύρωστης συνδιασποράς δίνεται από $\Sigma(\beta_0) = A(\beta_0)^{-1} B(\beta_0) A(\beta_0)^{-1}$ (Lin and Wei, 1989), ο οποίος μπορεί να εκτιμηθεί χρησιμοποιώντας το $\widehat{A}_n(\widehat{\beta}_n^A)$, που ορίστηκε παραπάνω και το $\widehat{B}_n(\widehat{\beta}_n^A)$ όπου

$$\widehat{B}_n(\beta) = n^{-1} \sum_{i=1}^n \left\{ \int_0^\tau [Z_i(s) - E(s; \beta)] d\widehat{M}_i(s; \beta) \right\}^{\times 2} \quad (2.2.5).$$

όπου $d\widehat{M}_i(s; \beta) = dN_i(s) - Y_i(s) e^{\beta^T Z_i(s)} d\widehat{\Lambda}_0(s; \beta)$.

Στο μοντέλο των Andersen-Gill (AG) κάθε άτομο αντιμετωπίζεται ως αριθμητική ανέλιξη ανεξάρτητων προσαυξήσεων (π.χ. , όπως η μη-ομογενής ανέλιξη Poisson (Chiang, 1968)), αφού το $\lambda_{ik}(t)$ είναι ανεξάρτητο του $N_i(t)$ εκτος από εξαρτήσεις που συμβαίνουν από κατάλληλες προσδιορισμένες χρόνο-εξαρτημένες συµμεταβλητές (π.χ ο αριθμός των προηγούμενων γεγονότων). Το AG μοντέλο είναι το πιο απλό μοντέλο τόσο στο να το αντιληφθούμε όσο και να το προγραμματίσουμε. Στην πράξη το AG μοντέλο μπορεί να προσεγγιστεί με ακρίβεια με τη βοήθεια του λογισμικού Poisson παλινδρόμησης ανάλογου της προσέγγισης για το μονομεταβλητό μοντέλο του

Cox (1972) των Laird και Olivier (1981). Σε μια ενδεικτική σύγκριση μερικών μοντέλων επαναλαμβανόμενων γεγονότων, ο Lin (1994) πρότεινε το μοντέλο των Andersen-Gill στο οποίο τον ερευνητή τον νοιάζει μόνο το συνολικό ποσοστό επανεμφάνισης γεγονότων, και όταν για ένα μικρό ποσοστό ατόμων που εξετάζονται ισχύει $N_i(t) \geq 2$.

Αν υποθέσουμε ότι το μοντέλο AG είναι σωστό, η μόνη αλλαγή που έχει, είναι ότι ο ασυμπτωτικός πίνακας συνδιασποράς δίνεται αντί από την γνωστή μας μορφή τύπου “σάντουιτς”. Παρόλα αυτά η υπόθεση των ανεξάρτητων προσαυξήσεων είναι ισχυρή και μπορεί να είναι αβάσιμη. Αν και τα προηγούμενα γεγονότα είναι θετικά συσχετισμένα με τα μελλούμενα γεγονότα, όταν η υπόθεση ανεξαρτησίας απορρίπτεται, τότε η $V(\widehat{\beta}_n)$ υποτιμάται από την αρκετά απλοϊκή διασπορά $A(\beta_0)^{-1}$, και η εύρωστη διασπορά $A(\beta_0)^{-1}B(\beta_0)A(\beta_0)^{-1}$ μπορεί να χρησιμοποιηθεί. Παρόλα αυτά όπως έχει συζητηθεί από τους Lin και Wei (1989) χωρίς τη δομή των ανεξάρτητων προσαυξήσεων, οι παρατηρούμενες πληροφορίες μπορεί να ακολουθούν ένα μοντέλο το οποίο δεν ικανοποιεί την υπόθεση των αναλογικών προσαυξήσεων, κάνοντας το $\widehat{\beta}_n^A$ δύσκολο στο να ερμηνευθεί. Ωστόσο, σε περιπτώσεις στις οποίες το μοντέλο μας είναι κατά προσέγγιση αληθές τότε το $\widehat{\beta}_n^A$ είναι στατιστικά χρήσιμο παρόλο που οι βασικές παραδοχές δεν ισχύουν. Στην πράξη για παράδειγμα τα αληθή μοντέλα αναλογικής διακινδύνευσης σπάνια παρατηρούνται.

Ο Πίνακας 2.2.1 περιέχει τα αποτελέσματα τριών διαφορετικών AG μοντέλων προσαρμοσμένα σε ένα σύνολο δεδομένων άσθματος. Κάθε μοντέλο περιέχει τις συμμεταβλητές πρωταρχικού ενδιαφέροντος οι οποίες είναι: χαμηλό βάρος κατά τη γέννηση (ΧΒΓ), συμπτώματα αναπνευστικής δυσκολίας (ΣΑΔ), παροδική ταχύπνοια νεογέννητων (ΠΤΝ), εκ γενετής ασφυξία (ΕΓΑ) και το φύλλο του ατόμου που εξετάζεται (ΦΑΠΕ). Όλα τα μοντέλα που εξετάζουμε σε αυτό το κεφάλαιο περιέχουν αυτές τις συμμεταβλητές. Πρέπει να σημειώσουμε ότι το HR αντιπροσωπεύει την αναλογία διακινδύνευσης. Και τα τρία μοντέλα του Πίνακα 2.2.1 διαφέρουν ως προς τη μορφή των μεταβλητών που χρειάζονται για την αναπαράσταση του ιστορικού των προηγούμενων γεγονότων. Κοιτώντας το πρώτο μοντέλο (μοντέλο 1) όλες οι συμμεταβλητές εκτός του (ΠΤΝ) έχουν ιδιαίτερα σημαντικές επιπτώσεις στη συνάρτηση έντασης του "άσθματος". Το προηγούμενο ιστορικό γεγονότων που

φαίνεται στο μοντέλο 1 αναπαρίσταται μέσω του αριθμού των προηγούμενων γεγονότων. Ο κίνδυνος άσθματος αυξάνεται μέσω ενός πολλαπλασιαστικού παράγοντα, ο οποίος είναι 1.20 (HR=1.20, $p < 10^{-4}$).

Πίνακας 2.2.1: Ανάλυση δεδομένων προσχολικού άσθματος- Εκτιμημένες αναλογίες διακινδύνευσης βασισμένες στο μοντέλο των Andersen-Gill

Συμμεταβλητή	Μοντέλο 1		Μοντέλο 2		Μοντέλο 3	
	HR	ρ	HR	ρ	HR	ρ
$N_i(t-)$	1.20	$< 10^{-4}$	2.01	$< 10^{-4}$	—	—
$N_i(t-)^2$	—	—	1.98	$< 10^{-4}$	—	—
$I(N_i(t-) = 1)$	—	—	—	—	24.52	$< 10^{-4}$
$I(N_i(t-) = 2)$	—	—	—	—	46.23	$< 10^{-4}$
$I(N_i(t-) \geq 3)$	—	—	—	—	139.67	$< 10^{-4}$
XBG	1.53	$< 10^{-2}$	1.21	0.10	1.04	0.77
ΣΑΔ	1.94	$< 10^{-3}$	1.00	0.98	1.25	0.18
ΠΤΝ	1.37	0.09	1.36	0.04	0.98	0.91
ΕΓΑ	1.42	$< 10^{-2}$	1.29	$< 10^{-3}$	1.14	0.20
ΦΑΠΕ	1.61	$< 10^{-4}$	1.33	$< 10^{-4}$	1.20	$< 10^{-3}$

Όταν ένας τετραγωνικός όρος του προηγούμενου αριθμού των γεγονότων προστίθεται στο ιστορικό των γεγονότων (μοντέλο 2), οι εκτιμημένες αναλογίες διακινδύνευσης για τις προαναφερθείσες συμμεταβλητές μειώνονται. Σε ορισμένες περιπτώσεις μάλιστα, η πτώση είναι δραματική. Για παράδειγμα, για το (ΦΑΠΕ) ισχύει HR=1.61 βασισμένο στο μοντέλο 1, όμως πέφτει στο HR=1.33 στο μοντέλο 2. Αξίζει επίσης να σημειωθεί ότι τα (XBG) και (ΣΑΔ) δεν είναι στατιστικά σημαντικά ($p > 0.05$) βασισμένα στο μοντέλο 2 πάντα. Τα χαρακτηριστικά του μοντέλου 3, του πιο ευέλικτου μοντέλου που περιγράφει το ιστορικό των προηγούμενων γεγονότων, χρησιμοποιούν δείκτριες συναρτήσεις, σε αντίθεση με την υπόθεση ότι υπάρχει λογαριθμογραμμική σχέση με την αρίθμηση των προηγούμενων γεγονότων. Αυτό έχει ως αποτέλεσμα το (ΦΑΠΕ) να γίνεται στατιστικά σημαντικό. Τα εκτιμημένα αποτελέσματα για το (ΦΑΠΕ) πέφτουν επιπλέον στο HR=1.20, το οποίο είναι ιδιαίτερα σημαντικό.

Το προηγούμενο παράδειγμα απεικονίζει έναν περιορισμό του AG μοντέλου. Αυτός είναι ότι είναι δύσκολο να μοντελοποιήσουμε το ιστορικό των προηγούμενων γεγονότων. Επιπλέον, δεν μπορούμε πότε να είμαστε σίγουροι αν το ιστορικό των προηγούμενων γεγονότων έχει προσδιοριστεί

σωστά. Είναι δύσκολο γενικά να ερμηνευτούν οι διαφορές στις εκτιμημένες παραμέτρους, που συμβαίνουν κατά τη χρησιμοποίηση διαφορετικών μορφών για να αναπαραστήσουμε το ιστορικό των γεγονότων. Ουσιαστικά, μοντελοποιούμε τη διαδικασία των γεγονότων, αναπροσαρμόζοντας παράλληλα τις συμεταβλητές που καθορίζονται από τις προηγούμενες τιμές της διαδικασίας των γεγονότων.

2.3 Μοντέλο PWP

Η πρώτη επέκταση του μοντέλου του Cox σε πολλαπλά δεδομένα γεγονότων δημιουργήθηκε από τον Prentice (1981). Σε αυτή την παράγραφο θα θεωρήσουμε δύο είδη PWP μοντέλων που χρησιμοποιούνται πολύ στην πράξη. Πιο συγκεκριμένα η συνάρτηση έντασης για το αντικείμενο i στο χρόνο t για k επαναλήψεις, δεδομένου του $N_i(t)$, παίρνει την ακόλουθη μορφή:

$$\lambda_{ik}(t) = Y_{ik}(t)\lambda_{0k}(t)e^{\beta_k^T Z_{ik}(t)} \quad (2.3.1)$$

$$\lambda_{ik}(t) = Y_{ik}(t)\lambda_{0k}(t - T_{i,k-1})e^{\beta_k^T Z_{ik}(t)} \quad (2.3.2)$$

Η πρώτη εξίσωση είναι για τους ολικούς χρόνους και η δεύτερη για τους ενδιάμεσους χρόνους. Ουσιαστικά αυτή η προσέγγιση παράγει ένα στρωματοποιημένο μοντέλο αναλογικής έντασης με χρόνο-εξαρτημένα στρωματοποιημένα δεδομένα, στο οποίο η εξάρτηση μεταξύ των χρόνων των γεγονότων επιτυγχάνεται στρωματοποιώντας τον προηγούμενο αριθμό των γεγονότων. Στο μοντέλο των ολικών χρόνων (PWP-TT) οι δείκτες του συνόλου που βρίσκεται σε κίνδυνο ορίζονται ακριβώς με τον ίδιο τρόπο όπως ορίστηκαν και για το μοντέλο AG (δηλαδή, $Y_{ik}(t) = I(X_{i,k-1} \leq t < X_{i,k})$). Για να αναλύσουμε τους ενδιάμεσους χρόνους, θέτουμε $Y_{ik}(t) = I(X_{i,k} \geq X_{i,k-1} + t)$ και αντικαθιστώ το $Z_{ik}(t)$ με $Z_{ik}(X_{i,k-1} + t)$. Αντίθετα με την προσέγγιση του AG μοντέλου, στα μοντέλα PWP οι προσδιορισμένοι παράμετροι παλινδρόμησης των γεγονότων και οι βασικές συναρτήσεις έντασης επιτρέπονται. Κατά συνέπεια όταν χρησιμοποιούμε ένα μοντέλο PWP είναι σημαντικό να κατασκευάσουμε ένα $Z_{ik}(t)$ ούτως ώστε η επίδραση του $N_i(t)$ στο $\lambda_{ik}(t)$ να φαίνεται με ακρίβεια. Οι παράμετροι

εκτιμώνται μέσω της μερικής πιθανοφάνειας. Για το μοντέλο (2.3.2) η παράμετρος παλινδρόμησης β_k εκτιμάται αν λυθεί η εκτιμώμενη εξίσωση $U_n^{PT}(\beta_k) = 0_{p_k \times 1}$, δηλαδή:

$$U_n^{PT}(\beta_k) = \sum_{i=1}^n \int_0^\tau \{Z_{ik}(s) - E_k^{PT}(s; \beta_k)\} dN_{ik}(s) \quad (2.3.3),$$

για $i = 1, \dots, K$,

$$\begin{aligned} \text{όπου } E_k^{PT}(s; \beta_k) &= Q_k^{(1)}(s; \beta_k) / Q_k^{(0)}(s; \beta_k), \\ Q_k^{(r)}(s; \beta_k) &= n^{-1} \times \sum_{i=1}^n Y_{ik}(s) Z_{ik}(s)^{\times r} e^{\beta_k^T Z_{ik}(s)} \\ \text{και } N_{ik}(t) &= I(T_{i,k} \leq t, \Delta_{ik} = 1). \end{aligned}$$

Για το PWP μοντέλο στη σχέση (3) ο εκτιμητής μερικής πιθανοφάνειας μπορεί να υπολογιστεί λύνοντας την $U_n^{PG}(\beta_k) = 0_{p_k \times 1}$ δηλαδή:

$$U_n^{PG}(\beta_k) = \sum_{i=1}^n \int_0^\tau \{Z_{ik}(s + T_{i,k-1}) - E_k^{PG}(s; \beta_k)\} d\tilde{N}_{ik}(s) \quad (2.3.4),$$

$$\begin{aligned} \text{όπου } E_k^{PG}(s; \beta_k) &= R_k^{(1)}(s; \beta_k) / R_k^{(0)}(s; \beta_k), \\ R_k^{(r)}(s; \beta_k) &= n^{-1} \times \sum_{i=1}^n Y_{ik}(s) Z_{ik} \times (T_{i,k-1} + s)^{\times r} e^{\beta_k^T Z_{ik}(T_{i,k-1} + s)} \\ \text{και } \tilde{N}_{ik}(t) &= I(\tilde{T}_{ik} \leq t, \Delta_{ik} = 1). \end{aligned}$$

Για την προσαρμογή των PWP μοντέλων πρέπει να σημειώσουμε ότι το K συνήθως επιλέγεται έτσι ώστε μια επαρκής ποσότητα εμπειρίας να μπορεί να παρατηρηθεί από το σύνολο δεδομένων το οποίο αναλύουμε.

Η ερμηνεία των αποτελεσμάτων συνήθως σε αυτά τα μοντέλα δυσχεραίνεται από την υπό συνθήκη φύση των μοντέλων. Η κύρια δυσκολία της ανάλυσης αυτών των παραμέτρων που βασίζονται σε κάποιο σύνολο που βρίσκεται σε κίνδυνο, είναι ότι η υπόθεση της ελλιπής τυχαίας ολοκληρωσιμότητας (ETO) παραβιάζεται από το γεγονός ότι τα άτομα που δεν έχουν βιώσει k γεγονότα εξαιρούνται από την ανάλυση, αναφορικά βέβαια με τη συνάρτηση έντασης των $k + 1$ γεγονότων. Παρόλα αυτά, η ανάλυση που βασίζεται σε μη περιορισμένα σύνολα διακινδύνευσης, μπορεί να αξιολογηθεί σε σχέση με την επίδραση “μεταφοράς υπολοίπου”. Οι προσδιορισμένοι ολικοί

χρόνοι ενός ατόμου τείνουν να είναι υψηλά συσχετισμένοι ακόμα και όταν οι ενδιάμεσοι χρόνοι δεν είναι υψηλά συσχετισμένοι, και αυτό διότι οι ολικόι χρόνοι για τα $k + 1$ διαστήματα διακινδύνευσης περιέχουν όλα τα προηγούμενα διαστήματα που βρίσκονται σε κίνδυνο (δηλαδή τα $1, \dots, k$). Όταν η επίδραση μιας συμμεταβλητής υφίσταται μόνο για το αρχικό γεγονός, αλλά όχι για τις επόμενες επαναλήψεις, τότε μια ανάλυση βασισμένη στους ολικούς χρόνους μπορεί εσφαλμένα να εκτιμήσει ότι το δεύτερο γεγονός, το τρίτο γεγονός κ.ο.κ ασκούν πολύ μεγάλη επίδραση στην θεραπεία και όλα αυτά εξαιτίας του φαινομένου “μεταφοράς υπολοίπου”. Η επιλογή της κλίμακας του χρόνου και η κατασκευή του συνόλου που βρίσκεται σε κίνδυνο εξαρτάται αποκλειστικά από τους στόχους του ερευνητή. Στο PWP μοντέλο όπως και στο μοντέλο των Andersen και Gill (1982) θεωρούμε ότι η πληροφορία στο $N_i(t)$ αποτυπώνεται στο διάνυσμα συνδιασποράς. Μια σωστή εκτίμηση της διασποράς μπορεί να πραγματοποιηθεί μέσω της εκτιμήτριας Sandwich.

Πίνακας 2.3.1: Ανάλυση δεδομένων προσχολικού άσθματος- Εκτιμημένες αναλογίες διακινδύνευσης βασισμένες στο μοντέλο ολικών χρόνων των Prentice, Williams και Petersen.

Συμμεταβλητή	1η Επίσκεψη		2η Επίσκεψη		3η Επίσκεψη	
	HR	ρ	HR	ρ	HR	ρ
$T_{i,1}$	—	—	0.98	$< 10^{-4}$	1.02	0.02
$T_{i,2}$	—	—	—	—	0.94	$< 10^{-4}$
ΧΒΓ	1.50	$< 10^{-3}$	1.11	0.57	1.01	0.98
ΣΑΔ	1.92	$< 10^{-4}$	1.47	0.09	1.45	0.20
ΠΤΝ	1.39	0.03	0.94	0.77	1.37	0.26
ΕΓΑ	1.23	0.02	1.23	0.09	0.96	0.80
ΦΑΠΕ	1.67	$< 10^{-4}$	1.23	$< 10^{-2}$	1.06	0.57

Ο Πίνακας 2.3.1 περιέχει τις εκτιμημένες αναλογίες διακινδύνευσης βασισμένες στο μοντέλο ολικών χρόνων του Prentice (1981). Όλες οι συμμεταβλητές καταδεικνύουν μια αξιόλογη σχέση που έχουν όλες οι πρώτες φυσικές επισκέψεις για έλεγχο άσθματος. Μέχρι τη χρονική στιγμή της δεύτερης επίσκεψης, μόνο το (ΦΑΠΕ) είναι στατιστικά σημαντικό όμως η επίδραση αυτής της συμμεταβλητής έχει μειωθεί πάρα πολύ σε σχέση με την πρώτη επίσκεψη. Καμία συμμεταβλητή δεν επηρεάζει τον ολικό χρόνο στην

τρίτη επίσκεψη. Για το δεύτερο και το τρίτο γεγονός, πρέπει να σημειώσουμε ότι υπάρχει στατιστικά σημαντική σχέση με τους χρόνους των γεγονότων που προηγούνται.

Πίνακας 2.3.2: Ανάλυση δεδομένων προσχολικού άσθματος- Εκτιμημένες αναλογίες διακινδύνευσης βασισμένες στο μοντέλο ενδιάμεσων χρόνων των Prentice, Williams και Petersen.

Συμμεταβλητή	1η Επίσκεψη		2η Επίσκεψη		3η Επίσκεψη	
	HR	p	HR	p	HR	p
$T_{i,1}$	—	—	1.01	0.05	1.00	0.44
$T_{i,2}$	—	—	—	—	0.98	$< 10^{-2}$
ΧΒΓ	1.50	$< 10^{-3}$	1.11	0.55	1.02	0.89
ΣΑΔ	1.92	$< 10^{-4}$	1.34	0.14	1.35	0.20
ΠΤΝ	1.39	0.03	0.90	0.59	1.24	0.32
ΕΓΑ	1.23	0.02	1.16	0.18	1.12	0.44
ΦΑΠΕ	1.67	$< 10^{-4}$	1.18	0.03	1.09	0.33

Στο PWP μοντέλο των ενδιάμεσων χρόνων του Πίνακα 2.3.2 βλέπουμε μερικές ομοιότητες. Όπως είναι φυσικό τα αποτελέσματα είναι ίδια για το γεγονός 1, αφού οι ολικοί και οι ενδιάμεσοι χρόνοι είναι ίσοι για την πρώτη επίσκεψη. Τα ίδια αποτελέσματα μπορούν να επιτευχθούν μέσω ενός μοντέλου παλινδρόμησης του Cox. Για τους ενδιάμεσους χρόνους της δεύτερης επίσκεψης, μόνο το (ΦΑΠΕ) παραμένει στατιστικά σημαντικό. Για τους ενδιάμεσους χρόνους της τρίτης επίσκεψης καμία συμμεταβλητή δεν είναι στατιστικά σημαντική. Κρίνοντας από το μέγεθος των συντελεστών παλινδρόμησης και από τις p τιμές τους, ο βαθμός σύνδεσης μεταξύ των ολικών και των ενδιάμεσων χρόνων είναι ισχυρός.

Είναι δύσκολο να ερμηνεύσουμε τη μείωση σημαντικότητας στο φύλο ενός ατόμου και των συνθηκών γέννησης του όσο αυξάνεται ο αριθμός των γεγονότων. Το γεγονός ότι το σύνολο που βρίσκεται σε κίνδυνο χρησιμοποιείται στο πρόβλημα είναι το κλειδί. Μια υπόθεση είναι ότι όσο προχωράμε στο πείραμα από γεγονός σε γεγονός, τα παιδιά που είναι υπό παρακολούθηση είναι σταδιακά λιγότερο υγιή. Υπό αυτή τη συνθήκη τα αποτελέσματα των συμμεταβλητών καθορίζονται από μια σταθερά αυξανόμενη βασική συνάρτηση διακινδύνευσης. Μια άλλη υπόθεση είναι ότι

οι συνθήκες γέννησης επηρεάζουν μόνο το χρόνο μέχρι την εκδήλωση άσθματος, όχι όμως το χρόνο εμφάνισης των μετέπειτα γεγονότων.

2.4 Μοντέλο WLW

Παρακινούμενος από την έλλειψη αξιοπιστίας των μοντέλων των AG και PWP, ο Wei (1989) πρότεινε τη μοντελοποίηση των περιθώριων κινδύνων ως εναλλακτική της μοντελοποίησης της συνάρτησης έντασης υπό την υπόθεση του $N_i(t)$. Η προσέγγιση του WLW συνεπάγεται τη μοντελοποίηση της περιθώριας κατανομής του k -οστού χρονικού γεγονότος χρησιμοποιώντας τις προδιαγραφές του μοντέλου του Cox για τη συνάρτηση διακινδύνευσης :

$$\lambda_{ik}(t) = \lambda_{0k}(t) e^{\beta_k^T Z_{ik}(t)} \quad (2.4.1)$$

για $k = 1, \dots, K$. Η συνάρτηση μερικής πιθανοφάνειας του k -στού γεγονότος δίνεται από:

$$PL_k \beta_k = \prod_{i=1}^n \left\{ \frac{e^{\beta_k^T Z_{ik}(t)(X_{ik})}}{\sum_{j=1}^n Y_{jk}(X_{ik}) e^{\beta_k^T Z_{jk}(t)}} \right\}^{\Delta_{ik}} \quad (2.4.2)$$

με αντίστοιχη συνάρτηση αποτελεσμάτων:

$$U_{k:n}(\beta_k) = \frac{\partial}{\partial \beta_k} \log PL_k(\beta_k) = \sum_{i=1}^n \{Z_{ik}(s) - E_k(s; \beta_k)\} dN_{ik}(s) \quad (2.4.3)$$

$$\text{όπου } \Delta_{ik} = I(T_{i,k} \leq C_i),$$

$$N_{ik}(t) = I(X_{i,k} \leq t, \Delta_{ik} = 1),$$

$$E_k(s; \beta_k) = S_k^{(1)}(s; \beta_k) / S_k^{(0)}(s; \beta_k),$$

$$S_k^{(r)}(s; \beta_k) = n^{-1} \times \sum_{i=1}^n Y_{ik}(s) Z_{ik}(s)^{\times r} e^{\beta_k^T Z_{ik}(s)}, \text{ για } r = 0, 1, 2$$

$$\text{και } Y_{ik}(t) = I(X_{i,k} \geq t).$$

Η βασική σωρευτική συνάρτηση διακινδύνευσης για το k -στο γεγονός εκτιμάται από τη σχέση $\widehat{\Lambda}_{0k}(t; \widehat{\beta}_{k:n}) = n^{-1} \int_0^t dN_{.k}(s) / S_k^{(0)}(s; \widehat{\beta}_{k:n})$,

$$\text{με } dN_{\cdot k}(s) = \sum_{i=1}^n dN_{ik}(s).$$

Θέτουμε $X_i = (X_{i,1}, \dots, X_{i,K})^T$ και $\Delta_i = (\Delta_{i,1}, \dots, \Delta_{i,K})^T$. Ανάμεσα στις άλλες συνθήκες κανονικότητας που παραθέτονται από τον Wei (1989), υποθέτουμε ότι τα (Z_i, X_i, Δ_i) είναι ανεξάρτητα και ισόνομα κατανομημένα. Η λύση της παραπάνω συνάρτησης αποτελεσμάτων, $\widehat{\beta}_{k:n}$, είναι σύμφωνη με το β_k , υπό την προϋπόθεση ότι τα περιθώρια μοντέλα έχουν καθοριστεί σωστά. Καθώς

$$n \rightarrow \infty, \quad n^{\frac{1}{2}}(\widehat{\beta}_{k:n} - \beta_0) \xrightarrow{D} N_p(0_{p \times 1} A_k(\beta_k)^{-1} B_k(\beta_k) A_k(\beta_k)^{-1}),$$

τότε ένας συνεπής εκτιμητής της ασυμπτωτικής διασποράς λαμβάνεται μέσω των $\widehat{A}_{k:n}(\widehat{\beta}_{k:n})$ και $\widehat{B}_{k:n}(\widehat{\beta}_{k:n})$, όπου:

$$\widehat{A}_{k:n}(\beta_k) = n^{-1} \sum_{i=1}^n \int_0^\tau \left\{ \frac{S_k^{(2)}(s; \beta_k)}{S_k^{(0)}(s; \beta_k)} - \left(\frac{S_k^{(1)}(s; \beta_k)}{S_k^{(0)}(s; \beta_k)} \right)^{\times 2} \right\} dN_{ik}(s) \quad (2.4.4).$$

$$\widehat{B}_{k:n}(\beta_k) = n^{-1} \sum_{i=1}^n \left\{ \int_0^\tau [Z_{ik}(s) - E_k(s; \beta_k)] d\widehat{M}_{ik}(s; \beta_k) \right\}^{\times 2} \quad (2.4.5)$$

με $d\widehat{M}_{ik}(s; \beta_k) = dN_{ik}(s) - Y_{ik}(s) e^{\beta_k^T Z_{ik}(s)} d\Lambda_{0k}(s; \beta_k)$. Πρέπει εδώ να σημειώσουμε ότι αν το περιθώριο μοντέλο είναι σωστά καθορισμένο, και τα γεγονότα για το ίδιο αντικείμενο μελέτης είναι ασυσχέτιστα, τότε το $\widehat{A}_{k:n}(\widehat{\beta}_{k:n})$ είναι ασυμπτωτικά ισοδύναμο με το $\widehat{B}_{k:n}(\widehat{\beta}_{k:n})$.

Οι παραπάνω τύποι ανακύπτουν στην περίπτωση των παραμέτρων παλινδρόμησης συγκεκριμένων γεγονότων, οι οποίοι επιτρέπουν την εξέταση της τάσης των επιπτώσεων κάθε επόμενη φορά. Στις περιπτώσεις για τις οποίες ισχύει $Z_{ik}(s) = Z_i(s)$ για $k = 1, \dots, K$, το μέσο αποτέλεσμα μιας συμμεταβλητής μπορεί να εκτιμηθεί προσαρμόζοντας ένα μοντέλο που χρησιμοποιεί μια συνηθισμένη παράμετρο παλινδρόμησης (π.χ. $\beta_k = \beta_0$, για $k = 1, \dots, K$). Στην πράξη, συγκεκριμένα στρωματοποιημένα δεδομένα μπορεί να χρειαστεί να συνυπολογισθούν. Για παράδειγμα εάν πολύ λίγα άτομα βιώνουν το μέγιστο αριθμό των γεγονότων για να προκαλέσουν τη σταθερότητα στο $\widehat{\beta}_{k:n}$ και να δικαιολογήσουν τις ασυμπτωτικές προσεγγίσεις.

Οι μέθοδοι των AG, PWP και WLW έχουν συγκριθεί από πάρα πολλούς συγγραφείς (όπως, Lin, 1994, Gao και Zhou, 1997, Clayton, 1994,

Therneau και Hamilton, 1997, Wei και Glidden, 1997) χρησιμοποιώντας πραγματικά και προσομοιωμένα δεδομένα, και είναι γνωστό ότι αυτά τα μοντέλα συχνά αποφέρουν διαφορετικά αποτελέσματα για το ίδιο σύνολο δεδομένων. Αυτό δεν είναι ένα απροσδόκητο εύρημα, δεδομένου ότι τα διαφορετικά μοντέλα απαιτούν διαφορετικές ερευνητικές ερωτήσεις.

Η μέθοδος WLW και οι υποπεριπτώσεις της (Wei κ.α., 1990, Lee κ.α., 1992, Liang κ.α., 1993, Cai και Prentice, 1995 και 1997) είναι ισχυρές και καλά ανεπτυγμένες θεωρητικά. Θεωρούνται ως εξαιρετικά λεπτομερείς διατάξεις για την εξαγωγή συμπερασμάτων σχετικά με τα αποτελέσματα της συμμεταβλητής "μέσος όρος του πληθυσμού". Ωστόσο, δεν παρέχουν καμία πληροφορία σχετικά με τις σχέσεις μεταξύ των χρόνων αποτυχίας.

Υπήρξε μεγάλη συζήτηση στη βιβλιογραφία σχετικά με το αν η μέθοδος WLW είναι κατάλληλη, κατ'αρχήν, για τα επαναλαμβανόμενα γεγονότα. Όταν η μέθοδος προτάθηκε, ο Wei κ.α. (1989), πρότεινε δύο ερευνητικές ρυθμίσεις στο ίδιο πλαίσιο. Αυτές ήταν:

- i) ομαδοποιημένα αντικείμενα, με ≤ 1 γεγονότα ανά αντικείμενο
- ii) ανεξάρτητα αντικείμενα με περισσότερα από ένα γεγονότα ανά αντικείμενο

Θεωρητικά, η μέθοδος WLW μπορεί βασίμως να εφαρμοστεί και στις δύο ρυθμίσεις. Στο παρελθόν έχει συχνά χρησιμοποιηθεί και προταθεί από πολλούς συγγραφείς (π.χ., Lin, 1994 Therneau και Hamilton, 1997, Wei και Glidden, 1997, Barai και Teoh, 1997, Kelly και Lim, 2000) για τα δεδομένα επαναλαμβανόμενων γεγονότων. Από την άλλη η μέθοδος WLW έχει επικριθεί όσον αφορά την ερμηνεία των εκτιμήσεων των παραμέτρων παλινδρόμησης. Δεδομένου ότι το $\lambda_{ik}(t)$ είναι ένας περιθώριος κίνδυνος, τα αντικείμενα είναι σε κίνδυνο για το $(k+1)$ γεγονός δεδομένου ότι έχουν βιώσει k γεγονότα. Οι συναρτήσεις διακινδύνευσης προσδιορισμένων γεγονότων μπορεί να είναι ιδιαίτερα αλληλοεξαρτώμενες. Χαμηλές τιμές που έχουν παρατηρηθεί για το $A_{ik}(t) = \int_0^t \lambda_{ik}(s) ds$ συνεπάγονται χαμηλό $A_{ik+1}(t)$ αφού $e^{A_{ik+1}(t)} \geq e^{A_{ik}(t)}$, χρησιμοποιώντας την συνάρτηση επιβίωσης και το γεγονός ότι το k -στο γεγονός προηγείται του $(k+1)$ γεγονότος. Συγγραφείς όπως οι Kelly και Lim (2000) υποστηρίζουν ότι η μέθοδος WLW υπερεκτιμά τους συντελεστές παλινδρόμησης λόγω του φαινομένου "μεταφοράς υπολοίπου" που

περιγράφηκε προηγουμένως. Οι Cook και Lawless (1997) σχολίασαν την λογική ασυνέπεια του ορισμού των αντικειμένων που βρίσκονται σε κίνδυνο για το $(k+1)$ γεγονός πριν βιώσουν το k -στο γεγονός, μια παράδοξη κατάσταση, που απορρέει φυσιολογικά από τον τύπο των WLW.

Πίνακας 2.4.1: Ανάλυση δεδομένων προσχολικού άσθματος- Εκτιμημένες αναλογίες διακινδύνευσης βασισμένες στο μοντέλο των Wei, Lin και Weissfeld.

Συμμεταβλητή	1ο γεγονός		2ο γεγονός		3ο γεγονός	
	HR	p	HR	p	HR	p
ΧΒΓ	1.50	$< 10^{-3}$	1.51	$< 10^{-3}$	1.58	$< 10^{-2}$
ΣΑΔ	1.92	$< 10^{-4}$	2.36	$< 10^{-4}$	2.68	$< 10^{-4}$
ΠΤΝ	1.39	0.03	1.32	0.18	1.52	0.07
ΕΓΑ	1.23	0.02	1.35	$< 10^{-2}$	1.44	$< 10^{-2}$
ΦΑΠΕ	1.67	$< 10^{-4}$	1.86	$< 10^{-4}$	1.95	$< 10^{-4}$

Η ανάλυση των δεδομένων προσχολικού άσθματος βασισμένη στο μοντέλο των Wei, Lin και Weissfeld παρουσιάζεται στον Πίνακα 2.4.1. Φυσικά οι παράμετροι εκτίμησης που αντιστοιχούν στην πρώτη επίσκεψη είναι ίδιες με εκείνες που αντιστοιχούν στο πρώτο γεγονός των Πινάκων 2.3.1 και 2.3.2, με όλες τις συμμεταβλητές να έχουν σημαντικές επιπτώσεις. Συγκρίνοντας τα αποτελέσματα μεταξύ των γεγονότων 1-3 τα HR για κάθε συμμεταβλητή αυξάνονται γενικά για κάθε διαδοχικό γεγονός, λαμβάνοντας υπόψη τα αποτελέσματα του PWP μοντέλου (Πίνακες 2.3.1 και 2.3.2), απεικονίζει το φαινόμενο μεταφοράς υπολοίπου που περιγράφηκε προηγουμένως. Για παράδειγμα, σκεφτείτε τη συμμεταβλητή ΕΓΑ. Με βάση το μοντέλο WLW, η εκτιμώμενη αναλογία διακινδύνευσης αυξάνεται από HR = 1,23 (1ο γεγονός) σε HR = 1,35 (2ο γεγονός) και σε HR = 1,44 (3ο γεγονός). Ωστόσο, με βάση τα μοντέλα PWP, φαίνεται ότι η ΕΓΑ δεν έχει σημαντική επίδραση πέρα από το πρώτο γεγονός. Το μοτίβο είναι γενικά ίδιο και για τις υπόλοιπες συμμεταβλητές.

2.5 Μοντέλο των Pepe και Cai

Οι Pepe και Cai (1993) ανέπτυξαν μια προσέγγιση που μπορεί να θεωρηθεί ενδιάμεση μεταξύ της υπό συνθήκη έντασης και των προσεγγίσεων περιθώριων κινδύνων. Πρότειναν μοντελοποίηση της συνάρτησης έντασης $\{r_{i1}(t), r_{i2}(t), \dots\}$, όπου $r_{ik}(t)$ είναι το ρυθμός (π.χ., η μέση ένταση) εμφάνισης του k -στου γεγονότος μεταξύ των ατόμων που βρίσκονται σε κίνδυνο στο χρόνο t και έχουν βιώσει $(k-1)$ γεγονότα. Το μοντέλο αυτό είναι:

$$r_{ik}(t) = \lim_{\delta \rightarrow 0} \frac{1}{\delta} P(t < T_{i,k} \leq t + \delta | T_{i,k} \geq t, T_{i,k-1} < t) \quad (2.5.1).$$

Παρόμοια με την περιθώρια προσέγγιση, οι συσχετίσεις των ενδοατομικών χρόνων γεγονότων είναι μη προσδιορισμένες. Η συνάρτηση $r_{ik}(t)$ είναι καλώς ορισμένη, και ανεξάρτητη από τις συσχετίσεις των χρόνων των γεγονότων, και επίσης παρουσιάζει συχνά ενδιαφέρον για τους ερευνητές. Σε αντίθεση με τις περιθώριες συναρτήσεις διακινδύνευσης WLW $\{\lambda_{ik}(t): k = 1, 2, \dots\}$, οι συναρτήσεις ρυθμού $\{r_{ik}(t): k = 1, 2, \dots\}$ εξαρτώνται από το γεγονός αν τα αντικείμενα της μελέτης έχουν βιώσει $(k-1)$ γεγονότα η οποία είναι μια πιο διαισθητική προσέγγιση για ένα επαναλαμβανόμενο γεγονός. Επιπλέον, λαμβάνοντας υπόψη ότι τα $\lambda_{i1}(t), \lambda_{i2}(t), \dots$ είχαν μια προαναφερθείσα αριθμητική σχέση το ένα με το άλλο, τα $r_{i1}(t), r_{i2}(t), \dots$ δεν έχουν καμία εγγενή σχέση το ένα με το άλλο. Έτσι κάθε υπό συνθήκη ρυθμός μπορεί να θεωρηθεί ως σύνοψη διαφορετικών συνιστωσών του συνόλου δεδομένων. Κάποιος θα μπορούσε να μοντελοποιήσει κάθε υπό συνθήκη ποσοστό χρησιμοποιώντας κάποια μορφή του μοντέλου του Cox, όπως $r_{ik}(t) = r_{0k}(t)e^{\beta_k^T Z_{ik}(t)}$ (όπου $\{r_{0k}(t): k = 1, 2, \dots\}$ είναι αυθαίρετες μη αρνητικές συναρτήσεις). Οι εκτιμήτριες συγκεκριμένων στιγμών, βασικών συναρτήσεων ρυθμού δίνονται από $\hat{r}_{0k}(t; \hat{\beta}_{k:n}) = \frac{dN_{k:n}}{S_k^{(0)}(t; \hat{\beta}_{k:n})}$, η οποία έχει την ίδια μορφή με αυτή του $\tilde{\lambda}_{0k}(t; \tilde{\beta}_{k:n})$ με το $Y_{ik}(t)$ να έχει τροποποιηθεί αναλόγως. Η βαθμιδωτή συνάρτηση είναι η εξής:

$$U_n^k(\beta_k) = \sum_{i=1}^n \int_0^\tau Z_i(s) \{dN_{ik}(s) - Y_{ik}(s) \hat{r}_{0k}(s) e^{\beta_k^T Z_{ik}(s)} ds\} \quad (2.5.2),$$

η οποία παρουσιάζει έντονη αντιστοιχία με την εκτιμήτρια μερικής πιθανοφάνειας που αναφέρθηκε προηγουμένως.

Πίνακας 2.5.1: Ανάλυση δεδομένων προσβολικού άσθματος- Εκτιμημένες ποσοστιαίες αναλογίες βασισμένες στο μοντέλο των Pepe και Cai.

Συμμεταβλητή	1η Επίσκεψη		2η Επίσκεψη		3η Επίσκεψη	
	RR	p	RR	p	RR	p
ΧΒΓ	1.50	$< 10^{-3}$	1.01	0.57	1.25	0.23
ΣΑΔ	1.92	$< 10^{-4}$	1.57	0.04	1.37	0.19
ΠΤΝ	1.39	0.03	0.97	0.89	1.28	0.29
ΕΓΑ	1.23	0.02	1.21	0.10	1.10	0.48
ΦΑΠΕ	1.67	$< 10^{-4}$	1.26	$< 10^{-2}$	1.16	0.09

Ο Πίνακας 2.5.1 εκτιμάει τις αναλογίες ρυθμού (RR) για την ανάλυση των δεδομένων του άσθματος με βάση το μοντέλο έντασης των Pepe και Cai. Όλες οι συμμεταβλητές σχετίζονται σημαντικά με τον χρόνο μέχρι την πρώτη επίσκεψη που οφείλεται στο άσθμα. Για τα ΦΑΠΕ, ΕΓΑ και ΣΑΔ, το RR μειώνεται σταθερά για το δεύτερο και τρίτο γεγονός. Και πάλι, δεδομένου ότι τα σύνολα διακινδύνευσης περιορίζονται βάσει αυτού του μοντέλου, αυτό μπορεί να αντανακλά την ανομοιογένεια μεταξύ του συνόλου που βρίσκεται σε κίνδυνο που αντιστοιχεί σε διαδοχικά επιτυχημένα γεγονότα.

2.6 Επαναλαμβανόμενα Γεγονότα: Χαρακτηριστικά και Μοντέλα

Είναι γνωστό ότι οποιαδήποτε συσχέτιση μεταξύ των γεγονότων που παράγονται είτε μεμονωμένα είτε από κοινού μέσω της ετερογένειας και της εξάρτησης των γεγονότων παραβιάζει την υπόθεση του μοντέλου του Cox ότι δηλαδή ο χρόνος εκδήλωσης των γεγονότων είναι ανεξάρτητος. Το μοντέλο του Cox είναι βασισμένο στα τυπικά πλαίσια των επαναλαμβανόμενων γεγονότων (Kelly και Lim, 2000, Aalen, 1988, Lawless και Nadeau 1995,

Nelson, 2003 και Andersen και Gill, 1982). Παραλλαγές του μοντέλου του Cox, δηλαδή τα μοντέλα διορθωμένης διασποράς και τα μοντέλα ευπάθειας έχουν προταθεί για την εκτίμηση των επαναλαμβανόμενων γεγονότων.

Τα μοντέλα διορθωμένης διασποράς παρουσιάζουν ένα τρόπο για να αντιμετωπίσουμε τα προβλήματα αποτελεσματικότητας που παράγονται από την ετερογένεια ανάμεσα στα άτομα που εξετάζονται. Ένα υποσύνολο αυτών των μοντέλων επιχειρεί να ενσωματώσει την εξάρτηση των γεγονότων που επιτρέπει οι βασικοί κίνδυνοι να ποικίλλουν ανάλογα με τον αριθμό των γεγονότων που κάθε άτομο βιώνει. Αυτές οι διαφορές εξηγούνται διότι τα μοντέλα διασποράς ενσωματώνουν την ανομοιογένεια στις εκτιμήσεις τους και ως εκ τούτου παραμένουν μεροληπτικές. Η έρευνα που χρησιμοποιεί προσομοιώσεις για να εξετάσει τις εκτιμήσεις των αποτελεσμάτων της θεραπείας για ένα υποσύνολο του μοντέλου διορθωμένης διασποράς υποθέτει ότι αυτές κατανέμονται κανονικά και ότι τα αποτελέσματα των επιδράσεων κατανέμονται ομοιόμορφα (Therneau και Grambsch 2000, Kelly και Lim 2000). Αντίστοιχα, έχει βρεθεί ότι η ετερογένεια προκαλεί αρνητικές αποκλίσεις στην εκτίμηση των αποτελεσμάτων της θεραπείας έτσι ώστε τα αποτελέσματα της θεραπείας να υποτιμώνται.

Σε αντίθεση με τα μοντέλα διορθωμένης διασποράς, **τα μοντέλα ευπάθειας** ενσωματώνουν την ετερογένεια στην εκτιμήτρια κάνοντας υποθέσεις για την κατανομή ευπάθειας και ενσωμάτωσης του στις εκτιμήσεις των μοντέλων, και ως εκ τούτου παρουσιάζει μια πιο πολλά υποσχόμενη εναλλακτική λύση για την αντιμετώπιση της ανομοιογένειας σε σχέση με αυτή του μοντέλου διορθωμένης διασποράς (Therneau και Grambsch 2000, Oakes 1992). Η βασική λογική των μοντέλων ευπάθειας είναι ότι ορισμένα άτομα (ή ομάδες) είναι ουσιαστικά λιγότερο ή περισσότερο επιρρεπείς στο να βιώνουν τα γεγονότα που μας ενδιαφέρουν σε σχέση με άλλα άτομα, και ότι η κατανομή αυτών των επιδράσεων μπορεί τουλάχιστον να προσεγγιστεί. Τα μοντέλα ευπάθειας αντιμετωπίζουν τα επαναλαμβανόμενα γεγονότα ως μια ειδική περίπτωση μιας πιο γενικής ανομοιογένειας. Το μοντέλο ευπάθειας αναλογικού διακινδύνευσης γράφεται:

$$\lambda_i(t) = \lambda_0(t)e^{X_i\beta + \omega_i} \quad (2.6.1)$$

όπου X_i είναι η i -στη γραμμή του πίνακα συμμεταβλητών X . Τα X και β αντιστοιχούν στις σταθερές επιδράσεις του μοντέλου, το ω είναι ένας δείκτης που περιέχει τυχαίες επιδράσεις ή ευπάθειες (Hougaard, 2000). Οι χρόνοι εμφάνισης των γεγονότων υποτίθεται ότι είναι ανεξάρτητοι δεδομένης μιας επιλεγμένης παραμετρικής κατανομής, έτσι ώστε η συμπερασματολογία να γίνεται με φυσικό τρόπο. Επειδή οι κίνδυνοι είναι κατ'ανάγκη θετικοί, η κατανομή συνήθως επιλέγεται από την κλάση των θετικών κατανομών. Οι πιο ευρέως χρησιμοποιούμενες κατανομές είναι η Γάμμα και η Αντίστροφη Γκαουσιανή και η t κατανομή με την Γάμμα να είναι η συνηθέστερη (Kosorok 2003, Wei και Glidden, 1997).

Τα μοντέλα ευπάθειας είναι καλύτερα από τα μοντέλα διορθωμένης διασποράς για την αντιμετώπιση της ανομοιογένειας αλλά αγνοούν τις αμερόληπτες επιδράσεις που προκαλούνται από την εξάρτηση των γεγονότων. Ο Lawless (Lawless 1995) έριξε λίγο φως σχετικά με τη σύνδεση μεταξύ της στρωματοποιημένης ανάλυσης, η οποία ασχολείται με την εξάρτηση μεταξύ των γεγονότων, και με των μοντέλων ευπάθειας. Συγκεκριμένα η περιθώρια συνάρτηση διακινδύνευσης δίνεται από:

$$\lim_{\Delta t \rightarrow 0} \frac{P(N(t + \Delta t^-) - N(t^-) | H(t))}{\Delta t} = \lambda(t) \frac{1 + \varphi N(t^-)}{1 + \varphi \Lambda(t)} \quad (2.6.2)$$

όπου $\Lambda(t) = \int_0^t \lambda(u) \, du$.

Ούτε τα μοντέλα ευπάθειας αλλά ούτε και τα μοντέλα διορθωμένης διασποράς παρουσιάζουν μια γενική στρατηγική για τη μοντελοποίηση των διαδικασιών των επαναλαμβανόμενων γεγονότων, που χαρακτηρίζονται τόσο από την εξάρτηση μεταξύ των γεγονότων όσο και από την ανομοιογένεια μεταξύ των μονάδων. Ούτε παρουσιάζει μια αξιόπιστη λύση, όταν οι αναλυτές είναι αβέβαιοι για το ποια χαρακτηριστικά των δεδομένων αποτελούν τη βάση της συσχέτισης. Και τα δύο αναγνωρίζουν τα προβλήματα που συνδέονται με τις παραβιάσεις της ανεξαρτησίας του μοντέλου του Cox. Αλλά, όπως προτείνεται επί του παρόντος, τα μοντέλα ευπάθειας αλλά και τα μοντέλα διορθωμένης διασποράς είναι ασυμβίβαστα με την τυπική διαδικασία παραγωγής δεδομένων για επαναλαμβανόμενα γεγονότα, τα οποία περιέχουν τόσο την εξάρτηση μεταξύ των γεγονότων όσο και την ανομοιογένεια. Επιζητούμε μια λύση μοντελοποίησης που είναι ανεξάρτητη από τα άγνωστα

χαρακτηριστικά της διαδικασίας παραγωγής δεδομένων, δηλαδή που αποδίδει σωστά είτε η εξάρτηση μεταξύ των γεγονότων είτε η ανομοιογένεια περιέχει χαρακτηριστικά της διαδικασίας που εξετάζουμε. Στην επόμενη παράγραφο θα μελετήσουμε τα υπό συνθήκη μοντέλα ευπάθειας ως έναν τρόπο για την επίτευξη αυτού του στόχου.

2.7 Το υπό συνθήκη μοντέλο ευπάθειας

Το υπό συνθήκη μοντέλο ευπάθειας συνδυάζει μια τυχαία επίδραση που ενσωματώνει την παρατηρήσιμη ανομοιογένεια με τη στρωματοποίηση που βασίζεται σε κάποιο γεγονός και που ενσωματώνει την εξάρτηση του γεγονότος. Το μοντέλο διαμορφώνεται στους ενδιάμεσους χρόνους έτσι ώστε οι εκτιμήσεις των παραμέτρων να μπορούν να ερμηνευθούν ως εκτιμημένος κίνδυνος εμφάνισης του k -στου γεγονότος σε σχέση με το προηγούμενο γεγονός.

Ο συνάρτηση διακινδύνευσης ενός συγκεκριμένου γεγονότος που συμβαίνει για ένα αντικείμενο i , (λ_{ik}) , για το υπό συνθήκη μοντέλο ευπάθειας (Therneau και Grambsch, 2000, Kelly και Lim, 2000) δίνεται από:

$$\lambda_{ik}(t) = \lambda_{0k}(t - t_{k-1})e^{X_{ik}\beta + \omega_i} \quad (2.7.1)$$

όπου το k δηλώνει τον αριθμό εμφάνισης του γεγονότος, το λ_{0k} είναι μια βασική συνάρτηση διακινδύνευσης και διαφέρει ανάλογα με τον αριθμό του γεγονότος, το $(t - t_{k-1})$ ενσωματώνει τον ενδιάμεσο χρόνο μιας δομής δεδομένων έτσι ώστε η συνάρτηση διακινδύνευσης να δίνει τον κίνδυνο εμφάνισης ενός k γεγονότος δεδομένου ότι το αντικείμενο έχει βιώσει το προηγούμενο γεγονός, το X είναι ένα δείκτης ανεξάρτητων μεταβλητών και το β δίνει τις επιδράσεις τους. Το υπόλοιπο τμήμα της συνάρτησης διακινδύνευσης ενσωματώνει την τυχαία επίδραση (Hougaard 2000). Εδώ κάθε άτομο i έχει μια τυχαία επίδραση που εξετάζεται από κοινού κατά τη διάρκεια των χρόνων μεταξύ των γεγονότων και το ω είναι ένα διάνυσμα που περιέχει άγνωστες τυχαίες επιδράσεις η ευπάθειες.

Η μερική πιθανοφάνεια (Kelly και Lim, 2000) αυτού του μοντέλου, εξαρτάται από τις ευπάθειες και δίνεται από:

$$L(\beta) = \prod_{i=1}^n \prod_{k=1}^K \left(\frac{e^{X_{ik}\beta + \omega_i}}{\sum_{i=1}^n \sum_{k=1}^K Y_{ik} e^{X_{ik}\beta + \omega_i}} \right)^{\delta_{ik}} \quad (2.7.2)$$

όπου το k αντιστοιχεί στον αριθμό του γεγονότος, το δ είναι ο συνηθισμένος δείκτης αποκοπής και είναι ίσος με 1 αν ο έχει παρατηρηθεί το γεγονός και ίσος με 0 αν όχι και το Y είναι ένας δείκτης διακινδύνευσης, και είναι ίσος με 1 αν το αντικείμενο βρίσκεται σε κίνδυνο εμφάνισης του γεγονότος και 0 σε διαφορετική περίπτωση.

Το υπό συνθήκη μοντέλο ευπάθειας μας δίνει τη δυνατότητα τόσο η ανομοιογένεια όσο και η εξάρτηση μεταξύ των γεγονότων να συμβάλλουν σημαντικά στη συνάρτηση διακινδύνευσης ή στο κίνδυνο ενός ατόμου για την (επαν)εμφάνιση ενός συγκεκριμένου γεγονότος. Ο ενδιαμέσος χρόνος ενός υπό συνθήκη μοντέλο ευπάθειας δεν είναι σαφής σε σχέση με τα συνηθισμένα μοντέλα επαναλαμβανόμενων γεγονότων και φαίνεται ιδιαίτερα ελπιδοφόρα για την αντιμετώπιση τόσο της ετερογένειας όσο και της εξάρτησης μεταξύ των γεγονότων, επειδή επιτρέπει και για τα δύο μια βασική συχνότητα διακινδύνευσης ενός συγκεκριμένου γεγονότος αλλά και μια τυχαία επίδραση.

Κεφάλαιο 3: Στατιστική Ανάλυση

3.1 Καρκίνος της ουροδόχου κύστης

Ένα σύνολο δεδομένων του καρκίνου της ουροδόχου κύστης (Andrews και Hertzberg 1985, Wei, Lin και Weissfeld 1989) χρησιμοποιείται ως βασικό παράδειγμα για την περιγραφή κυρίως της μεθόδου WLW. Στα δεδομένα καταγράφηκαν οι χρόνοι (σε μήνες) εμφάνισης έως και τεσσάρων επαναλαμβανόμενων γεγονότων για 86 άτομα. Μπορεί να υπάρχει συνέχεια της παρακολούθησης αυτών των ατόμων ακόμα και μετά την τελευταία επανάληψη.

	Αριθμός Επαναλήψεων				
	0	1	2	3	4
Αριθμός Ατόμων που Μελετώνται	39	18	7	8	14
Παρακολούθηση Μετά το Τελευταίο Γεγονός	38	17	5	6	12

Ένα από τα αντικείμενα της μελέτης δεν έχει κανένα επαναλαμβανόμενο γεγονός και 0 μήνες παρακολούθησης. Για λόγους απλότητας, αυτό το αντικείμενο διαγράφεται από το σύνολο δεδομένων, αφού δεν έχει να προσθέσει τίποτα επί της ουσίας στη συνάρτηση πιθανοφάνειας. Οι συμμεταβλητές που χρησιμοποιούμε είναι το αρχικό μέγεθος του μεγαλύτερου όγκου, ο αρχικός αριθμός όγκων και το είδος της θεραπείας που χρησιμοποιείται σε κάθε άτομο.

Το σύνολο δεδομένων bladder2 που θα παρουσιαστεί παρακάτω διαμορφώθηκε για την εφαρμογή των μοντέλων των Andersen-Gill αλλά και των Prentice-Williams-Peterson. Παρακάτω παρουσιάζεται ένα μικρό μέρος του συνόλου δεδομένων.

Group	futime	number	size	recurrence	times
1	0	1	1		
1	1	1	3		
1	4	2	1		
1	7	1	1		
1	10	5	1		
1	10	4	1	6	
1	14	1	1		
1	18	1	1		
1	18	1	3	5	
1	18	1	1	12	16

Πίνακας 3.1: Ένα δείγμα από τα δεδομένα των ασθενών που πάσχουν από καρκίνο της ουροδόχου κύστης

id	rx	number	size	start	stop	event	enum
1	1	1	3	0	1	0	1
2	1	2	1	0	4	0	1
3	1	1	1	0	7	0	1
4	1	5	1	0	10	0	1
5	1	4	1	0	6	1	1
5	1	4	1	6	10	0	2
6	1	1	1	0	14	0	1
7	1	1	1	0	18	0	1
8	1	1	3	0	5	1	1
8	1	1	3	5	18	0	2
9	1	1	1	0	12	1	1
9	1	1	1	12	16	1	2
9	1	1	1	16	18	0	3

Πίνακας 3.2: Τα δεδομένα του Πίνακα 3.1 για την ανάλυση των μεθόδων Andersen-Gill και Prentice-Williams-Peterson.

id: Αύξων αριθμός του κάθε ασθενή

rx: Το είδος της θεραπείας του κάθε ασθενή (1 για φάρμακο τύπου πλασίμπο και 2 για οποιαδήποτε άλλη θεραπεία)

number: Αρχικός αριθμός όγκων

size: Μέγεθος σε cm μεγαλύτερου αρχικού όγκου

recur: Αριθμός των γεγονότων

start,stop: Η αρχή και το τέλος κάθε διαστήματος μεταξύ επαναλαμβανόμενων γεγονότων

status: 0 σε περίπτωση αποκοπής και 1 σε περίπτωση κάποιου επαναλαμβανόμενου γεγονότος

enum: Αριθμός γεγονότος

Η μορφή του συνόλου δεδομένων για το μοντέλο του WLW περιέχει τέσσερις σειρές για κάθε αντικείμενο αφού ο μέγιστος αριθμός επαναλαμβανόμενων γεγονότων που έχει καταγραφεί για τουλάχιστον ένα αντικείμενο της μελέτης είναι το τέσσερα. Το σύνολο δεδομένων για την ανάλυση υπό τη μορφή της μεθόδου WLW περιέχει $85 \cdot 4 = 340$ παρατηρήσεις. Ένα μικρό του μέρος παρουσιάζεται παρακάτω:

id	rx	number	size	stop	event	enum
1	1	1	3	1	0	1
1	1	1	3	1	0	2
1	1	1	3	1	0	3
1	1	1	3	1	0	4
2	1	2	1	4	0	1
2	1	2	1	4	0	2
2	1	2	1	4	0	3
2	1	2	1	4	0	4
3	1	1	1	7	0	1
3	1	1	1	7	0	2
3	1	1	1	7	0	3
3	1	1	1	7	0	4
4	1	1	1	10	0	1
4	1	1	1	10	0	2
4	1	1	1	10	0	3
4	1	1	1	10	0	4

Πίνακας 3.3: Τα δεδομένα του Πίνακα 3.1 για την ανάλυση της μεθόδου Wei, Lin και Weissfeld.

Μετά την κατασκευή των πινάκων το επόμενο βήμα που έχουμε να κάνουμε είναι να προσαρμόσουμε τα τρία μοντέλα τα οποία αναφέραμε

προηγουμένως. Αυτό μπορεί να γίνει με τη βοήθεια της R και των ακόλουθων εντολών:

Οι εντολή που χρησιμοποιούμε για την ανάλυση AG είναι η ακόλουθη.

```
afit <- coxph(Surv(start, stop, event) ~ (rx + size + number + cluster(id)), data=bladder2) #AG
```

Τα αποτελέσματα που προκύπτουν από την προηγούμενη εντολή είναι τα ακόλουθα.

afit

Call:

```
coxph(formula = Surv(start, stop, event) ~ (rx + size + number + cluster(id)), data = bladder2)
```

	coef	exp(coef)	se(coef)	robust se	z	p
rx	-0.4647	0.628	0.1997	0.2656	-1.750	0.0800
size	-0.0437	0.957	0.0691	0.0776	-0.563	0.5700
number	0.1750	1.191	0.0471	0.0630	2.775	0.0055

Likelihood ratio test=17.5 on 3 df, p=0.000553 n= 178, number of events= 112

Για την ανάλυση του WLW χρειάζεται η ακόλουθη εντολή.

```
wfit <- coxph(Surv(stop, event)~ (rx + size + number + cluster(id) + strata(enum)), data=bladder) #WLW
```

Μετά την προσαρμογή προκύπτουν τα ακόλουθα αποτελέσματα.

wfit

Call:

```
coxph(formula = Surv(stop, event) ~ (rx + size + number + cluster(id) + strata(enum)), data = bladder)
```

	coef	exp(coef)	se(coef)	robust se	z	p
rx	-0.5848	0.557	0.2011	0.3079	-1.899	0.0580
size	-0.0516	0.950	0.0697	0.0946	-0.546	0.5900
number	0.2103	1.234	0.0468	0.0666	3.156	0.0016

Likelihood ratio test=25.3 on 3 df, p=1.36e-05 n= 340, number of events= 112

Η ακόλουθη εντολή είναι για την προσαρμογή του PWP μοντέλου.

```
fitc<- coxph(Surv(start, stop, event)~ (rx + size + number +  
cluster(id) + strata(enum)), data=bladder2) #PWP
```

Τα αποτελέσματα που προκύπτουν είναι τα εξής.

fitc

Call:

```
coxph(formula = Surv(start, stop, event) ~ (rx + size + number +  
cluster(id) + strata(enum)), data = bladder2)
```

	coef	exp(coef)	se(coef)	robust se	z	p
rx	-0.33349	0.716	0.2162	0.2048	-1.628	0.10
size	-0.00849	0.992	0.0728	0.0616	-0.138	0.89
number	0.11962	1.127	0.0533	0.0514	2.328	0.02

```
Likelihood ratio test=6.51 on 3 df, p=0.0893 n= 178, number of  
events= 112
```

Αν παρατηρήσουμε τις p-τιμές της συμμεταβλητής rx, τότε θα παρατηρήσουμε ότι η rx της μεθόδου WLW ($p=0.0580$) έχει πιο ισχυρή επίδραση σε σχέση με αυτή των AG ($p=0.800$) και PWP ($p=0.10$). Επίσης, οι διαφορές ανάμεσα στο μοντέλα δεν είναι τόσο ισχυρές όσο θα περιμέναμε να είναι. Παρατηρώντας επίσης τις p-τιμές μετά από κάθε έλεγχο βλέπουμε ότι μόνο οι p-τιμές των μεθόδων AG ($p=0.000553$) και WLW ($p=1.36e-05$) είναι μικρότερες του 5% πράγμα που σημαίνει ότι απορρίπτουμε την μηδενική υπόθεση δηλαδή οι $rx=size=number=0$. Αυτό σημαίνει ότι και οι τρεις συντελεστές της μεθόδου PWP είναι διάφοροι του μηδενός.

Αν προσαρμόσουμε ξεχωριστά τους συντελεστές που αφορούν το είδος της θεραπείας σε κάθε στρώμα (enum) τότε από τα αποτελέσματα που παίρνουμε παρακάτω παρατηρούμε ότι υπάρχει μια τάση οι συντελεστές της μεθόδου WLW να αυξάνονται μέχρι το τελευταίο στρώμα και οι συντελεστές της μεθόδου PWP να μειώνονται με ιδιαίτερη έμφαση στο τελευταίο στρώμα όπου έχουμε και αλλαγή πρόσημου στο συντελεστή rx4.

	rx
	coef
PWP	
1η επίσκεψη	-0,41835
2η επίσκεψη	-0,45184
3η επίσκεψη	-0,06741
4η επίσκεψη	0,20436
WLW	
1η επίσκεψη	-0,4857
2η επίσκεψη	-0,6621
3η επίσκεψη	-0,7163
4η επίσκεψη	-0,5537

Πίνακας 3.4: Αποτελέσματα της ανάλυσης του συντελεστή θεραπείας για κάθε επίσκεψη ξεχωριστά με βάση τα μοντέλα των PWP και WLW.

Οι αντίστοιχες εντολές που χρησιμοποιήθηκαν για το μοντέλο WLW είναι :

```
rx1<- bladder$rx * (bladder$enum==1)
rx2<- bladder$rx * (bladder$enum==2)
rx3<- bladder$rx * (bladder$enum==3)
rx4<- bladder$rx * (bladder$enum==4)

wfit14 <- coxph(Surv(stop, event)~ (rx1 + rx2 + rx3 + rx4 + size +
number + strata(enum) + cluster(id)), data=bladder)
```

Αποτελέσματα από την εφαρμογή δίνονται ως:

wfit14

Call:

```
coxph(formula = Surv(stop, event) ~ (rx1 + rx2 + rx3 + rx4 +
size + number + strata(enum) + cluster(id)), data = bladder)
```

	coef	exp(coef)	se(coef)	robust se	z	p
rx1	-0.4857	0.615	0.3057	0.2900	-1.675	0.0940
rx2	-0.6621	0.516	0.3922	0.3728	-1.776	0.0760
rx3	-0.7163	0.489	0.4597	0.4264	-1.680	0.0930

```

rx4      -0.5537      0.575      0.5612      0.5026 -1.102 0.2700
size     -0.0526      0.949      0.0697      0.0945 -0.556 0.5800
number   0.2092      1.233      0.0467      0.0667  3.136 0.0017

```

Likelihood ratio test=25.5 on 6 df, p=0.000277 n= 340, number of events= 112

Οι εντολές που αφορούν το μοντέλο του PWP είναι οι παρακάτω:

```

rx1<- bladder2$rx *(bladder2$enum==1)
rx2<- bladder2$rx *(bladder2$enum==2)
rx3<- bladder2$rx *(bladder2$enum==3)
rx4<- bladder2$rx *(bladder2$enum==4)

fitc14 <- coxph(Surv(start, stop, event)~ (rx1 + rx2 + rx3 + rx4 +
size + number + strata(enum)), data=bladder2)

```

Τα αποτελέσματα που προκύπτουν με τη βοήθεια της R είναι τα ακόλουθα:

```

fitc14
Call:
coxph(formula = Surv(start, stop, event) ~ (rx1 + rx2 + rx3 +
      rx4 + size + number + strata(enum)), data = bladder2)

```

	coef	exp(coef)	se(coef)	z	p
rx1	-0.41835	0.658	0.3042	-1.375	0.170
rx2	-0.45184	0.636	0.3981	-1.135	0.260
rx3	-0.06741	0.935	0.5985	-0.113	0.910
rx4	0.20436	1.227	0.6881	0.297	0.770
size	-0.00745	0.993	0.0732	-0.102	0.920
number	0.11421	1.121	0.0540	2.115	0.034

Likelihood ratio test=7.47 on 6 df, p=0.28 n= 178, number of events= 112

Παρακάτω θα βάλουμε στην ερέυνα μας δύο ακόμα προσαρμογές. Αυτές επί της ουσίας είναι προσαρμογές που αφορούν το μοντέλο PWP. Η

πρώτη προσαρμόζει τα δεδομένα μας μέχρι την εμφάνιση του πρώτου γεγονότος ενώ η δεύτερη περιλαμβάνει μόνο τους ασθενείς που έχουν βιώσει τουλάχιστον ένα γεγονός. Οι εντολές και τα αποτελέσματα παρουσιάζονται παρακάτω.

```
fit2pa <- coxph(Surv(stop, event)~ rx + size + number, bladder2,
+ subset=(enum==1))
fit2pa
Call:
coxph(formula = Surv(stop, event) ~ rx + size + number, data =
bladder2, subset = (enum == 1))
```

	coef	exp(coef)	se(coef)	z	p
rx	-0.5260	0.591	0.3158	-1.665	0.0960
size	0.0696	1.072	0.1016	0.685	0.4900
number	0.2382	1.269	0.0759	3.139	0.0017

Likelihood ratio test=9.92 on 3 df, p=0.0193 n= 85, number of events= 47

```
fit2pb <- coxph(Surv(stop-start, event) ~ rx + size + number,
bladder2 + subset=(enum==2))
```

```
fit2pb
```

```
Call:
```

```
coxph(formula = Surv(stop - start, event) ~ rx + size + number, data
= bladder2, subset = (enum == 2))
```

	coef	exp(coef)	se(coef)	z	p
rx	-0.27097	0.763	0.4050	-0.6691	0.50
size	-0.11927	0.888	0.1191	-1.0014	0.32
number	-0.00641	0.994	0.0964	-0.0666	0.95

Likelihood ratio test=1.35 on 3 df, p=0.716 n= 46, number of events= 29

Είναι φανερό από τις p-τιμές ($p_{fit2pa}(rx) = 0.0960$ και $p_{fit2pb}(rx) = 0.50$) και των δύο προσαρμογών για το rx, ότι οι δύο θεραπείες δεν διαφέρουν μεταξύ τους. Επίσης από τις p-τιμές κάθε προσαρμογής βλέπουμε ότι στην πρώτη μέθοδο ($p = 0.0193$) και οι τρίτες συντελεστές είναι διάφοροι του μηδενός κάτι που δε συμβαίνει για τη δεύτερη μέθοδο ($p = 0.716$).

Μια τελευταία προσαρμογή είναι η προσαρμογή ενδιάμεσων χρόνων.

```
gapfit <- coxph(Surv(stop-start, event) ~ (rx + size + number +
strata(enum)), data=bladder2, subset=(enum<5)) #gap time fit
```

gapfit

Call:

```
coxph(formula = Surv(stop - start, event) ~ (rx + size + number +
      strata(enum)), data = bladder2, subset = (enum < 5))
```

	coef	exp(coef)	se(coef)	z	p
rx	-0.27900	0.757	0.2073	-1.346	0.1800
size	0.00742	1.007	0.0700	0.106	0.9200
number	0.15805	1.171	0.0519	3.043	0.0023

```
Likelihood ratio test=9.33 on 3 df, p=0.0252 n= 178, number of
events= 112
```

Και αυτή η μέθοδος μας δείχνει ότι δεν υπάρχουν διαφορές μεταξύ των δύο τρόπων θεραπείας ($p=0.1800$) αλλά οι συντελεστές rx, size και number είναι και οι τρεις διάφοροι του μηδενός, αφού η p-τιμή της προσαρμογής ενδιάμεσων χρόνων είναι $p=0.0252$.

Κάτι που έχει επίσης ιδιαίτερο ενδιαφέρον είναι η προσαρμογή των μοντέλων AG και PWP με τη βοήθεια των μοντέλων ευπάθειας. Συγκεκριμένα θα επιλέξουμε το ρόλο του βασικού μοντέλου να τον παίξει η κατανομή Γάμμα. Με τη βοήθεια της R λοιπόν θα πάρουμε τα παρακάτω αποτελέσματα.

Οι εντολές λοιπόν για το μοντέλο AG είναι:

```
fracfit <- coxph(Surv(start, stop, event) ~ (rx + size + number +
frailty(id, dist='gamma')), data=bladder2) #fitting AG model with gamma
frailty
```

Και τα αποτελέσματα που μας δίνει η R είναι:

```
frafit
```

```
Call:
```

```
coxph(formula = Surv(start, stop, event) ~ (rx + size + number +  
      frailty(id, dist = "gamma")), data = bladder2)
```

	coef	se(coef)	se2	Chisq	DF	p
rx	-0.6077	0.3330	0.2197	3.33	1.0	0.06800
size	-0.0215	0.1140	0.0717	0.04	1.0	0.85000
number	0.2387	0.0932	0.0570	6.56	1.0	0.01000
frailty(id, dist = "gamma				82.92	45.4	0.00056

```
Iterations: 6 outer, 29 Newton-Raphson
```

```
Variance of random effect= 1.08 I-likelihood = -436.8
```

```
Degrees of freedom for terms= 0.4 0.4 0.4 45.4
```

```
Likelihood ratio test=144 on 46.6 df, p=7.87e-12 n= 178
```

Η ακόλουθη εντολή είναι για την προσαρμογή του PWP μοντέλου ευπάθειας.

```
frfitc<- coxph(Surv(start, stop, event)~ (rx + size + number +  
frailty(id, dist='gamma') + strata(enum)), data=bladder2) #fitting  
conditional model with gamma frailty
```

Τα αποτελέσματα της παραπάνω εφαρμογής δίνονται ως.

```
frfitc
```

```
Call:
```

```
coxph(formula = Surv(start, stop, event) ~ (rx + size + number +  
      frailty(id, dist = "gamma") + strata(enum)), data = bladder2)
```

	coef	se(coef)	se2	Chisq	DF	p
rx	-0.33349	0.2162	0.2162	2.38	1	0.120
size	-0.00849	0.0728	0.0728	0.01	1	0.910
number	0.11962	0.0533	0.0533	5.03	1	0.025
frailty(id, dist = "gamma				0.00	0	0.920

```
Iterations: 6 outer, 29 Newton-Raphson
```

```
Variance of random effect= 5e-07 I-likelihood = -316
```

```
Degrees of freedom for terms= 1 1 1 0
```

```
Likelihood ratio test=6.51 on 3 df, p=0.0893 n= 178
```

Παρατηρώντας τα παραπάνω αποτελέσματα βλέπουμε ότι δεν έχουμε αξιολογές διαφορές σε σχέση με τα αποτελέσματα που είχαμε για αυτές τις δύο μεθόδους πριν τη χρήση μοντέλων ευπάθειας. Και στα δύο μοντέλα οι τρόποι θεραπείας δε διαφέρουν, με $p=0.068$ για το συντελεστή rx της μεθόδου AG να είναι ελαφρώς ισχυρότερος του αντίστοιχου της μεθόδου PWP και $p=0.120$ με τη μέθοδο PWP.

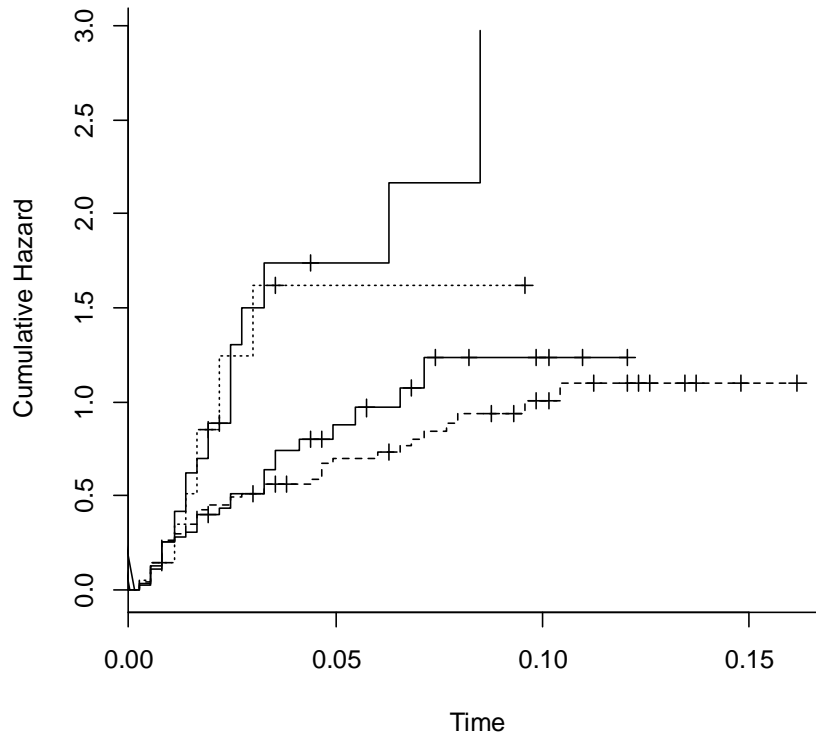
Είναι σημαντικό να δούμε ένα διάγραμμα ενδιάμεσων χρόνων όπου θα παρατηρήσουμε την εξέλιξη των γεγονότων με την πάροδο του χρόνου. Το διάγραμμα που χρειαζόμαστε είναι το παρακάτω:

```
gapfit <- coxph(Surv(stop-start, event) ~ (rx + size + number +
strata(enum)), data=bladder2, subset=(enum<5))

curve <- survfit(gapfit)

plot(curve, lty=c(2,1,1,3,3,3,3), fun='cumhaz', xscale=365, ylab =
"Cumulative Hazard", xlab = "Time", bty="L", main= "Gap Fit Cumulative
Hazard Function" )
```

Gap Fit Cumulative Hazard Function



Σχήμα 3.1: Διάγραμμα ενδιάμεσων χρόνων (-- Γεγονός 1, —Γεγονότα 2 και 3, ... Γεγονός 4)

Σύμφωνα με το Σχήμα 3.1 η διάρκεια του πρώτου γεγονότος είναι πολύ μεγαλύτερη από τη διάρκεια των άλλων ενδιάμεσων χρόνων. Αυτός συμβαίνει διότι για την δεύτερη εμφάνιση κάποιου γεγονότος, επιλέγονται συνήθως άτομα που έχουν υψηλό ποσοστό εμφάνισης κάποιου γεγονότος. Επίσης και ο παράγοντας number είναι πιθανό να παίζει ένα σημαντικό ρόλο στην διάρκεια των γεγονότων.

3.2 *rhDNase* σε ασθενείς με κυστική ίνωση

Στους ασθενείς με κυστική ίνωση, εξωκυτταρικό DNA απελευθερώνεται από τα λευκά αιμοσφαίρια που συσσωρεύονται στους αεραγωγούς για την αντιμετώπιση χρόνιων βακτηριακών μολύνσεων. Αυτό το DNA που περισσεύει αυξάνει τον όγκο της βλέννας το οποίο δεν μπορεί στην συνέχεια να καθαριστεί από τους πνεύμονες. Αυτή η συσσώρευση οδηγεί σε εξάρσεις των αναπνευστικών συμπτωμάτων και σταδιακή επιδείνωση της αναπνευστικής λειτουργίας. Περισσότερο από το 90% των ασθενών με κυστική ίνωση πεθαίνουν από πνευμονοπάθεια.

Το δεοξυριβονουκλεοτίδιο 1 (DNase 1) είναι ένα ανθρώπινο ένζυμο που φυσιολογικά εμφανίζεται στη βλέννα των ανθρώπινων πνευμόνων και αφομοιώνει το εξωκυτταρικό DNA. Χορηγήθηκε σε ασθενείς σε μια κλινική δοκιμή.

Το σύνολο δεδομένων rhDNase1 (Cook και Lawless 2007) διαμορφώθηκε για την προσαρμογή των Andersen-Gill αλλά και των Prentice-Williams-Peterson. Παρακάτω παρουσιάζεται ένα μικρό μέρος του συνόλου δεδομένων.

id	trt	fev1	fev2	time1	time2	status	enum	etype
493301	1	28,8	28,1	0	168	0	1	1
493303	1	64	63	0	169	0	1	1
493305	0	67,2	68,7	0	65	1	1	1
493305	0	67,2	68,7	0	65	0	2	2
493305	0	67,2	68,7	0	65	0	3	1
493309	1	57,6	56,5	0	168	0	1	1
493310	0	57,6	56,3	0	171	0	1	1
493311	1	25,6	25,3	0	166	0	1	1
493312	0	86,4	85,4	0	168	0	1	1
493313	0	32	32,4	0	90	1	1	1
493313	0	32	32,4	0	90	0	2	2
493313	0	32	32,4	0	90	0	3	1
589301	1	86,4	86	0	169	0	1	1
589302	0	28,8	29,2	0	8	1	1	1
589302	0	28,8	29,2	0	8	0	2	2
589302	0	28,8	29,2	0	8	0	3	1
589302	0	28,8	29,2	0	8	0	4	2
589302	0	28,8	29,2	0	8	0	5	1

Πίνακας 3.5: Τα δεδομένα rhDNase 1 για την ανάλυση των μεθόδων Andersen-Gill και Prentice-Williams-Peterson.

id: Αύξων αριθμός ασθενούς

trt: το είδος της θεραπείας, trt=1 αν ο ασθενής λαμβάνει το φάρμακο rhDNase και 0 αν ο ασθενής λαμβάνει το πλασίμπο

fev1: μέτρηση του βιαίως εκπνεόμενου όγκου αέρα

fev2: δεύτερη μέτρηση του βιαίως εκπνεόμενου όγκου αέρα

time1: η αρχή ενός διαστήματος

time2: το τέλος του διαστήματος. Η μεταβλητή etype δείχνει το είδος του διαστήματος

status: όταν etype=1 το status είναι ίσο με 1 αν το time2 είναι ένας ο χρόνος ενός γεγονότος και ίσο με 0 όταν το time2 είναι χρόνος αποκοπής

etype: ο δείκτης του είδους του διαστήματος (time1, time2). Αν etype=1, το αντικείμενο βρίσκεται σε κίνδυνο

enum: ο συνολικός αριθμός των γραμμών στα δεδομένα κάθε ατόμου που εξετάζεται

Το να κατασκευαστεί ο πίνακας για WLW σε αυτό το πείραμα ήταν πιο δύσκολος απ' ό τι για το σύνολο των δεδομένων που περιγράφηκε στην προηγούμενη παράγραφο. Η δυσκολία με τα δεδομένα rhDNase είναι ότι μετά από κάθε γεγονός, ο ασθενής λαμβάνει αντιβιοτικά για ένα χρονικό διάστημα, εντός του οποίου δε βρίσκεται σε κίνδυνο. Έστω ο ασθενής 493305. Ο ασθενής 493305 έχει μια έξαρση τη μέρα 65 με διάρκεια 4 μέρες (ο κάθε ασθενής δεν είναι σε κίνδυνο για 6 μέρες μετά τη λήψη του αντιβιοτικού) και η παρακολούθηση τερματίζεται την ημέρα 165. Έτσι για τον ασθενή 493305 τα δεδομένα είναι:

id	trt	fev1	fev2	time1	time2	status	enum
493305	0	67,2	68,7	0	65	1	1
493305	0	67,2	68,7	65	75	1	2
493305	0	67,2	68,7	75	168	0	3

Πίνακας 3.6: Ένα δείγμα από τον Πίνακα 3.5.

Για την ανάλυση του WLW θα έχουμε 9 παρατηρήσεις πλέον και τα παραπάνω δεδομένα του ασθενούς 493305 θα πάρουν τη παρακάτω μορφή:

id	trt	fev1	fev2	time1	time2	status	enum
493305	0	67.2	68.7	0	65	1	1
493305	0	67.2	68.7	0	65	0	2
493305	0	67.2	68.7	0	65	0	3
493305	0	67.2	68.7	0	65	0	4
493305	0	67.2	68.7	0	65	0	5
493305	0	67.2	68.7	75	168	0	2
493305	0	67.2	68.7	75	168	0	3
493305	0	67.2	68.7	75	168	0	4
493305	0	67.2	68.7	75	168	0	5

Πίνακας 3.7: Ανάλυση του Πίνακα 3.6 για να μπορέσει να γίνει η ανάλυση της μεθόδου των Wei, Lin και Weissfeld.

Μετά την κατασκευή των πινάκων το επόμενο βήμα είναι να προσαρμόσουμε τα τρία μοντέλα τα οποία αναφέραμε προηγουμένως.

Οι εντολή που χρησιμοποιούμε για την ανάλυση AG είναι η ακόλουθη.

```
afit <- coxph(Surv(time1, time2, status) ~ (trt + fev1 + fev2 +  
cluster(id)), data=rhDNase1 )
```

Τα αποτελέσματα που προκύπτουν από την προηγούμενη εντολή είναι τα ακόλουθα.

```
afit
```

```
Call:
```

```
coxph(formula = Surv(time1, time2, status) ~ (trt + fev1 + fev2 +  
cluster(id)), data = rhDNase1)
```

	coef	exp(coef)	se(coef)	robust se	z	p
trt	-0.2640	0.768	0.0752	0.1210	-2.182	0.029
fev1	0.0604	1.062	0.0401	0.0672	0.899	0.370
fev2	-0.0770	0.926	0.0401	0.0672	-1.147	0.250

```
Likelihood ratio test=131 on 3 df, p=0 n= 1326, number of events=  
721
```

Η ακόλουθη εντολή είναι για την προσαρμογή του PWP μοντέλου.

```
fitc <- coxph(Surv(time1, time2, status) ~ (trt + fev1 + fev2 +  
cluster(id) + strata(enum)), data=rhDNase1)
```

Τα αποτελέσματα που προκύπτουν είναι τα εξής.

```
fitc
```

```
Call:
```

```
coxph(formula = Surv(time1, time2, status) ~ (trt + fev1 + fev2 +  
cluster(id) + strata(enum)), data = rhDNase1)
```

	coef	exp(coef)	se(coef)	robust se	z	p
trt	-0.1890	0.828	0.0774	0.0866	-2.184	0.029
fev1	0.0174	1.018	0.0411	0.0414	0.419	0.680
fev2	-0.0240	0.976	0.0412	0.0415	-0.579	0.560

Likelihood ratio test=21.8 on 3 df, p=7.21e-05 n= 1326, number of events= 721

Για την ανάλυση του WLW χρειάζεται η ακόλουθη εντολή.

```
wfit <- coxph(Surv(time2, status) ~ (trt + fev1 + fev2 + cluster(id) + strata(enum)), data=rhDNase3 )
```

Μετά την προσαρμογή προκύπτουν τα ακόλουθα αποτελέσματα.

```
wfit
```

```
Call:
```

```
coxph(formula = Surv(time2, status) ~ (trt + fev1 + fev2 + cluster(id) + strata(enum)), data = rhDNase3)
```

	coef	exp(coef)	se(coef)	robust se	z	p
trt	-0.3546	0.701	0.1067	0.1450	-2.445	0.014
fev1	0.0728	1.076	0.0570	0.0813	0.896	0.370
fev2	-0.0923	0.912	0.0571	0.0813	-1.135	0.260

Likelihood ratio test=91.8 on 3 df, p=0 n= 4388, number of events= 360

Η τελευταία ανάλυση και η εντολή που θα χρησιμοποιηθεί είναι για την ανάλυση μέχρι την εκδήλωση του πρώτου γεγονότος.

```
fit1 <- coxph(Surv(time1, time2, status) ~ (trt + fev1 + fev2 + cluster(id)), data=rhDNase1, subset=(enum==1))#first event
```

Τα αποτελέσματα που προκύπτουν είναι.

```
fit1
```

```
Call:
```

```
coxph(formula = Surv(time1, time2, status) ~ (trt + fev1 + fev2 + cluster(id)), data = rhDNase1, subset = (enum == 1))
```

	coef	exp(coef)	se(coef)	robust se	z	p
--	------	-----------	----------	-----------	---	---

```

trt -0.3611      0.697      0.1293      0.1302 -2.773 0.0056

fev1 0.0421      1.043      0.0687      0.0706 0.596 0.5500

fev2 -0.0629      0.939      0.0687      0.0708 -0.889 0.3700

```

Likelihood ratio test=70.6 on 3 df, p=3.11e-15 n= 645, number of events= 245

Συγκρίνοντας τα αποτελέσματα του μοντέλου των Andersen Gill με το αποτέλεσμα που προκύπτει από το μοντέλο που εστιάζει στο πρώτο γεγονός, παρατηρούμε μια αισθητή μείωση στην επίδραση της θεραπείας μετά το πρώτο γεγονός. Το ίδιο προκύπτει και συγκρίνοντας το μοντέλο του πρώτου γεγονότος και με το PWP. Συγκρίνοντας το βέβαια με το WLW παρατηρούμε ότι οι συντελεστές είναι σχεδόν πανομοιότυποι. Κάτι ιδιαίτερα σημαντικό που προκύπτει και από τις τέσσερις προσαρμογές που πραγματοποιήθηκαν παραπάνω, είναι ότι οι p-τιμές ($p_{AG} = 0.029$, $p_{PWP} = 0.029$, $p_{WLW} = 0.014$, $p_{first} = 0.0056$) μας δείχνουν ότι τα δύο διαφορετικά είδη θεραπείας διαφέρουν σημαντικά μεταξύ τους.

	fev1	fev2	trt
	coef	coef	coef
First Event	0,0421	-0,0629	-0,3611
Andersen-Gill	0,0604	-0,077	-0,264
PWP	0,0174	-0,024	-0,189
WLW	0,0728	-0,0923	-0,3546

Πίνακας 3.7: Αποτελέσματα που προέκυψαν από την ανάλυση των τεσσάρων μεθόδων.

Παρακάτω θα προσπαθήσουμε να αναλύσουμε χρησιμοποιώντας τα μοντέλα των PWP και WLW, τα αποτελέσματα του φαρμάκου rhDNase ξεχωριστά για την πρώτη, τη δεύτερη και μετά την τρίτη επίσκεψη με σκοπό να διακρίνουμε αν όντως υπάρχουν σημαντικές επιδράσεις του.

Οι εντολές που αφορούν το μοντέλο του PWP είναι οι παρακάτω:

```

trt1<- rhDNase1$trt *(rhDNase1$enum==1)

trt2<- rhDNase1$trt *(rhDNase1$enum==2)

```

```

trt3<- rhDNase1$trt *(rhDNase1$enum>=3)

fev11<- rhDNase1$fev1 *(rhDNase1$enum==1)

fev12<- rhDNase1$fev1 *(rhDNase1$enum==2)

fev13<- rhDNase1$fev1 *(rhDNase1$enum>=3)

fev21<- rhDNase1$fev2 *(rhDNase1$enum==1)

fev22<- rhDNase1$fev2 *(rhDNase1$enum==2)

fev23<- rhDNase1$fev2 *(rhDNase1$enum>=3)

fitc13 <- coxph(Surv(time1, time2, status) ~ (trt1 + trt2 + trt3
+fev11 + fev12 + fev13 + fev21 + fev22 + fev23 + cluster(id) +
strata(enum)), data=rhDNase1 )

```

Τα αποτελέσματα που προκύπτουν είναι τα ακόλουθα:

```
fitc13
```

```
Call:
```

```

coxph(formula = Surv(time1, time2, status) ~ (trt1 + trt2 + trt3 +
fev11 + fev12 + fev13 + fev21 + fev22 + fev23 + cluster(id) +
strata(enum)), data = rhDNase1)

```

	coef	exp(coef)	se(coef)	robust se	z	p
trt1	-0.36113	0.697	0.1293	0.1302	-2.7729	0.0056
trt2	-0.19911	0.819	0.1342	0.1604	-1.2414	0.2100
trt3	-0.00416	0.996	0.1395	0.1238	-0.0336	0.9700
fev11	0.04212	1.043	0.0687	0.0706	0.5963	0.5500
fev12	-0.03469	0.966	0.0719	0.0464	-0.7471	0.4600
fev13	0.05351	1.055	0.0725	0.0645	0.8298	0.4100
fev21	-0.06294	0.939	0.0687	0.0708	-0.8885	0.3700
fev22	0.04128	1.042	0.0720	0.0466	0.8864	0.3800
fev23	-0.05188	0.949	0.0727	0.0648	-0.8000	0.4200

```

Likelihood ratio test=78 on 9 df, p=3.93e-13 n= 1326, number of
events= 721

```

Οι αντίστοιχες εντολές που χρησιμοποιήθηκαν για το μοντέλο PWP, για το μοντέλο WLW είναι :

```

trt1<- rhDNase3 $trt *(rhDNase3 $enum==1)
trt2<- rhDNase3 $trt *(rhDNase3 $enum==2)
trt3<- rhDNase3 $trt *(rhDNase3 $enum>=3)
fev11<- rhDNase3$fev1 *(rhDNase3$enum==1)
fev12<- rhDNase3$fev1 *(rhDNase3$enum==2)
fev13<- rhDNase3$fev1 *(rhDNase3$enum>=3)
fev21<- rhDNase3$fev2 *(rhDNase3$enum==1)
fev22<- rhDNase3$fev2 *(rhDNase3$enum==2)
fev23<- rhDNase3$fev2 *(rhDNase3$enum>=3)

wfit13 <- coxph(Surv(time2, status) ~ (trt1 + trt2 + trt3 + fev11 +
fev12 + fev13 + fev21 + fev22 + fev23 + cluster(id) + strata(enum)),
data=rhDNase3)

```

Αποτελέσματα από την εφαρμογή δίνονται ως:

wfit13

Call:

```

coxph(formula = Surv(time2, status) ~ (trt1 + trt2 + trt3 + fev11 +
fev12 + fev13 + fev21 + fev22 + fev23 + cluster(id) + strata(enum)),data =
rhDNase3)

```

	coef	exp(coef)	se(coef)	robust se	z	p
trt1	-0.3910	0.676	0.1300	0.1307	-2.992	0.0028
trt2	-0.0804	0.923	0.2230	0.2270	-0.354	0.7200
trt3	-0.7295	0.482	0.3537	0.4353	-1.676	0.0940
fev11	0.0419	1.043	0.0690	0.0714	0.587	0.5600
fev12	0.0572	1.059	0.1216	0.1318	0.434	0.6600
fev13	0.3152	1.371	0.1852	0.2339	1.348	0.1800
fev21	-0.0626	0.939	0.0691	0.0716	-0.875	0.3800
fev22	-0.0748	0.928	0.1217	0.1320	-0.567	0.5700
fev23	-0.3299	0.719	0.1854	0.2329	-1.417	0.1600

Likelihood ratio test=97.3 on 9 df, p=1.11e-16 n= 4388, number of events= 360

Παρατηρώντας τα παραπάνω αποτελέσματα θα δούμε ότι και στη μέθοδο WLW αλλά και στη μέθοδο PWP κατά την πρώτη επίσκεψη υπάρχει μια αισθητή επίδραση της θεραπείας όπως παρατηρούμε και από τον Πίνακα 3.8 κάτι που δεν συμβαίνει στη δεύτερη αλλά και στην τρίτη επίσκεψη. Σε αντίθεση με την μέθοδο PWP, η μέθοδος WLW δείχνει ότι όντως κατά την πρώτη επίσκεψη υπάρχει μια αισθητή επίδραση της θεραπείας, όμως από την τρίτη επίσκεψη και έπειτα οι δύο μέθοδοι δε δείχνουν να συμφωνούν. Αξίζει πάντως να αναφερθεί ότι οι p-τιμές που αφορούν τη θεραπεία στην πρώτη επίσκεψη ($p_{PWP} = 0.0056$, $p_{WLW} = 0.0028$) διαφέρουν σημαντικά με τις p-τιμές που αφορούν την τρίτη επίσκεψη και μετά ($p_{PWP} = 0.9700$, $p_{WLW} = 0.0940$) κάτι που σημαίνει ότι οι θεραπείες στην πρώτη επίσκεψη είχε αποτελέσματα αλλά κατά τη διάρκεια των επόμενων επισκέψεων δε δείχνει να διαφέρει και πολύ από το πλασίμπο φάρμακο που χορηγούνταν σε ομάδα ασθενών.

	fev1	fev2	trt
	coef	coef	coef
PWP			
1η επίσκεψη	0,04212	-0,0624	-0,3613
2η επίσκεψη	-0,03469	0,04128	-0,19911
≥3	0,05351	-0,05188	-0,00416
WLW			
1η επίσκεψη	0,0419	-0,0626	-0,391
2η επίσκεψη	0,0572	-0,0748	-0,0804
≥3	0,3152	-0,329	-0,7295

Πίνακας 3.8: Αποτελέσματα από την ανάλυση των μεθόδων PWP και WLW για κάθε επίσκεψη ξεχωριστά.

Παρακάτω θα κάνουμε μια τελευταία ανάλυση των δεδομένων με τη μέθοδο AG, PWP και πρώτης επίσκεψης, κάνοντας χρήση των μοντέλων ευπάθειας.

Οι εντολές λοιπόν για το μοντέλο AG είναι:

```
fracfit <- coxph(Surv(time1, time2, status) ~ (trt + fev1 + fev2 + frailty(id, dist='gamma')), data=rhDNase1 )
```

Και τα αποτελέσματα είναι:

frfit

Call:

```
coxph(formula = Surv(time1, time2, status) ~ (trt + fev1 + fev2 +  
      frailty(id, dist = "gamma")), data = rhDNase1)
```

	coef	se(coef)	se2	Chisq	DF	p
trt	-0.2838	0.1451	0.0789	3.83	1	0.05
fev1	0.0938	0.0759	0.0412	1.53	1	0.22
fev2	-0.1122	0.0760	0.0413	2.18	1	0.14
frailty(id, dist = "gamma				969.83	443	0.00

Iterations: 7 outer, 26 Newton-Raphson

Variance of random effect= 2.29 I-likelihood = -4366.3

Degrees of freedom for terms= 0.3 0.3 0.3 443.5

Likelihood ratio test=1346 on 444 df, p=0 n= 1326

**Η ακόλουθη εντολή είναι για την προσαρμογή του PWP μοντέλου
ευπάθειας:**

```
frfitc <- coxph(Surv(time1, time2, status) ~ (trt + fev1 + fev2 +  
frailty(id, dist='gamma') + strata(enum)), data=rhDNase1 )
```

Τα αποτελέσματα της παραπάνω εφαρμογής δίνονται ως:

frfitc

Call:

```
coxph(formula = Surv(time1, time2, status) ~ (trt + fev1 + fev2 +  
frailty(id, dist = "gamma") + strata(enum)), data = rhDNase1)
```

	coef	se(coef)	se2	Chisq	DF	p
trt	-0.1832	0.0817	0.0777	5.03	1	0.025
fev1	0.0200	0.0430	0.0410	0.22	1	0.640
fev2	-0.0274	0.0431	0.0411	0.40	1	0.530
frailty(id, dist = "gamma				40.47	34	0.210

Iterations: 9 outer, 50 Newton-Raphson

Variance of random effect= 0.0562 I-likelihood = -3009.8

Degrees of freedom for terms= 0.9 0.9 0.9 34.0

Likelihood ratio test=95.3 on 36.8 df, p=4.34e-07 n= 1326

Η εντολή για το μοντέλο ευπάθειας που αφορά την ανάλυση μέχρι την εκδήλωση του πρώτου γεγονότος είναι:

```
frfit1 <- coxph(Surv(time1, time2, status) ~ (trt + fev1 + fev2 + frailty(id, dist='gamma')), data=rhDNase1, subset=(enum==1))
```

Τα αντίστοιχα αποτελέσματα της ανάλυσης είναι:

```
frfit1
```

```
Call:
```

```
coxph(formula = Surv(time1, time2, status) ~ (trt + fev1 + fev2 + frailty(id, dist = "gamma")), data = rhDNase1, subset = (enum == 1))
```

	coef	se(coef)	se2	Chisq	DF	p
trt	-0.4507	0.1587	0.1326	8.07	1	0.0045
fev1	0.0736	0.0845	0.0704	0.76	1	0.3800
fev2	-0.0991	0.0847	0.0706	1.37	1	0.2400
frailty(id, dist = "gamma")				188.93	170	0.1500

```
Iterations: 6 outer, 37 Newton-Raphson
```

```
Variance of random effect= 0.873 I-likelihood = -1491.7
```

```
Degrees of freedom for terms= 0.7 0.7 0.7 170.2
```

```
Likelihood ratio test=379 on 172 df, p=0 n= 645
```

	trt
	coef
First Event	-0,4507
Andersen-Gill	-0,2838
PWP	-0,1832

Πίνακας 3.9: Τα αποτελέσματα για την ανάλυση των μεθόδων Andersen-Gill, Prentice-Williams-Peterson και First Event με χρήση μοντέλων ευπάθειας.

Συγκρίνοντας τα αποτελέσματα του μοντέλου των Andersen Gill που βλέπουμε και στον Πίνακα 3.9 με το αποτέλεσμα που προκύπτει από το μοντέλο που εστιάζει στο πρώτο γεγονός, παρατηρούμε μια αισθητή μείωση στην επίδραση της θεραπείας μετά το πρώτο γεγονός. Το ίδιο προκύπτει και συγκρίνοντας το μοντέλο του πρώτου γεγονότος και με το μοντέλο των PWP

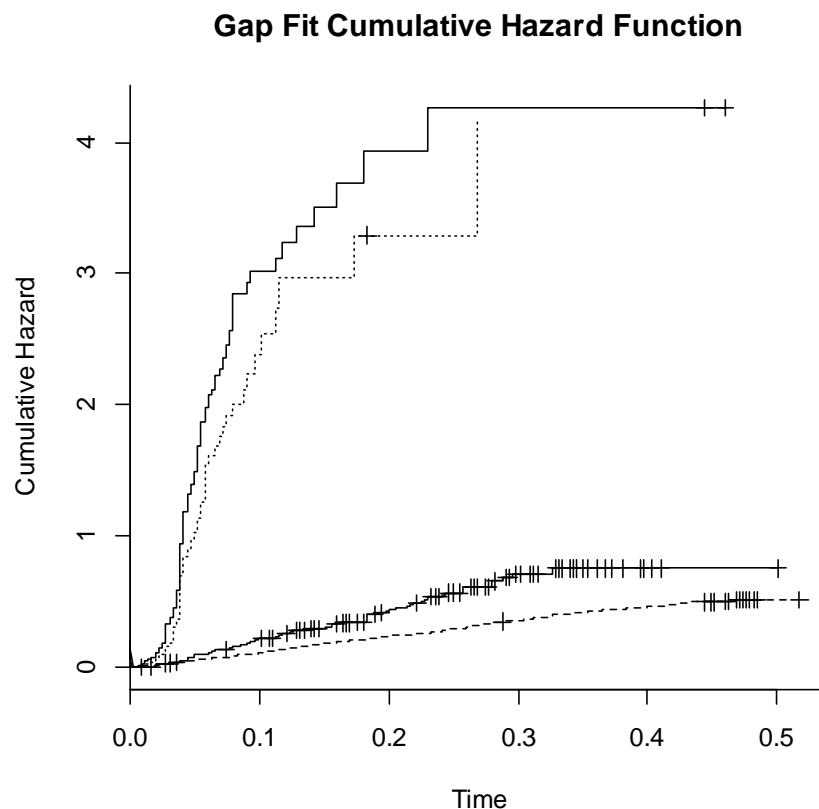
όπου εκεί παρατηρείτε μια μεγαλύτερη πτώση. Αξίζει να σημειωθεί πάντως ότι η χρήση των μοντέλων ευπάθειας έδειξε μια πιο αισθητή μείωση στην επίδραση της θεραπείας με τη μέθοδο PWP. Το ίδιο συμβαίνει και εδώ με το βαθμό βιαίως εκπεμπόμενου όγκου αέρα όπου μετά την πρώτη επίσκεψη μειώνεται αρκετά.

Τέλος θα δούμε ένα διάγραμμα ενδιάμεσων χρόνων όπου θα παρατηρήσουμε την εξέλιξη των γεγονότων με την πάροδο του χρόνου. Το διάγραμμα που χρειαζόμαστε είναι το παρακάτω:

```
gapfit <- coxph(Surv(time2-time1, status) ~ (trt + fev1 + fev2 +
strata(enum)), data=rhDNase1, subset=(enum<5))

curve <- survfit(gapfit)

plot(curve, lty=c(2,1,1,3,3,3,3), fun='cumhaz', xscale=365, ylab =
"Cumulative Hazard", xlab = "Time", bty="L", main= "Gap Fit Cumulative
Hazard Function" )
```



Σχήμα 3.2: Διάγραμμα ενδιάμεσων χρόνων (-- Γεγονός 1, - Γεγονότα 2 και 3, ... Γεγονός 4)

Σύμφωνα με το Σχήμα 3.2 η διάρκεια του πρώτου γεγονότος είναι πολύ μεγαλύτερη από τη διάρκεια των άλλων ενδιάμεσων χρόνων. Πάντως αξίζει να σημειωθεί ότι οι διάρκειες του δεύτερου αλλά και του τρίτου γεγονότος είναι εξίσου μεγάλη με την διάρκεια του πρώτου γεγονότος.

3.3 rIFN-g σε ασθενείς με χρόνια κοκκιωματώδη ασθένεια

Η χρόνια κοκκιωματώδης ασθένεια (CGD) είναι μια ετερογενής ομάδα ασυνήθιστων κληρονομικών ανωμαλιών που χαρακτηρίζονται από επαναλαμβανόμενες πυογενείς μολύνσεις, οι οποίες συνήθως ξεκινούν κατά τα πρώτα στάδια της παιδικής ηλικίας και πολλές φορές καταλήγει στο θάνατο. Η ιντερφερόνη γάμμα είναι ένας πρωταρχικός παράγοντας που εμφανίζεται για να διορθώσει μερικώς διάφορες μεταβολικές ατέλειες των φαγοκυττάρων, και για αυτό το λόγο υποθέτουμε ότι θα μειώσει τη συχνότητα των μολύνσεων σε ασθενείς με CGD (Fleming και Harrington 1991). Το 1986 η εταιρεία Genentech εκτέλεσε μια τυχαιοποιημένη μελέτη όπου περιλαμβάνονταν ασθενείς που λάμβαναν το φάρμακο rIFN-g και ασθενείς που λάμβαναν κάποιο πλασίμπο φάρμακο. Συνολικά στη μελέτη πήραν μέρος 128 ασθενείς που έπασχαν από CGD και λάμβαναν τη θεραπευτική αγωγή 3 φορές τη μέρα και με συνολική διάρκεια ένα χρόνο. Το πρώτο τερματικό σημείο της μελέτης ήταν ο χρόνος μέχρι την πρώτη σοβαρή μόλυνση. Τριάντα από τους 65 ασθενείς στην ομάδα των πλασίμπο και 14 από του 63 ασθενείς στην ομάδα των rIFN-g είχαν τουλάχιστον μια σοβαρή μόλυνση. Ο συνολικός αριθμός των μολύνσεων ήταν 56 και 20 στην ομάδα των πλασίμπο και των rIFN-g αντίστοιχα.

Το σύνολο δεδομένων *cgd1* έχει 203 παρατηρήσεις για 128 αντικείμενα μελέτης. Υπάρχει ένας ασθενής με 7 επαναλαμβανόμενες μολύνσεις. Αυτό σημαίνει ότι το σύνολο δεδομένων *cgd2* για την ανάλυση της μεθόδου WLW θα έχει 7 επαναλήψεις για κάθε ασθενή, δηλαδή $128 \times 7 = 896$ παρατηρήσεις. Το σύνολο δεδομένων rIFN-g που θα παρουσιαστεί παρακάτω διαμορφώθηκε για το μοντέλο των Andersen-Gill αλλά και των Prentice-

Williams-Peterson. Παρακάτω παρουσιάζεται ένα μικρό μέρος του συνόλου δεδομένων.

id	tstart	tstop	enum	treat	center	age	steroids	height	weight	status
1	0	219	1	1	204	12	0	147	62	1
1	219	373	2	1	204	12	0	147	62	1
1	373	414	3	1	204	12	0	147	62	0
2	0	8	1	0	204	15	0	159	47,5	1
2	8	26	2	0	204	15	0	159	47,5	1
2	26	152	3	0	204	15	0	159	47,5	1
2	152	241	4	0	204	15	0	159	47,5	1
2	241	249	5	0	204	15	0	159	47,5	1
2	249	322	6	0	204	15	0	159	47,5	1
2	322	350	7	0	204	15	0	159	47,5	1
2	350	439	8	0	204	15	0	159	47,5	0
3	0	382	1	1	204	19	0	171	72,7	0
4	0	388	1	1	204	12	0	142	34	0
5	0	246	1	0	238	17	0	162,5	52,7	1
5	246	253	2	0	238	17	0	162,5	52,7	1
5	253	383	3	0	238	17	0	162,5	52,7	0

Πίνακας 3.10: Τα δεδομένα cgd1 για την ανάλυση των μεθόδων Andersen-Gill, Prentice-Williams-Peterson και First Event.

id: Αύξων αριθμός ασθενούς

center: κέντρο νοσηλείας

treat: πλασίμπο ή ιντερφερόνη γάμμα

age: ηλικία ασθενούς

height: ύψος ασθενούς σε cm

weight: βάρος ασθενούς σε kg

steroids: χρήση στεροειδών από ασθενή, 1=ναι, 0=όχι

tstart, tstop: έναρξη και λήξη κάθε διαστήματος χρόνου

status: 1=όταν υπάρχει μόλυνση στο συγκεκριμένο διάστημα χρόνου

enum: παρατηρούμενος αριθμός μόλυνσης

Ένα μικρό μέρος των δεδομένων σε μορφή κατάλληλη για τη μέθοδο WLW παρουσιάζεται παρακάτω:

id	centre	treat	age	height	weight	enum	tstop	status
1	204	1	12	147,0	62,0	1	219	1
1	204	1	12	147,0	62,0	2	373	1
1	204	1	12	147,0	62,0	3	414	0
1	204	1	12	147,0	62,0	4	414	0
1	204	1	12	147,0	62,0	5	414	0
1	204	1	12	147,0	62,0	6	414	0
1	204	1	12	147,0	62,0	7	414	0
2	204	0	15	159,0	47,5	1	8	1
2	204	0	15	159,0	47,5	2	26	1
2	204	0	15	159,0	47,5	3	152	1
2	204	0	15	159,0	47,5	4	241	1
2	204	0	15	159,0	47,5	5	249	1
2	204	0	15	159,0	47,5	6	322	1
2	204	0	15	159,0	47,5	7	350	1
3	204	1	19	171,0	72,7	1	382	0
3	204	1	19	171,0	72,7	2	382	0
3	204	1	19	171,0	72,7	3	382	0
3	204	1	19	171,0	72,7	4	382	0
3	204	1	19	171,0	72,7	5	382	0
3	204	1	19	171,0	72,7	6	382	0
3	204	1	19	171,0	72,7	7	382	0
4	204	1	12	142,0	34,0	1	388	0
4	204	1	12	142,0	34,0	2	388	0
4	204	1	12	142,0	34,0	3	388	0
4	204	1	12	142,0	34,0	4	388	0
4	204	1	12	142,0	34,0	5	388	0
4	204	1	12	142,0	34,0	6	388	0
4	204	1	12	142,0	34,0	7	388	0

Πίνακας 3.11: Τα δεδομένα cgd2 για την ανάλυση της μεθόδου των Wei, Lin και Weissfeld.

Μετά την κατασκευή των πινάκων το επόμενο βήμα είναι να προσαρμόσουμε τα τρία μοντέλα τα οποία αναφέραμε προηγουμένως. Χρησιμοποιήθηκαν στην ανάλυση των δεδομένων οι μεταβλητές age, height, weight και steroids.

Οι εντολή για την ανάλυση AG είναι η ακόλουθη:

```
afit <- coxph(Surv(tstart, tstop, status) ~ (treat + height + weight
+ age + steroids + cluster(id)), data=cgd1)
```

Τα αποτελέσματα που προκύπτουν από την προηγούμενη εντολή είναι τα ακόλουθα:

```
afit
```

```
Call:
```

```
coxph(formula = Surv(tstart, tstop, status) ~ (treat + height +
weight + age + steroids + cluster(id)), data = cgd1)
```

	coef	exp(coef)	se(coef)	robust se	z	p
treat	-1.1095	0.330	0.3182	0.402	-2.761	0.0058
height	0.0138	1.014	0.0140	0.017	0.809	0.4200
weight	0.0203	1.020	0.0192	0.020	1.014	0.3100
age	-0.1385	0.871	0.0518	0.051	-2.718	0.0066
steroids	0.5748	1.777	1.0588	1.097	0.524	0.6000

```
Likelihood ratio test=21.3 on 5 df, p=0.000718 n= 128, number of
events= 56
```

Η ακόλουθη εντολή είναι για την προσαρμογή του PWP μοντέλου:

```
fitc <- coxph(Surv(tstart, tstop, status) ~ (treat + height + weight
+ age + steroids + cluster(id) + strata(enum)), data=cgd1)
```

Τα αποτελέσματα που προκύπτουν είναι τα εξής:

```
fitc
```

```
Call:
```

```
coxph(formula = Surv(tstart, tstop, status) ~ (treat + height +
weight + age + steroids + cluster(id) + strata(enum)), data =
cgd1)
```

	coef	exp(coef)	se(coef)	robust se	z	p
treat	-0.85838	0.424	0.3499	0.3764	-2.280	0.023
height	0.02157	1.022	0.0154	0.0140	1.544	0.120


```

weight    0.00351    1.004    0.0209    0.0193    0.182    0.860
age       -0.11076    0.895    0.0518    0.0403   -2.746    0.006
steroids  0.61234     1.845    1.0774    0.9520    0.643    0.520

```

Likelihood ratio test=9.68 on 5 df, p=0.0847 n= 128, number of events= 56

Για την ανάλυση του WLW χρειάζεται η ακόλουθη εντολή:

```
wfit <- coxph(Surv(tstop, status) ~ (treat + height + weight + age +
cluster(id) + strata(enum)), data=cgd2)
```

Μετά την προσαρμογή προκύπτουν τα ακόλουθα αποτελέσματα:

```
wfit
```

```
Call:
```

```
coxph(formula = Surv(tstop, status) ~ (treat + height + weight +
age + cluster(id) + strata(enum)), data = cgd2)
```

	coef	exp(coef)	se(coef)	robust se	z	p
treat	-1.468994	0.230	0.27763	0.3799	-3.87	0.00011
height	-0.000392	1.000	0.00937	0.0130	-0.03	0.98000
weight	0.027841	1.028	0.01493	0.0172	1.62	0.11000
age	-0.100724	0.904	0.03408	0.0410	-2.46	0.01400

Likelihood ratio test=43 on 4 df, p=1.05e-08 n= 896, number of events= 76

Η τελευταία ανάλυση και η εντολή που θα χρησιμοποιηθεί είναι για την ανάλυση μέχρι την εκδήλωση του πρώτου γεγονότος:

```
fit1 <- coxph(Surv(tstart, tstop, status) ~ (treat + height + weight
+ age + steroids + cluster(id)), data=cgd1, subset=(enum==1))
```

Τα αποτελέσματα που προκύπτουν είναι:

```
fit1
```

```
Call:
```

```
coxph(formula = Surv(tstart, tstop, status) ~ (treat + height +
```

```
weight + age + steroids + cluster(id)), data = cgd1, subset =
(enum == 1))
```

	coef	exp(coef)	se(coef)	robust se	z	p
treat	-0.98915	0.372	0.4299	0.4538	-2.180	0.029
height	-0.00361	0.996	0.0190	0.0186	-0.195	0.850
weight	0.03327	1.034	0.0261	0.0255	1.305	0.190
age	-0.10280	0.902	0.0595	0.0523	-1.966	0.049
steroids	0.64618	1.908	1.0865	1.0678	0.605	0.550

Likelihood ratio test=9.84 on 5 df, p=0.0798 n= 74, number of events= 30

Με μια πρώτη εκτίμηση θα λέγαμε ότι ο συντελεστής της θεραπείας του μοντέλου των Andersen-Gill είναι ελαφρώς μεγαλύτερος από αυτόν του μοντέλου που εστιάζει στο πρώτο γεγονός. Μείωση σε σχέση με το πρώτο γεγονός βλέπουμε ότι παρουσιάζει ο συντελεστής της θεραπείας με βάση το μοντέλο των PWP. Αν το επίπεδο σημαντικότητας μας στην ανάλυση είναι το 5% τότε με σιγουριά μπορούμε να πούμε ότι με βάση τις p-τιμές κάθε προσαρμογής ($p_{AG} = 0.0058$, $p_{PWP} = 0.023$, $p_{WLW} = 0.00011$, $p_{first} = 0.029$) τα δύο είδη θεραπείας διαφέρουν μεταξύ τους.

	treat	age	height
	coef	coef	coef
AG	-1,1095	-0,1385	0,0138
PWP	-0,85838	-0,11076	0,02157
WLW	-1,468994	-0,100724	-0,000392
1η επίσκεψη	-0,98915	-0,1028	-0,00361

Πίνακας 3.12: Αποτελέσματα που προκύπτουν μετά την ανάλυση των τεσσάρων μεθόδων.

Παρακάτω θα προσπαθήσουμε να αναλύσουμε χρησιμοποιώντας τα μοντέλα των PWP και WLW, τα αποτελέσματα του φαρμάκου rIFN-g

ξεχωριστά για την πρώτη, τη δεύτερη και την τρίτη επίσκεψη με σκοπό να διακρίνουμε αν όντως υπάρχουν σημαντικές επιδράσεις του.

Οι εντολές που αφορούν το μοντέλο του PWP είναι οι παρακάτω:

```
treat1<- cgd1$treat *(cgd1$enum==1)
treat2<- cgd1$treat *(cgd1$enum==2)
treat3<- cgd1$treat *(cgd1$enum>=3)

fitc13 <- coxph(Surv(tstart, tstop, status) ~ (treat1 + treat2 +
treat3 + height + weight + age + steroids + cluster(id) + strata(enum)),
data=cgd1)
```

Τα αποτελέσματα που προκύπτουν είναι τα ακόλουθα:

```
fitc13
```

```
Call:
```

```
coxph(formula = Surv(tstart, tstop, status) ~ (treat1 + treat2 +
treat3 + height + weight + age + steroids + cluster(id) +
strata(enum)), data = cgd1)
```

	coef	exp(coef)	se(coef)	robust se	z	p
treat1	-1.005006	0.366	0.4158	0.4382	-2.2933	0.0220
treat2	-0.155789	0.856	0.6577	0.6575	-0.2369	0.8100
treat3	-1.426464	0.240	1.1687	0.5792	-2.4627	0.0140
height	0.023370	1.024	0.0156	0.0146	1.5956	0.1100
weight	-0.000421	1.000	0.0212	0.0203	-0.0207	0.9800
age	-0.108080	0.898	0.0515	0.0386	-2.7987	0.0051
steroids	0.678904	1.972	1.0835	0.9295	0.7304	0.4700

```
Likelihood ratio test=11.2 on 7 df, p=0.132 n= 128, number of
events= 56
```

Οι αντίστοιχες εντολές για το μοντέλο WLW είναι :

```
treat1<- cgd2$treat *(cgd2$enum==1)
treat2<- cgd2$treat *(cgd2$enum==2)
```

```
treat3<- cgd2$treat *(cgd2$enum>=3)
```

Αποτελέσματα από την εφαρμογή δίνονται ως:

```
wfit13 <- coxph(Surv(tstop, status) ~ (treat1 + treat2 + treat3 +  
height + weight + age + cluster(id) + strata(enum)), data=cgd2)
```

```
wfit13
```

```
Call:
```

```
coxph(formula = Surv(tstop, status) ~ (treat1 + treat2 + treat3 +  
height + weight + age + cluster(id) + strata(enum)), data = cgd2)
```

	coef	exp(coef)	se(coef)	robust se	z	p
treat1	-1.22e+00	0.2959	0.34000	0.3521	-3.45863	0.00054
treat2	-1.41e+00	0.2449	0.57675	0.5821	-2.41679	0.01600
treat3	-2.84e+00	0.0585	1.03686	1.0620	-2.67379	0.00750
height	-6.84e-05	0.9999	0.00941	0.0131	-0.00523	1.00000
weight	2.77e-02	1.0281	0.01500	0.0173	1.60610	0.11000
age	-1.01e-01	0.9042	0.03418	0.0412	-2.44575	0.01400

```
Likelihood ratio test=46.1 on 6 df, p=2.79e-08 n= 896, number of  
events= 76
```

Παρατηρώντας τα παραπάνω αποτελέσματα αλλά και τον πίνακα που ακολουθεί, βλέπουμε ότι με τη μέθοδο PWP στην πρώτη επίσκεψη καθώς και στην τρίτη υπάρχει μια αισθητή επίδραση της θεραπείας, κάτι όμως που δεν ισχύει και για τη δεύτερη επίσκεψη. Σε αντίθεση με τη μέθοδο PWP, η μέθοδος WLW δείχνει ότι και στις τρεις επισκέψεις υπάρχει μια αισθητή διαδοχική αυξανόμενη επίδραση της θεραπείας.

	treat
	coef
PWP	
1η επίσκεψη	-1,00501
2η επίσκεψη	-0,15579
3η επίσκεψη	-1,42646
WLW	
1η επίσκεψη	-1,22
2η επίσκεψη	-1,41
3η επίσκεψη	-2,84

Πίνακας 3.13: Τα αποτελέσματα που προκύπτουν μετά την ανάλυση των μεθόδων PWP και WLW ξεχωριστά για κάθε επίσκεψη.

Παρακάτω θα κάνουμε μια τελευταία ανάλυση των δεδομένων με την βοήθεια των μεθόδων AG, PWP και πρώτης επίσκεψης, κάνοντας χρήση των μοντέλων ευπάθειας.

Οι εντολές λοιπόν για το μοντέλο AG είναι:

```
frafit <- coxph(Surv(tstart, tstop, status) ~ (treat + height +
weight + age + steroids + frailty(id, dist='gamma')), data=cgd)
```

Και τα αποτελέσματα είναι:

```
frafit
```

```
Call:
```

```
coxph(formula = Surv(tstart, tstop, status) ~ (treat + height +
weight + age + steroids + frailty(id, dist = "gamma")), data =
cgd)
```

	coef	se(coef)	se2	Chisq	DF	p
treatrIFN-g	-1.10631	0.3059	0.2691	13.08	1.0	0.0003
height	0.00705	0.0127	0.0108	0.31	1.0	0.5800
weight	0.02221	0.0189	0.0162	1.39	1.0	0.2400
age	-0.11077	0.0439	0.0392	6.38	1.0	0.0120
steroids	1.48100	0.7535	0.5864	3.86	1.0	0.0490

```
frailty(id, dist = "gamma" 37.90 26.7 0.0740
```

```
Iterations: 7 outer, 31 Newton-Raphson
```

```
Variance of random effect= 0.557 I-likelihood = -321.8
```

```
Degrees of freedom for terms= 0.8 0.7 0.7 0.8 0.6 26.7
```

```
Likelihood ratio test=92.5 on 30.4 df, p=3.31e-08 n= 203
```

Η ακόλουθη εντολή είναι για την προσαρμογή του PWP μοντέλου ευπάθειας:

```
frfitc <- coxph(Surv(tstart, tstop, status) ~ (treat + height +  
weight + age + steroids + strata(enum) + frailty(id, dist='gamma')),  
data=cgd)
```

Τα αποτελέσματα της παραπάνω εφαρμογής δίνονται ως:

```
frfitc
```

```
Call:
```

```
coxph(formula = Surv(tstart, tstop, status) ~ (treat + height +  
weight + age + steroids + strata(enum) + frailty(id, dist =  
"gamma")), data = cgd)
```

	coef	se(coef)	se2	Chisq	DF	p
treatrIFN-g	-1.01021	0.2919	0.2919	11.98	1	0.00054
height	0.00846	0.0113	0.0113	0.56	1	0.45000
weight	0.01685	0.0166	0.0166	1.03	1	0.31000
age	-0.09848	0.0383	0.0383	6.60	1	0.01000
steroids	1.12222	0.6058	0.6058	3.43	1	0.06400
frailty(id, dist = "gamma"				0.00	0	0.93000

```
Iterations: 6 outer, 25 Newton-Raphson
```

```
Variance of random effect= 5e-07 I-likelihood = -242.3
```

```
Degrees of freedom for terms= 1 1 1 1 1 0
```

```
Likelihood ratio test=19.9 on 5 df, p=0.00132 n= 203
```

Η εντολή για το μοντέλο ευπάθειας που αφορά την ανάλυση μέχρι την εκδήλωση του πρώτου γεγονότος είναι:

```
frfit1 <- coxph(Surv(tstart, tstop, status) ~ (treat + height +
weight + age + steroids + frailty(id, dist='gamma')), data=cgd,
subset=(enum==1))
```

Τα αντίστοιχα αποτελέσματα της ανάλυσης είναι:

```
frfit1
```

```
Call:
```

```
coxph(formula = Surv(tstart, tstop, status) ~ (treat + height +
weight + age + steroids + frailty(id, dist = "gamma")), data = cgd, subset
=(enum == 1))
```

	coef	se(coef)	se2	Chisq	DF	p
treatrIFN-g	-1.21738	0.3474	0.3474	12.28	1	0.00046
height	-0.00362	0.0118	0.0118	0.09	1	0.76000
weight	0.02799	0.0189	0.0189	2.18	1	0.14000
age	-0.08133	0.0404	0.0404	4.05	1	0.04400
steroids	1.22226	0.7678	0.7678	2.53	1	0.11000
frailty(id, dist = "gamma				0.00	0	0.94000

```
Iterations: 6 outer, 23 Newton-Raphson
```

```
Variance of random effect= 5e-07 I-likelihood = -184.8
```

```
Degrees of freedom for terms= 1 1 1 1 1 0
```

```
Likelihood ratio test=18.6 on 5 df, p=0.00223 n= 128
```

Συγκρίνοντας τα αποτελέσματα του μοντέλου των Andersen Gill και στον Πίνακα 3.14 με το αποτέλεσμα που προκύπτει από το μοντέλο που εστιάζει στο πρώτο γεγονός, παρατηρούμε μια αισθητή μείωση στην επίδραση της θεραπείας μετά το πρώτο γεγονός. Το ίδιο προκύπτει και συγκρίνοντας το μοντέλο του πρώτου γεγονότος και με το μοντέλο των PWP όπου εκεί παρατηρείτε μια μεγαλύτερη πτώση. Αξίζει να σημειωθεί πάντως ότι η χρήση των μοντέλων ευπάθειας έδειξε μια πιο αισθητή μείωση στην επίδραση της θεραπείας με τη μέθοδο PWP.

	treat
	coef
Andersen-Gill	-1,10631
PWP	-1,01021
First Event	-1,21738

Πίνακας 3.14: Τα δεδομένα cgd1 για την ανάλυση των μεθόδων Andersen-Gill, Prentice-Williams-Peterson και First Event με χρήση μοντέλων ευπάθειας

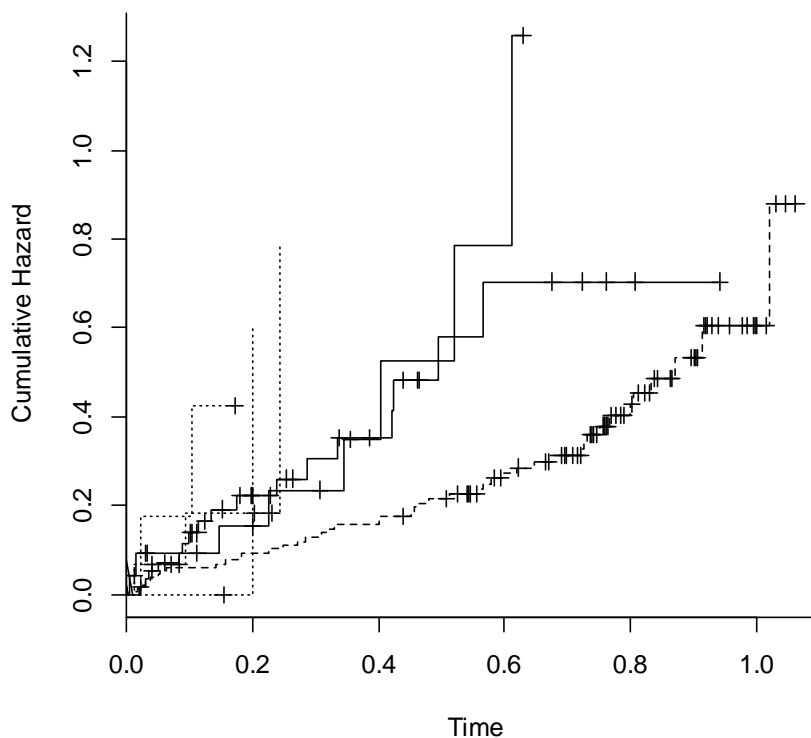
Είναι σημαντικό να δούμε ένα διάγραμμα ενδιάμεσων χρόνων όπου θα παρατηρήσουμε την εξέλιξη των γεγονότων με την πάροδο του χρόνου. Το διάγραμμα που χρειαζόμαστε είναι το παρακάτω:

```
gapfit <- coxph(Surv(tstop-tstart, status) ~ (treat + height + weight
+ age + steroids + cluster(id) + strata(enum)), data=cgd, subset=(enum<7))

curve <- survfit(gapfit)

plot(curve, lty=c(2,1,1,3,3,3,3), fun='cumhaz', xscale=365, ylab =
"Cumulative Hazard", xlab = "Time", bty="L", main= "Gap Fit Cumulative
Hazard Function" )
```


Gap Fit Cumulative Hazard Function

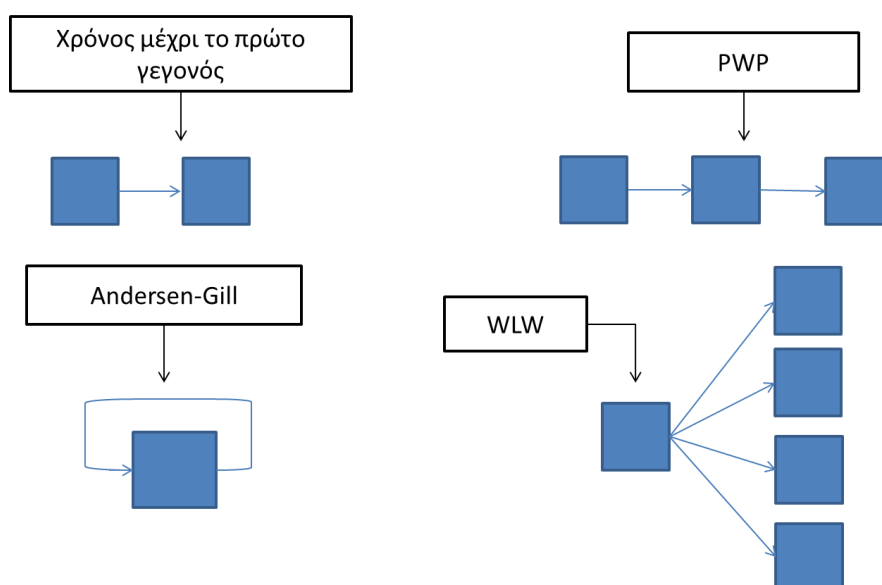


Σχήμα 3.3: Διάγραμμα ενδιάμεσων χρόνων (-- Γεγονός 1, —Γεγονότα 2 και 3, ... Γεγονότα 4, 5 και 6)

Σύμφωνα με το Σχήμα 3.3 η διάρκεια του πρώτου γεγονότος είναι πολύ μεγαλύτερη από τη διάρκεια των άλλων ενδιάμεσων χρόνων. Ένας πιθανός λόγος που μπορεί να συμβαίνει αυτό, είναι ότι για τη δεύτερη εμφάνιση κάποιου γεγονότος, επιλέγονται συνήθως άτομα που έχουν υψηλό ποσοστό εμφάνισης κάποιου γεγονότος.

Συμπεράσματα

Στην παρούσα διπλωματική εργασία είδαμε αρκετές μεθόδους που είναι χρήσιμες στην ανάλυση των επαναλαμβανόμενων γεγονότων. Η ανάλυση για παράδειγμα των δεδομένων του προσχολικού άσθματος απεικονίζει τις διαφορές των μεθόδων. Η τελική επιλογή για το ποιο μοντέλο ταιριάζει για την ανάλυση των δεδομένων που έχουμε κάθε φορά στην κατοχή μας, εξαρτάται από τους στόχους του ερευνητή. Αυτό σημαίνει ότι αν ο ερευνητής θέλει να διακρίνει τις επιδράσεις της θεραπείας σύμφωνα με τη σειρά που το γεγονός πραγματοποιείται, τότε θα επιλέξει το μοντέλο των Andersen-Gill. Ας υποθέσουμε ότι ένα αντικείμενο δεν μπορεί να είναι σε κίνδυνο για ένα k -στο γεγονός αν δεν έχει βιώσει $k-1$ γεγονότα. Σε αυτή την περίπτωση ο ερευνητής επιλέγει το μοντέλο των Prentice, Williams και Peterson και αποδίδει κάθε γεγονός ως ένα ξεχωριστό στρώμα. Η περιθώρια προσέγγιση χρησιμοποιείται όταν ο ερευνητής θέλει να εξετάσει τα γεγονότα που συμβαίνουν σε διαφορετική σειρά, όπως διαφορετικά είδη γεγονότων, για παράδειγμα διαφορετικές συνθήκες ασθένειας. Σε αυτήν την περίπτωση προτιμάται το μοντέλο των Wei, Lin και Weissfeld. Τέλος αν το ενδιαφέρον του ερευνητή επικεντρώνεται στο πρώτο γεγονός τότε μπορεί να γίνει ανάλυση του πρώτου γεγονότος. Ο καλύτερος τρόπος για να καταλάβουμε την διαφορά αυτών των μεθόδων είναι το παρακάτω σχήμα.



Αυτό που παρουσίασε ιδιαίτερο ενδιαφέρον στην ανάλυση των τριών μεθόδων και στα τρία σύνολα δεδομένων, είναι ότι η μέθοδος WLW έδειξε πιο ισχυρή επίδραση στη θεραπεία, με βάση την p-τιμή, απ' ότι οι μέθοδοι AG και PWP.

	p-value		
	bladder cancer	rhDNase	rIFN-g
	treat	treat	treat
AG	0,08	0,029	0,0058
WLW	0,058	0,014	0,00011
PWP	0,1	0,029	0,023

Βιβλιογραφία

- [1] Aalen, O.O., Husebye E. (1991). Statistical analysis of repeated events forming renewal processes. *Statistics in Medicine* 10, 1227-1240.
- [2] Andersen P.K., Gill, R.D. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics*. **10**, 1100-1120.
- [3] Andersen, P.K., Borgan, O., Gill, R.D., Keiding, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.
- [4] Andrews, D.F., Hertzberg, A.M. (1985). *DATA: A Collection of Problems from Many Fields for the Student and Research Worker*, New York: Springer-Verlag.
- [5] Barai, U., Teoh, N. (1997). Multiple statistics for multiple events, with application to repeated infections in the growth factor studies. *Statistics in Medicine* **16**, 941-949.
- [6] Berman, M., Turner, T.R. (1992). Approximating point process likelihoods. *Applied Statistics*.
- [7] Bowman, M.E. (1996). An evaluation of statistical models for the analysis of recurrent events data: with application to needlestick injuries among a cohort of female veterinarians, PhD thesis Ohio State University.
- [8] Clayton D. (1994). Some approaches to the analysis of recurrent event data. *Statistical Methods in Medical Research* 3, 244 –262.
- [9] Cook, R.J., Lawless, J.F., Nadeau, J.C. (1996). Robust tests for treatment comparisons based on recurrent event responses. *Biometrics* 52, 732-739.
- [10] Cook, R.J., Lawless, J.F. (1997) Discussion of paper by Wei and Glidden. *Statistics in Medicine* **17**, 841-851.
- [11] Cook, R.J., Lawless, J.F. (2007). *The Statistical Analysis of Recurrent Events*. New York: Springer.
- [12] Cox, D.R. (1972). Regression models and life-tables (with discussion). *J. Roy. Statistical Society Series B*. **34**, 187-220.
- [13] Cox, D.R. (1975). Partial likelihood. *Biometrika* **62**, 262-276.
- [14] Dabrowska, D.M., Sun, G., Horwitz, M.M. (1994). Cox regression in a Markov renewal model: an application to the analysis of bone marrow transplant data. *Journal of the American Statistical Association* 89, 867-877.
- [15] Dean, C.B., Balshaw, R. (1997). Efficiency lost by analysis counts rather than event times in Poisson and overdispersed Poisson regression models. *Journal of the American Statistical Association* 92, 1387-1398.
- [16] Fan, J., Lin, F. (2002). Variable selection for Cox's proportional hazards model and frailty model. *The Annals of Statistics* 30, 74 –99.
- [17] Fleming, T.R., Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. New York: John Wiley and Sons.

- [18] Follman, D., Goldberg M. (1988). Distinguishing heterogeneity from decreasing hazard rates. *Technometrics* 30, 389-396.
- [19] Gail, M., Santner, T., Brown, C. (1980). An analysis of comparative carcinogenesis experiments based on multiple times to tumor. *Biometrics* 36, 255-266.
- [20] Gao, S., Zhou, X.H. (1997). An empirical comparison of two semi-parametric approaches for the estimation of covariate effects from multivariate failure time data. *Statistics in Medicine* 1997; 16, 2049–2062.
- [21] Godambe, V.P., Kale, B.K. (1991). Estimating functions: an overview. In: Godambe V.P., ed. *Estimating Functions*. Oxford: Clarendon Press.
- [22] Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. New York: Springer-Verlag.
- [23] Huber, P.J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In: *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability*. University of California Press, Berkeley, CA, pp.221-233.
- [24] Kelly, P., Lim, L.L.-Y. (2000). Survival analysis for recurrent event data: An application to childhood infectious diseases. *Statistics in Medicine* 19, 13-33.
- [25] Klein, J.P., Moeschberger, M.L. (1997). *Survival Analysis Techniques for Censored and Truncated Data*. New York Springer.
- [26] Laird, N.M., Olivier, D. (1981). Covariance analysis of censored survival data using log-linear analysis techniques. *Journal of the American Statistical Association* 76, 231-240.
- [27] Lawless, J.F., Nadeau, J.C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics* 37, 158-168.
- [28] Lawless, J.F., Fong, D. (1999). State duration models in clinical and observational studies. *Statistics in Medicine* 18, 2365-2376.
- [29] Lawless, J.F., Wigg M.B., Tuli, S., Drake, J., Lamberti-Pasculli, M. (2001). Analysis of repeated shunt failures or durations, with applications to shunt failures for patients with pediatric hydrocephalus. *Applied Statistics* 50, 449-465.
- [30] Lee, E.W., Wei, L.J., Amato, D.A. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In: Klein, J.P., Goel, P.K. (Eds.), *Survival Analysis: State of the Art*. Kluwer Academic, Dordrecht, pp. 237-247.
- [31] Liang, K.Y., Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* 73, 13-22.
- [32] Liang, K.Y., Self, S.G., Chang, Y.C. (1993). Modeling marginal hazards in multivariate failure time data. *Journal of the Royal Statistical Society Series B.* 55, 441-453.
- [33] Lin, D.Y., Wei, L.J. (1989). The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association* 84, 1074-1078.

- [34] Lipschutz, K.H., Snapinn, S.M. (1997). Discussion of paper by Wei and Glidden. *Statistics in Medicine* **16**, 846-848.
- [35] Nelson, W.B. (2003). Recurrent Events Data Analysis for Product Repairs, Disease Recurrences, and Other Applications. ASA-SIAM Series on Statistics and Applied Probability, Philadelphia, PA.
- [36] Pepe, M.S., Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time-dependent covariates. *Journal of the American Statistical Association* **88**, 811-820.
- [37] Oakes, D.A. (1992). Frailty models for multiple event times. In Survival Analysis, State of the Art, Klein, J.P., Goel, P.K. (eds). Kluwer Academic Publishers: Dordrecht.
- [38] Ohman-Strickland, P. (2003). Fitting the frailty distribution using empirical Bayes. Presented at the International Conference on Reliability and Survival Analysis, Columbia, S.C.
- [39] Prentice, R.L., Williams, B.J., Peterson, A.V. (1981). On the regression analysis of multivariate failure time data. *Biometrika* **68**, 373-389.
- [40] Thall, P.F., Vail, S.C. (1995). Some covariate models for longitudinal count data with overdispersion. *Biometrics* **46**, 657-671.
- [41] Therneau, T.M., Grambsch P.M. (2000). Modeling Survival Data: Extending the Cox Model. New York: Springer.
- [42] Tsiatis, A.A. (1981). A large sample study of Cox's regression model. *The Annals of Statistics* **9**, 93-108.
- [43] Visser, M. (1996). Nonparametric estimation of the bivariate survival functions with an application to vertically transmitted AIDS. *Biometrika* **83**, 507-518.
- [44] Wang, W., Wells, M.T. (1998). Nonparametric estimation of successive duration times under dependent censoring. *Biometrika* **85**, 561-572.
- [45] Wei, L.J., Lin, D.Y., Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* **84**, 1065-1073.
- [46] Wei, L.J., Glidden, D.V. (1997). An overview of statistical methods for multiple failure time data in clinical trials. *Statistics in Medicine* **16**, 833-839.
- [47] Wei, L.J., Ying, Z., Lin, D.Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77**, 845-851.
- [48] White, H. (1982). Maximum likelihood estimations of misspecified models. *Econometrica* **50**, 1-25.