

ΜΠΣ ΒΙΟΣΤΑΤΙΣΤΙΚΗ

**ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ**

ΙΑΤΡΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΓΡΗΓΟΡΙΟΥ ΜΑΡΙΑ

«Ανάπτυξη κλίμακας για την εκτίμηση κινδύνου (points-based risk score)
υπό την παρουσία ανταγωνιστικών κινδύνων»

ΑΘΗΝΑ, 2019

Η παρούσα διπλωματική εργασία εκπονήθηκε στο πλαίσιο των σπουδών για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στη

ΒΙΟΣΤΑΤΙΣΤΙΚΗ

που απονέμει η Ιατρική Σχολή και το Τμήμα Μαθηματικών του Εθνικού & Καποδιστριακού Πανεπιστημίου Αθηνών

Εγκρίθηκε την *1/7/2019* από την εξεταστική επιτροπή:

ΟΝΟΜΑΤΕΠΩΝΥΜΟ	ΒΑΘΜΙΑ	ΥΠΟΓΡΑΦΗ
Β. ΣΥΨΑ (Επιβλέπουσα)	ΑΝΑΠΛ. ΚΑΘΗΓΗΤΡΙΑ
Γ. ΤΟΥΛΟΥΜΗ	ΚΑΘΗΓΗΤΡΙΑ
Γ. ΠΑΠΑΘΕΟΔΩΡΙΔΗΣ	ΚΑΘΗΓΗΤΗΣ

Ευχαριστίες

Θα ήθελα να πω ένα μεγάλο ευχαριστώ στην επιβλέπουσα καθηγήτρια της παρούσας διπλωματικής εργασίας, κυρία Βασιλική – Αναστασία (Βάνα) Σύψα, για τη στήριξη, την καθοδήγηση και τις πολύτιμες συμβουλές της. Οφείλω επίσης να ευχαριστήσω και τα μέλη της Επιτροπής, την καθηγήτρια κυρία Γιώτα Τουλούμη και τον καθηγητή κύριο Γιώργο Παπαθεοδωρίδη για την πολύτιμη συμβολή τους στην ολοκλήρωση της εργασίας. Τέλος, ένα τεράστιο ευχαριστώ στην οικογένεια μου η οποία ήταν δίπλα μου σε κάθε μου βήμα.

Περιεχόμενα

1. Εισαγωγή.....	7
1.1. Προγνωστικά μοντέλα.....	7
1.2. Προγνωστικά μοντέλα στην ανάλυση επιβίωσης και ανταγωνιστικοί κίνδυνοι.....	9
1.3. Προγνωστικά μοντέλα και χρόνια ηπατίτιδα Β.....	11
1.4. Σκοπός της εργασίας.....	13
2. Ανάλυση επιβίωσης και Ανταγωνιστικοί κίνδυνοι.....	14
2.1. Δεδομένα επιβίωσης.....	14
2.2. Βασικές έννοιες στην ανάλυση επιβίωσης.....	17
2.3. Μη παραμετρικές μέθοδοι για την εκτίμηση των συναρτήσεων επιβίωσης....	18
2.3.1. Εκτιμήτρια Kaplan-Meier.....	18
2.3.2. Πίνακες επιβίωσης – Life tables.....	22
2.4. Σύγκριση καμπυλών επιβίωσης.....	24
2.5. Μοντέλο αναλογικών κινδύνων Cox – Proportional hazard model.....	26
2.5.1. Έλεγχοι υποθέσεων.....	28
2.5.1.1. Έλεγχος του λόγου πιθανοφανειών – Likelihood ratio test.....	28
2.5.1.2. Έλεγχος του Wald – Wald test.....	29
2.5.2. Έλεγχοι της υπόθεσης αναλογικότητας των κινδύνων.....	29
2.5.2.1. Γραφικός έλεγχος.....	30
2.5.2.2. Έλεγχος μέσω καταλοίπων.....	31
2.6. Ανταγωνιστικοί κίνδυνοι.....	31
2.6.1. Στατιστική ανάλυση δεδομένων παρουσία ανταγωνιστικών κινδύνων....	32
2.6.1.1. Αθροιστική συνάρτηση κινδύνου (Cumulative Incidence Function).....	32
2.6.1.2. Συνάρτηση κινδύνου συγκεκριμένης αιτίας (cause-specific hazard	

function) και συνάρτηση κινδύνου υπό-κατανομής (subdistribution hazard function).....	34
2.6.2. Μοντέλα παλινδρόμησης κινδύνου παρουσία ανταγωνιστικών κινδύνων.....	35
2.6.2.1. Μοντελοποίηση κινδύνου συγκεκριμένης αιτίας.....	35
2.6.2.2. Μοντελοποίηση αθροιστικής επίπτωσης.....	36
3. Ανάπτυξη και Επικύρωση ενός προγνωστικού μοντέλου.....	39
3.1. Ανάπτυξη ενός προγνωστικού μοντέλου.....	39
3.1.1. Ερευνητικό ζήτημα και έλεγχος δεδομένων.....	39
3.1.2. Κωδικοποίηση των προγνωστικών μεταβλητών.....	40
3.1.3. Επιλογή τελικού μοντέλου.....	43
3.1.4. Εκτίμηση του μοντέλου.....	46
3.2. Αξιολόγηση του μοντέλου.....	49
3.2.1. Βαθμονόμηση του μοντέλου – Calibration.....	49
3.2.2. Ικανότητα διαχωριστικότητας του μοντέλου – Discrimination.....	51
3.2.2.1. Ευαισθησία και Ειδικότητα.....	52
3.2.2.2. Καμπύλη ROC.....	52
3.2.2.3. Το στατιστικό c (c-statistic ή c-index).....	53
3.2.3. Κλινική χρησιμότητα του μοντέλου – clinical usefulness.....	54
3.3. Επικύρωση του μοντέλου – Validation.....	57
3.3.1. Εσωτερική επικύρωση – Internal validation.....	58
3.3.1.1. Μέθοδος bootstrap.....	58
3.3.1.2. Μέθοδος διεπικύρωσης – Cross validation.....	59
3.3.2. Εξωτερική επικύρωση – External validation.....	60
3.3.2.1. Χρονική επικύρωση – Temporal validation.....	60
3.3.2.2. Γεωγραφική επικύρωση – Geographic validation.....	61

4. Δημιουργία κλίμακας (risk score) για την εκτίμηση του κινδύνου από ένα προγνωστικό μοντέλο παρουσία ανταγωνιστικών κινδύνων	62
4.1. Κλίμακα εκτίμησης κινδύνου απουσία ανταγωνιστικών κινδύνων	62
4.2. Κλίμακα εκτίμησης κινδύνου παρουσία ανταγωνιστικών κινδύνων	65
4.3. Αξιολόγηση κλίμακας εκτίμησης κινδύνου	69
5. Εφαρμογή σε δεδομένα	71
5.1. Ανάλυση δεδομένων	71
5.1.1. Παρακολούθηση (follow-up) και ορισμοί	72
5.1.2. Μονοπαραγοντική ανάλυση	74
5.1.3. Επιλογή τελικού μοντέλου	76
5.1.4. Τελικό μοντέλο	77
5.1.5. Σύγκριση τελικού μοντέλου με το μοντέλο από PAGE-B	77
5.2. Δημιουργία κλίμακας εκτίμησης κινδύνου εμφάνισης ΗΚΚ σε ασθενείς με χρόνια ηπατίτιδα Β	79
5.2.1. Κλίμακα εκτίμησης κινδύνου-Risk score	80
6. Συζήτηση	84
Περίληψη	
Ελληνικά.....	86
Αγγλικά.....	88
Ευρετήριο αγγλικών-ελληνικών	90
Βιβλιογραφία	95

Κεφάλαιο 1

Εισαγωγή

1.1. Προγνωστικά μοντέλα

Η πρόγνωση είναι μια πρόβλεψη η οποία εκτιμά τον κίνδυνο μελλοντικών εκβάσεων σε άτομα βάσει των κλινικών και μη κλινικών χαρακτηριστικών τους. Τα πιο συνηθισμένα παραδείγματα πρόβλεψης είναι οι καιρικές συνθήκες και οι οικονομικές προβλέψεις. Οι εκβάσεις είναι συνήθως ειδικά γεγονότα, όπως ο θάνατος ή επιπλοκές, αλλά επίσης μπορούν να έχουν διαφορετικό χαρακτήρα, όπως η πορεία της νόσου και αλλαγές στον πόνο ή στην ποιότητα ζωής. Η πρόγνωση μπορεί να διαμορφωθεί από την ηλικία, το φύλο, το ιστορικό και άλλα χαρακτηριστικά των ασθενών. Κατά τον Ιπποκράτη, το πιο σημαντικό για έναν γιατρό είναι η πρόγνωση μιας ασθένειας, αφού είναι αυτή που του επιτρέπει να προλάβει την εξέλιξη μιας νόσου, να κερδίσει την εμπιστοσύνη του ασθενούς αλλά και να καθορίσει την θεραπεία με πιο σωστό τρόπο (*Adams F, 1939*).

Οι προγνωστικές μελέτες πρέπει να χρησιμοποιούν μια πολυπαραγοντική προσέγγιση στον σχεδιασμό και στην ανάλυση επειδή μια μονοπαραγοντική προσέγγιση σπάνια δίνει μια επαρκή εκτίμηση της πρόγνωσης (*Karel et al, 2009*). Παρόλο που ένα προγνωστικό μοντέλο μπορεί να χρησιμοποιηθεί για να δώσει μια εικόνα για την αιτιότητα και την παθοφυσιολογία της μελετώμενης έκβασης, αυτός δεν είναι ούτε στόχος ούτε απαίτηση. Όλες οι μεταβλητές που σχετίζονται ενδεχομένως με την έκβαση, όχι απαραίτητα αιτιώδεις, μπορούν να ληφθούν υπόψη σε μια προγνωστική μελέτη. Σημειώνουμε ότι κάθε αιτιώδης παράγοντας είναι προγνωστικός αλλά κάθε προγνωστικός δεν είναι αιτιώδης. Για την εκτίμηση τέτοιων πιθανοτήτων έχουμε κάποια εργαλεία τα οποία ονομάζονται συνήθως προγνωστικά μοντέλα, κανόνες πρόβλεψης ή κλίμακες κινδύνου (*risk scores*). Τα μοντέλα αυτά παρέχουν την δυνατότητα στους γιατρούς να χρησιμοποιήσουν συνδυασμούς τιμών πρόβλεψης για να υπολογίσουν ένα απόλυτο κίνδυνο ή την πιθανότητα να εμφανίσει ένα άτομο μια έκβαση.

Τα προγνωστικά μοντέλα μπορούν να χρησιμοποιηθούν με πολλούς τρόπους και για πολλούς λόγους. Οι κυριότεροι λόγοι χρήσης τους είναι η ενημέρωση των ασθενών σχετικά με την μελλοντική πορεία της νόσου τους ή τον κίνδυνο εμφάνισης

ασθενειών και την καθοδήγηση των γιατρών και των ασθενών σε κοινές αποφάσεις για περαιτέρω θεραπεία. Για παράδειγμα, η κλίμακα καρδιαγγειακού κινδύνου Framingham (Kannel et al, 1976) χρησιμοποιείται ευρέως στην πρωτοβάθμια περίθαλψη για να αποφασίσουν οι γιατροί αν πρέπει να χορηγήσουν φάρμακα για χοληστερόλη και υπέρταση στους ασθενείς. Ένα παράδειγμα στη δευτεροβάθμια περίθαλψη είναι η χρήση του προγνωστικού δείκτη Nottingham για την εκτίμηση του μακροχρόνιου κινδύνου υποτροπής του καρκίνου ή του θανάτου σε ασθενείς με καρκίνο του μαστού (Galea et al, 1992). Ένας άλλος λόγος χρήσης των προγνωστικών μοντέλων είναι η επιλογή ασθενών σε κλινικές δοκιμές. Για παράδειγμα, ερευνητές χρησιμοποίησαν ένα υπάρχον προγνωστικό μοντέλο για να επιλέξουν γυναίκες με υψηλό κίνδυνο εμφάνισης καρκίνου για να συμμετάσχουν σε μια τυχαιοποιημένη κλινική δοκιμή αναφορικά με την ταμοξιφαίνη στην πρόληψη του καρκίνου του μαστού (Fisher et al, 1998). Επίσης, τα μοντέλα πρόβλεψης χρησιμοποιούνται για να συγκρίνουν τις διάφορες επιδόσεις μεταξύ των νοσοκομείων όπως για παράδειγμα, ο κλινικός δείκτης κινδύνου για βρέφη που αναπτύχθηκε για να συγκρίνει τις επιδόσεις και τη θνησιμότητα μεταξύ των μονάδων εντατικής θεραπείας των νεογνών (Cockburn et al, 2004). Τέλος, τα προγνωστικά μοντέλα μπορούν να χρησιμοποιηθούν για τον καθορισμό της συχνότητας παρακολούθησης του κάθε ασθενούς ανάλογα με την πρόγνωση του κινδύνου (Papatheodoridis et al, 2016).

Τα μοντέλα πρόβλεψης χρησιμοποιούνται ευρέως και έχουν αναπτυχθεί αρκετά τόσο με χρήση λογαριθμιστικής παλινδρόμησης (logistic regression) όσο και με χρήση ανάλυσης επιβίωσης μέσω του μοντέλου παλινδρόμησης αναλογικών κινδύνων Cox (Cox Proportional hazards model). Το μοντέλο λογαριθμιστικής παλινδρόμησης χρησιμοποιείται όταν η έκβαση είναι δίτιμη ενώ το μοντέλο αναλογικών κινδύνων χρησιμοποιείται όταν η έκβαση αφορά τον χρόνο μέχρι την εμφάνιση του γεγονότος ενδιαφέροντος. Μερικά παραδείγματα μοντέλων πρόβλεψης που έχουν ήδη αναπτυχθεί είναι το μοντέλο Framingham το οποίο υπολογίζει τον κίνδυνο εμφάνισης καρδιακών παθήσεων (Mahmood et al, 2014), το EUROSCORE το οποίο υπολογίζει τον κίνδυνο πρόωρης θνησιμότητας σε καρδιακά χειρουργημένους ασθενείς στην Ευρώπη (Nashef et al, 1999), το SCORES το οποίο υπολογίζει τον δεκαετή κίνδυνο εμφάνισης θανατηφόρας καρδιαγγειακής πάθησης σε ευρωπαϊκό πληθυσμό (Conroy et al, 2003), το PROCAM το οποίο υπολογίζει τον κίνδυνο εμφάνισης στεφανιαίας νόσου στον Γερμανικό πληθυσμό, βελτιώνοντας την πρόβλεψη και την έγκαιρη διάγνωση της νόσου (Assmann et al, 2002) και το

ASSIGN το οποίο υπολογίζει τον κίνδυνο εμφάνισης καρδιαγγειακής νόσου προσθέτοντας ως παράγοντες την κοινωνική στέρηση και το οικογενειακό ιστορικό (Woodward et al 2007). Στην παρούσα εργασία θα γίνει ανασκόπηση και περιγραφή των μοντέλων αναλογικών κινδύνων Cox λαμβάνοντας υπόψη και τους ανταγωνιστικούς κινδύνους. Επίσης, θα αναφερθούμε στους τρόπους επικύρωσης (validation) του μοντέλου πρόβλεψης, όπως η εξωτερική (external) και εσωτερική (internal) επικύρωση (Royston et al, 2013) και σε μέτρα αξιολόγησης της πρόβλεψης του μοντέλου όπως η βαθμονόμηση (calibration), η ικανότητα διαχωριστικότητας του μοντέλου (discrimination) και η κλινική του χρησιμότητα (clinical usefulness). Τέλος θα αναπτυχθεί μια κλίμακα εκτίμησης κινδύνου (risk score) παρουσία ανταγωνιστικών κινδύνων.

1.2. Προγνωστικά μοντέλα στην ανάλυση επιβίωσης και ανταγωνιστικοί κίνδυνοι

Οι ανταγωνιστικοί κίνδυνοι εμφανίζονται συχνά στην ανάλυση δεδομένων επιβίωσης. Ένας ανταγωνιστικός κίνδυνος είναι ένα γεγονός το οποίο αποκλείει την εμφάνιση του γεγονότος ενδιαφέροντος. Για παράδειγμα, σε μια μελέτη που εξετάζει τον χρόνο μέχρι τον θάνατο από καρδιαγγειακά αίτια, ο θάνατος που οφείλεται σε μη καρδιαγγειακά αίτια είναι ένας ανταγωνιστικός κίνδυνος. Σημαντικό είναι το γεγονός ότι κατά την εκτίμηση της επίπτωσης των εκβάσεων, οι αναλυτές θα πρέπει να χρησιμοποιήσουν τη θροιστική συνάρτηση επίπτωσης (CIF) και όχι τη συνάρτηση επιβίωσης επιβίωσης Kaplan-Meier. Η χρήση της συνάρτησης επιβίωσης Kaplan-Meier οδηγεί σε εκτιμήσεις της επίπτωσης που είναι μεροληπτικές προς τα πάνω, ανεξάρτητα από το αν τα ανταγωνιστικά γεγονότα είναι ανεξάρτητα το ένα από το άλλο.

Όταν εφαρμόζουν μοντέλα παλινδρόμησης με την παρουσία ανταγωνιστικών κινδύνων, οι ερευνητές μπορούν να επιλέξουν από 2 διαφορετικές οικογένειες μοντέλων: μοντελοποίηση της επίδρασης των προγνωστικών παραγόντων για τον κίνδυνο εμφάνισης της έκβασης ή μοντελοποίηση της επίδρασης των παραγόντων στην αθροιστική συνάρτηση επίπτωσης. Το πρώτο επιτρέπει σε κάποιον να εκτιμήσει την επίδραση των παραγόντων επί του ποσοστού εμφάνισης της έκβασης σε εκείνα

τα άτομα που είναι επί του παρόντος χωρίς το γεγονός ενδιαφέροντος. Το δεύτερο επιτρέπει την εκτίμηση της επίδρασης των μεταβλητών στον απόλυτο κίνδυνο της έκβασης με την πάροδο του χρόνου. Η πρώτη προσέγγιση μπορεί να είναι πιο κατάλληλη για την αντιμετώπιση των αιτιωδών ζητημάτων, ενώ η δεύτερη προσέγγιση μπορεί να είναι πιο κατάλληλη για την εκτίμηση της κλινικής πρόγνωσης ενός ασθενούς.

Οι ανταγωνιστικοί κίνδυνοι επικρατούν σε μεγάλο μέρος της καρδιαγγειακής έρευνας (*Austin et al, 2016*). Αγνοώντας τα ανταγωνιστικά γεγονότα έχουμε αρνητικές συνέπειες συμπεριλαμβανομένης της εσφαλμένης εκτίμησης της επίδρασης των παραγόντων στην εμφάνιση της έκβασης. Οι *Koller et al* διαπίστωσαν ότι οι ανταγωνιστικοί κίνδυνοι ήταν παρόντες σε μια μεγάλη πλειοψηφία των μελετών που δημοσιεύθηκαν σε ένα δείγμα περιοδικών μεγάλης εμβέλειας (*Koller et al, 2012*). Αυτό υποδηλώνει ότι είναι σημαντικό οι ερευνητές να γνωρίζουν τις κατάλληλες μεθόδους για την αντιμετώπιση των ανταγωνιστικών κινδύνων κατά την ανάλυση των δεδομένων επιβίωσης. Επίσης, οι *Wolbers et al* σύγκριναν την απόδοση ενός τυπικού μοντέλου επιβίωσης Cox με ένα μοντέλο Fine-Gray για την πρόβλεψη της επίπτωσης της στεφανιαίας νόσου στις μεγαλύτερες γυναίκες (*Wolbers et al, 2009*). Διαπίστωσαν ότι το μοντέλο Cox υπερεκτίμησε τον 10ετή κίνδυνο στεφανιαίας νόσου σε σύγκριση με την εκτίμηση από το μοντέλο Fine-Gray. Το 18% των γυναικών ταξινομήθηκαν ως υψηλού κινδύνου με βάση το μοντέλο Cox, ενώ μόνο το 8% των γυναικών ταξινομήθηκε ως υψηλού κινδύνου όταν χρησιμοποιήθηκε το μοντέλο Fine-Gray. Αποδίδουν αυτή την ασυμφωνία στον αυξημένο κίνδυνο θανάτου που οφείλεται σε ανταγωνιστικούς κινδύνους σε αυτόν τον ηλικιωμένο πληθυσμό. Τα πιο πάνω παραδείγματα φανερώνουν το πόσο σημαντικό είναι να λαμβάνουμε υπόψη την παρουσία ανταγωνιστικών κινδύνων κατά την ανάλυση των δεδομένων μας, ούτως ώστε οι προβλέψεις των μοντέλων να είναι ορθές μιας και από αυτές εξαρτάται πολλές φορές η πορεία της νόσου ενός ασθενή ή/και η θεραπεία του.

1.3. Προγνωστικά μοντέλα και χρόνια ηπατίτιδα Β

Η χρόνια λοίμωξη από τον ιό της ηπατίτιδας Β (HBV) παραμένει ένα σημαντικό παγκόσμιο πρόβλημα υγείας. Η ηπατίτιδα Β αποκαλείται χρόνια, όταν η λοίμωξη διαρκεί περισσότερο από έξι μήνες ελεγχόμενη με τους εργαστηριακούς δείκτες του ιού στο αίμα. Από τα άτομα με οξεία ηπατίτιδα Β, ποσοστό 5-15% μεταπίπτει σε χρονιότητα, εφόσον έχουν μολυνθεί ως ενήλικες. Μπορεί μάλιστα να έχουν περάσει στη φάση της χρόνιας λοίμωξης χωρίς να το γνωρίζουν καν, σε περίπτωση που η νόσος δεν έχει δώσει συμπτώματα κατά τη φάση της οξείας λοίμωξης ή αν τα συμπτώματα δεν ήταν τέτοια ώστε να αναζητήσουν ιατρική βοήθεια. Μερικοί άνθρωποι που πάσχουν από χρόνια λοίμωξη ηπατίτιδας Β, είναι γνωστοί ως «ανενεργοί» φορείς. Αυτό σημαίνει ότι έχουν τον ιό της ηπατίτιδας Β στο αίμα τους και μπορούν να τον μεταδώσουν σε άλλους, αλλά ο ιός δεν προκαλεί εξελισσόμενη ηπατική βλάβη, τα ιικά επίπεδα είναι μονίμως χαμηλά και δεν απαιτείται καμία φαρμακευτική αγωγή. Κάποιες φορές ο ιός μπορεί να επανενεργοποιηθεί. Συνεπώς, οι ανενεργοί φορείς είναι σημαντικό να παρακολουθούνται τακτικά με σκοπό τη φαρμακευτική αντιμετώπιση της ηπατίτιδας Β, καθώς αν η λοίμωξη γίνει χρόνια ενεργός εξακολουθεί να υπάρχει αυξημένος κίνδυνος κίρρωσης και καρκίνου του ήπατος. Ορισμένοι ανενεργοί φορείς αποβάλλουν φυσικά τον ιό από το σώμα τους, αλλά για αυτό ενδέχεται να χρειαστούν πολλά χρόνια.

Η μόλυνση από τον ιό της ηπατίτιδας Β είναι ένας παράγοντας που θεωρείται υπεύθυνος για την ανάπτυξη ηπατοκυτταρικού καρκίνου. Η νοσηρότητα και η θνητότητα στους ασθενείς με χρόνια ηπατίτιδα Β σχετίζεται με την εξέλιξη σε κίρρωση ή ηπατοκυτταρικό καρκίνωμα (ΗΚΚ), αφού η ετήσια επίπτωση ΗΚΚ σε ασθενείς με χρόνια ηπατίτιδα Β είναι 2-5% ιδιαίτερα όταν συνυπάρχει κίρρωση ήπατος και ο συνολικός κίνδυνος είναι 100 φορές μεγαλύτερος σε σχέση με άτομα που δεν έχουν χρόνια ηπατίτιδα Β (Ε.Σιβρίδης, 2007).

Ένα προγνωστικό μοντέλο είναι αυτό που θα εκτιμήσει τον κίνδυνο εμφάνισης ηπατοκυτταρικού καρκίνου σε ασθενείς με χρόνια ηπατίτιδα Β λαμβάνοντας υπόψη την πληροφορία θανάτου. Ο θάνατος είναι ένα ανταγωνιστικό γεγονός αφού η εμφάνιση του αποτρέπει την εμφάνιση της έκβασης που μελετάμε, δηλαδή του ηπατοκυτταρικού καρκίνου. Η μελέτη της ηπατίτιδας Β και της ανάπτυξης ΗΚΚ είναι ένα υπαρκτό θέμα και αυτό αποδεικνύεται από το γεγονός ότι έχουν γίνει πολλές

μελέτες για το θέμα αυτό και έχουν ήδη αναπτυχθεί μοντέλα πρόβλεψης και κλίμακες εκτίμησης του κινδύνου εμφάνισης ΗΚΚ σε ασθενείς με χρόνια ηπατίτιδα Β. Οι Paratheodoridis et al, ανέπτυξαν την κλίμακα κινδύνου PAGE-B η οποία προβλέπει τον κίνδυνο ανάπτυξης ΗΚΚ σε καυκάσιους ασθενείς με χρόνια ηπατίτιδα Β πέντε έτη από την έναρξη αντικής θεραπείας, και βασίστηκε μόνο στην ηλικία, το φύλο και τα αιμοπετάλια των ασθενών (Paratheodoridis et al, 2016). Η αθροιστική πιθανότητα εμφάνισης ΗΚΚ στην 5ετία στα δεδομένα ασθενών βάσει των οποίων αναπτύχθηκε το μοντέλο πρόβλεψης ήταν 5,7% ενώ στα δεδομένα που χρησιμοποιήθηκαν για τη επικύρωση του μοντέλου ήταν 8.4%. Οι ασθενείς με PAGE-B ≤ 9 , 10-17, ≥ 18 είχαν αθροιστική επίπτωση ΗΚΚ στην πενταετία 0%, 3% και 17%. Η μελέτη αυτή ήταν η πρώτη που προσπάθησε να αναπτύξει μια κλίμακα κινδύνου για ΗΚΚ σε καυκάσιους ασθενείς με χρόνια ηπατίτιδα Β. Σημαντικό είναι ότι στην μελέτη αυτή υπήρξαν αρκετές διαφορές στα χαρακτηριστικά των ασθενών στο σύνολο δεδομένων της ανάπτυξης του μοντέλου και στο σύνολο δεδομένων της επικύρωσης του μοντέλου γεγονός που ενισχύει την αξιοπιστία της κλίμακας κινδύνου αφού δείχνει ότι προσφέρει παρόμοια προβλεπτικότητα σε διαφορετικούς πληθυσμούς ασθενών.

Μια άλλη κλίμακα που αναπτύχθηκε με σκοπό την εκτίμηση κινδύνου εμφάνισης ηπατοκυτταρικού καρκινώματος σε ασθενείς με χρόνια ηπατίτιδα Β είναι η REACH B (Yang et al, 2011). Στην μελέτη αυτή το δείγμα που χρησιμοποιήθηκε στην ανάπτυξη του μοντέλου περιλάμβανε 3584 ασθενείς χωρίς κίρρωση από την μελέτη της Ταϊβάν REVEAL-HBV (εκ των οποίων 131 εμφάνισαν ΗΚΚ κατά τη διάρκεια της παρακολούθησης) και μία ομάδα 1505 ασθενών για επικύρωση από τρία νοσοκομεία στο Χονγκ Κονγκ και τη Νότια Κορέα. Χρησιμοποιήθηκε πολυπαραγοντικό μοντέλο αναλογικών κινδύνων Cox για την πρόβλεψη κινδύνου ΗΚΚ σε 3, 5 και 10 χρόνια. Αναπτύχθηκε μια κλίμακα 17 σημείων με τον κίνδυνο για ΗΚΚ να κυμαίνεται από 0, 0% έως 23,6% στα 3 χρόνια, 0,0% έως 47,4% στα 5 έτη και 0,01% έως 81,6% σε 10 χρόνια για τους ασθενείς με τον χαμηλότερο και υψηλότερο κίνδυνο ΗΚΚ, αντίστοιχα. Οι κλινικοί γιατροί μπορούν να χρησιμοποιήσουν αυτή την κλίμακα για να αξιολογήσουν τον κίνδυνο ΗΚΚ σε ασθενείς με χρόνια ηπατίτιδα Β και στη συνέχεια να λάβουν αποφάσεις σχετικά με την κλινική τους διαχείριση.

Άλλες κλίμακες που έχουν αναπτυχθεί για το θέμα αυτό είναι οι CU-HCC και GAG-HCC. Στην πρώτη κατασκευάστηκε μια απλή κλίμακα πρόβλεψης από συνήθεις κλινικές και εργαστηριακές παραμέτρους για την πρόβλεψη της ανάπτυξης ΗΚΚ σε ασθενείς με χρόνια ηπατίτιδα Β. Χρησιμοποιήθηκαν 5 μεταβλητές (ηλικία, λευκωματίνη, χολερυθρίνη, HBV DNA και κίρρωση) για να κατασκευαστεί η κλίμακα πρόγνωσης που κυμαίνεται από 0 έως 44.5. Το δείγμα βάσει του οποίου αναπτύχθηκε η κλίμακα αποτελείται από 1005 ασθενείς και το δείγμα επικύρωσης από 424. Κατά τη διάρκεια παρακολούθησης 10 ετών 105 (10%) ασθενείς από την κοορτή ανάπτυξης του μοντέλου και 45 (10.6%) από την κοορτή επικύρωσης εμφάνισαν ΗΚΚ (*Wong et al, 2010*). Στην δεύτερη (GAG-HCC), 820 ασθενείς με χρόνια ηπατίτιδα Β παρακολούθηθηκαν (μέση διάρκεια 76,8 μηνών) για την εμφάνιση ηπατοκυτταρικού καρκίνου. Αναπτύχθηκε μια κλίμακα εκτίμησης 5ετή και 10ετή κινδύνου εμφάνισης ΗΚΚ, σε άτομα με χρόνια ηπατίτιδα Β, η οποία επικυρώθηκε με ευαισθησία >84% και ειδικότητα >76%. Η κλίμακα αυτή μπορεί να εκτιμήσει την πιθανότητα εμφάνισης ΗΚΚ σε 5 και 10 χρόνια μετά την εμφάνιση ηπατίτιδας Β και μπορεί να χρησιμοποιηθεί για την ταυτοποίηση ασθενών με χρόνια ηπατίτιδα Β υψηλού κινδύνου για τη θεραπεία και τη διάγνωση του ΗΚΚ (*Yuen et al, 2009*).

1.4. Σκοπός της εργασίας

Σκοπός της παρούσας εργασίας είναι να πραγματοποιηθεί ανασκόπηση της μεθοδολογίας για την ανάπτυξη κλίμακας εκτίμησης κινδύνου (points-based risk score) υπό την παρουσία ανταγωνιστικών κινδύνων, καθώς επίσης και η εφαρμογή σε διαθέσιμα δεδομένα σε ασθενείς με χρόνια ηπατίτιδα Β για την εκτίμηση του κινδύνου εμφάνισης ηπατοκυτταρικού καρκίνου λαμβάνοντας υπόψη την πληροφορία θανάτου.

Κεφάλαιο 2

Ανάλυση Επιβίωσης και Ανταγωνιστικοί Κίνδυνοι

2.1. Δεδομένα Επιβίωσης

Η ανάλυση επιβίωσης επικεντρώνεται στον χρόνο από μια καθορισμένη έναρξη μέχρι να συμβεί το υπό μελέτη γεγονός (time to event data). Δηλαδή, αναφέρεται σε ένα σύνολο μεθόδων για την ανάλυση της χρονικής διάρκειας μέχρι την εμφάνιση ενός καλά καθορισμένου τελικού σημείου ενδιαφέροντος (*Schober et al, 2018*). Η ανάλυση επιβίωσης είναι η μεθοδολογία για την ανάλυση της αναμενόμενης διάρκειας του χρόνου μέχρι να συμβούν ένα ή περισσότερα γεγονότα, όπως ο θάνατος σε βιολογικούς οργανισμούς και η αποτυχία στα μηχανικά συστήματα. Γενικά έχει σκοπό να μοντελοποιήσει τον χρόνο και επιχειρεί να απαντήσει σε ερωτήματα όπως για παράδειγμα: ποιο είναι το ποσοστό ενός πληθυσμού που θα επιβιώσει μετά από ένα συγκεκριμένο χρονικό διάστημα;

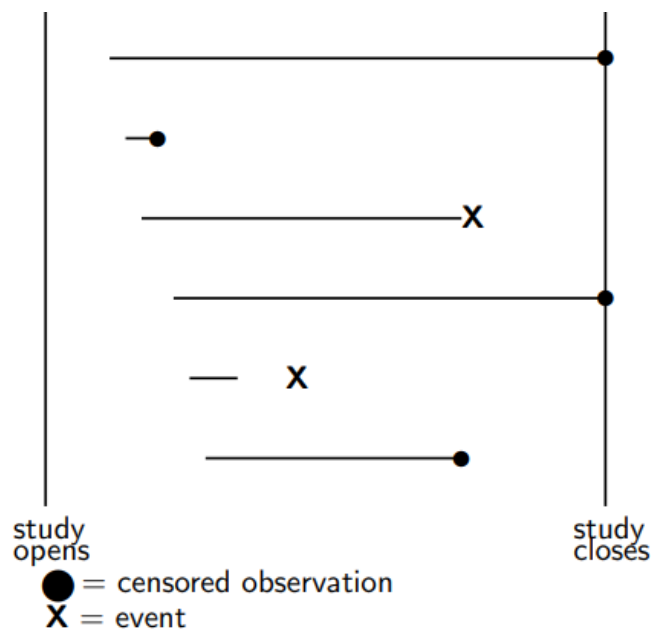
Η επιβίωση αποτελείται από θετικές τυχαίες μεταβλητές όπως, ο χρόνος μέχρι τον θάνατο, την έναρξη μιας ασθένειας (onset) ή την υποτροπή μιας νόσου (relapse) κλπ. Ο χρόνος επιβίωσης χρήζει ιδιαίτερης μεταχείρισης για τον λόγο ότι περιορίζεται στο να είναι πάντα θετικός καθώς και στο ότι τα δεδομένα περιέχουν αποκομμένες (censored) παρατηρήσεις. Με T συμβολίζουμε τον χρόνο επιβίωσης και παίρνει τιμές από 0 έως ∞ , όπου το 0 είναι η έναρξη (onset). Για να ορίσουμε μια τυχαία μεταβλητή χρόνου αποτυχίας χρειαζόμαστε τρία βασικά πράγματα (*D. Collett, 2003*):

- 1) αδιαμφισβήτητη έναρξη, δηλαδή πότε ξεκινάμε να μετράμε τον χρόνο (time origin)
- 2) ίδια κλίμακα χρόνου (time scale)
- 3) καλός ορισμός του γεγονότος (event)

Ένα ξεχωριστό χαρακτηριστικό των δεδομένων επιβίωσης είναι η έννοια της αποκοπής (censoring). Η αποκοπή είναι μια μορφή έλλειψης πληροφορίας (missing data) και συμβαίνει όταν για παράδειγμα το γεγονός ενδιαφέροντος είναι ο θάνατος και κάποια άτομα είναι ζωντανά μέχρι το τέλος της μελέτης. Συμβολίζεται με U και

το μόνο που ξέρουμε για τα αποκομμένα άτομα είναι ότι μέχρι το τέλος της μελέτης δεν είχαν εμφανίσει το γεγονός ενδιαφέροντος. Εμείς παρατηρούμε $X = \min(T, U)$, δηλαδή έχουμε δύο διαδικασίες (failure T και censoring U) και καταγράφουμε όποιον από τους δύο χρόνους έρθει πρώτος. Εκτός από την παρατήρηση X μπορούμε επίσης να δούμε τον δείκτη αποτυχίας (failure indicator) δ_i : $\delta_i = 1$ αν $T_i \leq U_i$ (αποτυχία) ή $\delta_i = 0$ αν $T_i > U_i$ (αποκοπή).

Σχήμα 2.1: Απεικόνιση δεδομένων επιβίωσης με αποκομμένες παρατηρήσεις



Η παραπάνω απεικόνιση παρουσιάζει αρκετά χαρακτηριστικά των δεδομένων επιβίωσης, όπως το ότι τα άτομα δεν συμμετέχουν στην μελέτη ταυτόχρονα αφού επιτρέπεται να μπαίνουν στην μελέτη σταδιακά (staggered entry). Επίσης απεικονίζεται πολύ καλά η έννοια της αποκοπής μιας και βλέπουμε ότι όταν τελειώνει η μελέτη μερικά άτομα δεν έχουν πάθει ακόμα το γεγονός και άλλα άτομα αποχωρούν ή χάνονται στη μέση της μελέτης και το μόνο που ξέρουμε για αυτά είναι ότι μέχρι την στιγμή που χάθηκαν δεν είχαν πάθει το γεγονός.

Υπάρχουν διάφοροι τύποι αποκοπής (Darity et al, 2008). Η δεξιά αποκοπή (right censoring) παρατηρείται όταν το άτομο χάνεται ή αποχωρεί από την μελέτη πριν εμφανιστεί το γεγονός ενδιαφέροντος, δηλαδή το αληθινό γεγονός βρίσκεται δεξιά από την αποκομμένη παρατήρηση. Εμείς παρατηρούμε $X_i = \min(U_i, T_i)$. Στην περίπτωση της δεξιάς αποκοπής όταν συμβεί το γεγονός που μελετάμε διακόπτεται η

διαδικασία (δηλαδή είναι καταληκτικό). Η δεξιά αποκοπή παρατηρείται όταν δεν παρακολουθείται πλέον το άτομο λόγω εξωτερικών παραγόντων (loss to follow up), όταν το άτομο φεύγει οικειοθελώς από τη μελέτη (drop out) ή όταν απλά τερματίζεται η μελέτη. Ένα παράδειγμα δεξιάς αποκοπής είναι όταν διεξάγεται μια μελέτη σχετικά με τη διάρκεια της εγκυμοσύνης αλλά στο τέλος της μελέτης κάποιες γυναίκες εξακολουθούν να είναι έγκυες και επομένως δεν γνωρίζουμε ακριβώς πόσο θα διαρκέσει η εγκυμοσύνη τους. Οι παρατηρήσεις αυτές θα ήταν δεξιά αποκομμένες επειδή η «αποτυχία» ή η γέννηση σε αυτή την περίπτωση, θα συμβεί μετά τον χρόνο που καταγράφεται.

Ένας άλλος τύπος είναι η αριστερή αποκοπή (left censoring). Στην περίπτωση αυτή το αληθινό γεγονός βρίσκεται αριστερά από την αποκομμένη παρατήρηση, δηλαδή ξεκινάμε να παρακολουθήσουμε το άτομο αφού έχει αναπτύξει τη νόσο. Η στιγμή που παρατήρησα το γεγονός έχει συμβεί πριν, το γεγονός συμβαίνει αλλά ταυτόχρονα δεν διακόπτει τη διαδικασία (όχι θάνατος). Εδώ παρατηρούμε $Y_i = \max(T_i, U_i)$, δηλαδή η αποτυχία συμβαίνει πριν την αποκοπή (Lawless, 1982). Για παράδειγμα όταν ένα μόνιμο δόντι έχει ήδη εμφανιστεί πριν από την έναρξη μιας οδοντιατρικής μελέτης που στοχεύει στην εκτίμηση της κατανομής της εμφάνισης. Επίσης μπορεί να υπάρξει και αποκοπή κατά διαστήματα (interval censoring) όπου το γεγονός παρατηρείται κάπου ανάμεσα σε ένα χρονικό διάστημα. Το γεγονός που μας ενδιαφέρει συμβαίνει μεταξύ του L_i και του R_i : (L_i, R_i) όπου $T_i \in (L_i, R_i)$. Το αποκομμένο διάστημα συμβαίνει συχνά στις μελέτες HIV / AIDS. Πράγματι, ο χρόνος για την ορομετατροπή του HIV μπορεί να προσδιοριστεί μόνο με εργαστηριακή αξιολόγηση η οποία συνήθως ξεκινά μετά από μια επίσκεψη στον ιατρό. Στη συνέχεια, μπορεί κανείς να συμπεράνει ότι η ορομετατροπή του HIV έχει συμβεί μεταξύ δύο εξετάσεων. Το ίδιο ισχύει για τη διάγνωση του AIDS, η οποία βασίζεται σε κλινικά συμπτώματα και πρέπει να επιβεβαιωθεί με ιατρική εξέταση.

2.2. Βασικές έννοιες στην ανάλυση επιβίωσης

Συμβολίζοντας το χρόνο επιβίωσης με T , η συνάρτηση επιβίωσης (survival function) $S(t)$ ορίζεται ως η πιθανότητα επιβίωσης ενός ατόμου πέραν τη χρονική στιγμή t και δίνεται από τη σχέση :

$$S(t) = P(T > t) = 1 - F(t)$$

όπου $F(t)$ η αθροιστική συνάρτηση πιθανότητας. Η συνάρτηση επιβίωσης είναι μη αρνητική και μη αύξουσα συνάρτηση του t με $S(0) = 1$ και $S(\infty) = 0$. Η γραφική παράσταση της $S(t)$ συναρτήσει του t είναι γνωστή ως καμπύλη επιβίωσης και είναι πολύ σημαντική στην ανάλυση δεδομένων χρόνου επιβίωσης (*Miller et al, 1997*).

Η συνάρτηση κινδύνου, $\lambda(t)$, ορίζεται ως η πιθανότητα αποβίωσης (ή πραγμάτωσης του γεγονότος που εξετάζεται) τη χρονική στιγμή t , δεδομένου ότι το άτομο έχει επιβιώσει μέχρι τη χρονική στιγμή t . Δηλαδή,

$$\lambda(t) = \lim_{s \rightarrow 0} \frac{P(t \leq T < t+s | T \geq t)}{s}$$

Η συνάρτηση κινδύνου δίνει ένα μέτρο του πόσο πιθανό είναι ένα άτομο να αποβιώσει ως συνάρτηση της ηλικίας του, για παράδειγμα ο κίνδυνος θανάτου ανάμεσα σε αυτούς που είναι ζωντανοί τη συγκεκριμένη χρονική στιγμή. Ο αθροιστικός κίνδυνος ορίζεται ως

$$A(t) = \int_0^t \lambda(u) du \quad \text{ή} \quad A(t) = \sum_{ak < t} \lambda_k$$

για συνεχή και διακριτή περίπτωση αντίστοιχα. Ο αθροιστικός κίνδυνος δεν έχει πολύ διαισθητική ερμηνεία αλλά είναι χρήσιμος για ορισμένες γραφικές εκτιμήσεις.

Ενδιαφέρον επίσης παρουσιάζει η σχέση μεταξύ του $S(t)$ και του $\Lambda(t)$. Ο Cox (*Cox, 1984*) ορίζει: $S(t) = e^{-\Lambda(t)}$. Στις εφαρμογές παρατηρούνται διάφορες μορφές κινδύνου όπως:

- Αυξανόμενη καμπύλη (increasing): πχ γήρανση μετά τα 65
- Φθίνουσα καμπύλη (decreasing): πχ επιβίωση μετά από χειρουργική επέμβαση
- Σχήμα μπανιέρας (bathtub): πχ θνησιμότητα λόγω ηλικίας

- **2.3. Μη παραμετρικές μέθοδοι για την εκτίμηση των συναρτήσεων επιβίωσης**

Η χρήση αρχικά, μη παραμετρικών μεθόδων για ανάλυση δεδομένων γίνεται λόγω του ότι είναι πιο εύκολες τόσο στην εφαρμογή τους όσο και στην κατανόηση τους. Επιπλέον είναι περισσότερο αποδοτικές όταν δεν υπάρχουν γνωστές θεωρητικές κατανομές. Η παραμετρική στατιστική βρίσκει τρόπους να εκτιμήσει κάποιες ποσότητες (για παράδειγμα τη συνάρτηση επιβίωσης $S(t)$) χωρίς τη χρήση καμιάς κατανομής. Οι μη παραμετρικές στατιστικές μέθοδοι περιλαμβάνουν τόσο περιγραφικά στατιστικά στοιχεία όσο και στατιστική συμπερασματολογία (*Stuart et al, 1999*). Πιο κάτω θα αναπτύξουμε μη παραμετρικούς τρόπους εκτίμησης των συναρτήσεων επιβίωσης.

2.3.1. Εκτιμητήρια Kaplan-Meier

Η εκτιμητήρια Kaplan-Meier είναι ίσως η πιο δημοφιλής προσέγγιση εκτίμησης της συνάρτησης επιβίωσης. Αυτό οφείλεται στο ότι εκτιμά την επιβίωση με την πάροδο του χρόνου με πολύ απλό τρόπο (*Manish et al, 2010*). Στην περίπτωση όπου δεν υπάρχουν αποκομμένες παρατηρήσεις ο γενικός τύπος είναι:

$$\widehat{S}(t) = \frac{\text{αριθμός ατόμων με } T \geq t}{\text{συνολικό δείγμα}}$$

όπου t μια δεδομένη χρονική στιγμή και T ο χρόνος επιβίωσης.

Αυτό όμως δεν συμβαίνει όταν έχουμε αποκομμένα δεδομένα, διότι δεν ξέρουμε τι έχει συμβεί την συγκεκριμένη χρονική στιγμή που συμβαίνει η αποκοπή ώστε να μπορούμε να υπολογίσουμε την επιβίωση την επόμενη χρονική στιγμή. Για τον λόγο αυτό, οι Kaplan-Meier χρησιμοποίησαν δεσμευμένες πιθανότητες ώστε να μπορούμε να εκτιμήσουμε μη παραμετρικά το $S(t)$ ακόμη και όταν έχουμε αποκομμένες παρατηρήσεις. Στην περίπτωση αυτή, η καμπύλη επιβίωσης Kaplan-Meier ορίζεται ως η πιθανότητα επιβίωσης σε δεδομένο χρονικό διάστημα, ενώ εξετάζεται ο χρόνος σε πολλά μικρά διαστήματα (*Chapman and Hall, 1992*). Στην ανάλυση αυτή υπάρχουν τρεις παραδοχές:

- 1) Τα άτομα που «χάθηκαν» από την παρακολούθηση, έχουν την ίδια πιθανότητα επιβίωσης με τα άτομα που συνεχίζουν στην παρακολούθηση. Αυτό όμως δε μπορεί να ελεγχθεί και μπορεί να οδηγήσει σε μεροληψία που μειώνει τον εκτιμητή $S(t)$ της συνάρτησης επιβίωσης
- 2) Οι πιθανότητες επιβίωσης είναι οι ίδιες για άτομα που εισήλθαν στην αρχή της μελέτης με των ατόμων που εισήλθαν πιο αργά στη μελέτη. Αυτό μπορεί να ελεγχθεί.
- 3) Το γεγονός που μελετάται (π.χ. θάνατος) συμβαίνει στον καθορισμένο χρόνο. Καθυστερημένη καταγραφή του γεγονότος, θα προκαλέσει αύξηση της $S(t)$.

Η διαδικασία για την εκτιμήτρια Kaplan-Meier περιλαμβάνει τη διαίρεση του παρατηρούμενου χρονικού διαστήματος της μελέτης σε μια σειρά από λεπτά διαστήματα ώστε να υπάρχει ξεχωριστό διάστημα για κάθε χρόνο θανάτου ή αποκοπής. Για τον υπολογισμό της εκτιμήτριας K-M χρησιμοποιείται η δεσμευμένη πιθανότητα,

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

και ο πολλαπλασιαστικός νόμος της πιθανότητας,

$$P(A \cap B) = P(B)P(A/B)$$

Έστω ότι έχουμε A_1, A_2, \dots, A_k διαφορετικά γεγονότα, τότε

$$P(A_1 \cap A_2 \cap \dots \cap A_k) = P(A_k | A_{k-1} \cap \dots \cap A_1) \times \dots \times P(A_2 | A_1) \times P(A_1)$$

Εφαρμόζοντας το παραπάνω μπορούμε να εκτιμήσουμε το $\widehat{S}(t)$:

έστω $a_k < t < a_{k+1}$, τότε $S(t) = P(T \geq a_{k+1}) = P(T \geq a_1, T \geq a_2, \dots, T \geq a_{k+1})$

$$= P(T \geq a_1) \times \prod_{j=1}^k P(T \geq a_{j+1} | T \geq a_j) = \prod_{j=1}^k [1 - P(T = a_j | T \geq a_j)]$$

$$= \prod_{j=1}^k [1 - \lambda_j].$$

Άρα, τελικά η πιθανότητα να επιβιώσει κάποιος μέχρι κάποια χρονική στιγμή είναι το γινόμενο όλων των προηγούμενων στιγμών που είχε επιβιώσει, δηλαδή,

$$\widehat{S}(t) = \prod_{j=1}^k \frac{n_j - d_j}{n_j} = \prod_{j=1}^k (1 - \frac{d_j}{r_j})$$

Όπου, n_j : ο αριθμός των ατόμων που είναι σε κίνδυνο στο χρόνο t_j

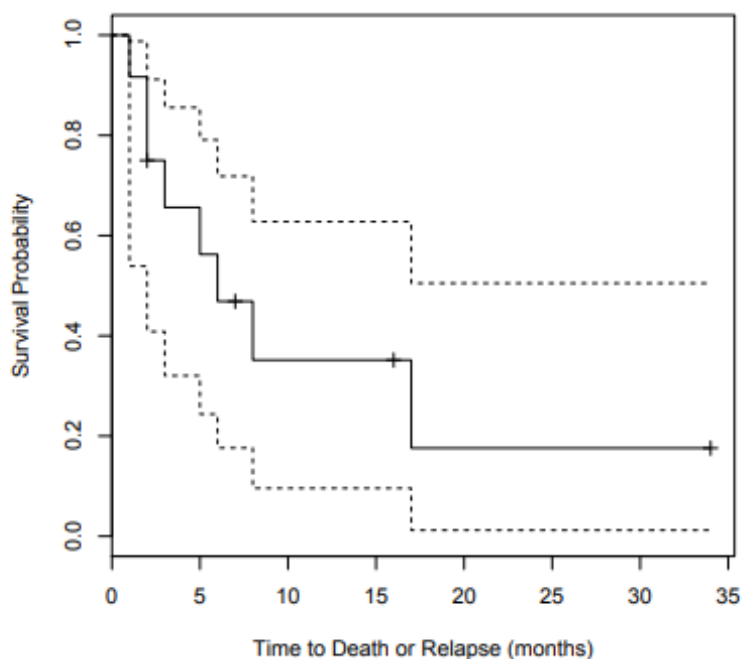
d_j : ο αριθμός των θανάτων στον χρόνο t_j

c_j : ο αριθμός των αποκομμένων παρατηρήσεων στον χρόνο t_j .

Στην πράξη, ο εκτιμητής Kaplan-Meier, μπορεί να εκτιμηθεί κατασκευάζοντας έναν πίνακα, ο οποίος θα έχει 5 στήλες:

1. Η πρώτη στήλη θα περιέχει τους διακεκριμένους πλήρεις χρόνους επιβίωσης, αποκομμένους και μη, σε αύξουσα σειρά.
2. Η δεύτερη στήλη θα έχει τον αριθμό των ατόμων που βρίσκονται σε κίνδυνο (r_j).
3. Η τρίτη στήλη θα έχει τον αριθμό των θανάτων (d_j) στους διάφορους χρόνους επιβίωσης (αποτυχίας).
4. Στην τέταρτη στήλη θα είναι οι τιμές των $p_j = 1 - \frac{d_j}{n_j}$.
5. Η τελευταία στήλη θα είναι οι εκτιμητές Kaplan - Meier.

Από την γραφική παράσταση $\widehat{S}(t)$ συναρτήσεως του χρόνου t παίρνουμε τον εκτιμητή Kaplan-Meier της καμπύλης επιβίωσης. Η καμπύλη αυτή, όπως φαίνεται και στο γράφημα 2.2, είναι μια σκαλωτή φθίνουσα συνάρτηση, η οποία στον άξονα ψ πάντα ξεκινάει από την τιμή 1 αφού στην αρχή της μελέτης όλοι είναι ζωντανοί (ή γενικότερα χωρίς το υπό μελέτη συμβάν). Η καμπύλη αυτή πέφτει κάθε φορά που έχουμε ένα γεγονός (event) και σιγά σιγά πλησιάζει το μηδέν στο τέλος της μελέτης.



Γράφημα 2.2. πιθανότητα επιβίωσης συναρτήσει του χρόνου μέχρι τον θάνατο ή την υποτροπή

Οι διακεκομμένες γραμμές αντιπροσωπεύουν τα 95% διαστήματα εμπιστοσύνης ενώ τα σημεία που σχηματίζουν σταυρό αντιπροσωπεύουν τις αποκομμένες παρατηρήσεις. Παρατηρούμε ότι εάν το τελευταίο γεγονός είναι αποκοπή θα έχουμε μια ευθεία γραμμή.

Γενικά η κατανομή του χρόνου επιβίωσης είναι λοξή προς τα δεξιά. Ο μέσος τραβιέται από αυτούς που είναι μακριά (από τις ακραίες τιμές), γι αυτό στην ανάλυση επιβίωσης χρησιμοποιείται ο διάμεσος χρόνος επιβίωσης (median survival time), ο οποίος δεν επηρεάζεται από τις ακραίες τιμές. Ο διάμεσος χρόνος επιβίωσης είναι ο χρόνος επιβίωσης, στον οποίο το 50% των υπό μελέτη ατόμων επιβιώνει $S(t)=0.5$. Η εκτίμηση του διάμεσου χρόνου επιβίωσης μπορεί να παρθεί μέσω του εκτιμητή Kaplan-Meier ως:

$$\widehat{S}(t) \leq 0.5.$$

Αφού πάρουμε τον εκτιμητή K-M μπορούμε να υπολογίσουμε και τα 95% διαστήματα εμπιστοσύνης. Τα διαστήματα εμπιστοσύνης που υπολογίζονται στους πλήρεις χρόνους, δίνουν μια καλή ένδειξη για την αξιοπιστία των εκτιμητών σε αυτούς τους χρόνους. Τα διαστήματα εμπιστοσύνης που θα αναπτυχθούν στην

παράγραφο αυτή είναι σημειακά διαστήματα εμπιστοσύνης. Ένα $(1-\alpha)100\%$ διάστημα εμπιστοσύνης στο χρόνο t για την ποσότητα $S(t)$, θεωρώντας ότι οι εκτιμητές $\widehat{S}(t)$ Kaplan – Meier ακολουθούν κανονική κατανομή, είναι

$$\widehat{S}(t) \pm Z_{1-\frac{\alpha}{2}} \text{se}[\widehat{S}(t)]$$

Υπάρχουν διάφοροι τρόποι για να εκτιμήσουμε το τυπικό σφάλμα $\text{se}[\widehat{S}(t)]$. Εδώ θα δούμε την μέθοδο Greenwood (*Pokhrel et al, 2008*). Κατά τη μέθοδο αυτή η εκτίμηση του τυπικού σφάλματος δίνεται από:

$$\text{se}[\widehat{S}(t)] = [\widehat{S}(t)] \left\{ \sum_{j=1}^k \frac{d_j}{n_j(n_j - d_j)} \right\}^{1/2}$$

όπου n_j ο αριθμός των ατόμων που βρίσκονται σε κίνδυνο την χρονική στιγμή t_j και d_j ο αριθμός των θανάτων κατά τη χρονική στιγμή t_j .

2.3.2. Πίνακες Επιβίωσης-Life tables

Η μέθοδος αυτή είναι μία από τις παλαιότερες τεχνικές για τη μέτρηση της θνησιμότητας και περιγράφει την εμπειρία της επιβίωσης ενός πληθυσμού. Είναι μια “επέκταση” των συνηθισμένων πινάκων συχνοτήτων (frequency tables) στην περίπτωση που υπάρχουν αποκομμένα δεδομένα (*E. T. Lee, 2003*). Εφαρμόζεται όταν έχουμε ομαδοποιημένα δεδομένα (grouped data) και η διαφορά με τον εκτιμητή Kaplan-Meier είναι ότι εδώ δεν έχουμε διαστήματα από γεγονός σε γεγονός, αλλά έχουμε χρονικά διαστήματα, πχ ανά μήνα. Δηλαδή, η μέθοδος Kaplan – Meier μπορεί να θεωρηθεί μια ειδική περίπτωση της μεθόδου πινάκων επιβίωσης, όπου κάθε διάστημα περιέχει μόνο μία παρατήρηση.

Υπάρχουν δύο ειδών πίνακες επιβίωσης, οι πίνακες επιβίωσης πληθυσμού (population life table) και οι κλινικοί πίνακες επιβίωσης-αναλογιστική μέθοδος (clinical life tables-the actuarial method). Στην περίπτωση των πινάκων επιβίωσης πληθυσμού έχουμε δύο κατηγορίες πινάκων: ο πίνακας της ζωής της ομάδας (cohort life table) και ο τρέχων πινάκων επιβίωσης (current life table). Ο πίνακας επιβίωσης της ομάδας περιγράφει την επιβίωση ή την εμπειρία θνησιμότητας από τη γέννηση

έως το θάνατο μιας συγκεκριμένης ομάδας των ατόμων που είχαν γεννηθεί περίπου την ίδια ώρα. Για παράδειγμα, ο πίνακας επιβίωσης όλων των ατόμων που έχουν γεννηθεί το 1960. Η ομάδα πρέπει να ακολουθείται από το 1960 έως ότου όλα αυτά τα άτομα πεθάνουν. Το ποσοστό θανάτου (ή επιβίωσης) χρησιμοποιείται στη συνέχεια για να κατασκευαστούν πίνακες επιβίωσης για διαδοχικά ημερολογιακά έτη. Αυτός ο τύπος πίνακα, είναι χρήσιμος στην προβολή του πληθυσμού και σε προοπτικές μελέτες αλλά δεν κατασκευάζεται συχνά επειδή απαιτεί μακρά περίοδο παρακολούθησης. Ο τρέχων πίνακας επιβίωσης κατασκευάζεται με την εφαρμογή ενός λόγου θνησιμότητας του πληθυσμού μιας συγκεκριμένης ηλικίας, σε μια δεδομένη χρονική περίοδο σε μια υποθετική ομάδα των 100.000 ή 1.000.000 ατόμων. Το σημείο εκκίνησης θεωρείται η γέννηση κατά το έτος 0. Για την κατασκευή ενός πίνακα επιβίωσης πληθυσμού απαιτούνται δύο πηγές δεδομένων. Η πρώτη είναι δεδομένα απογραφής όσον αφορά τον αριθμό των ατόμων κάθε ηλικίας που ζουν για ένα δεδομένο έτος, στα τέλη του πρώτου εξαμήνου, και η δεύτερη, δημογραφικές στατιστικές για τον αριθμό των θανάτων κατά το συγκεκριμένο έτος για κάθε ηλικία.

Γενικά, οι πίνακες επιβίωσης κοορτής βασίζονται στις εξής οκτώ παραδοχές:

- 1.** Μια υποθετική κοορτή του πίνακα επιβίωσης αποτελείται συνήθως από 1.000 ή 10.000 ή 1.00.000 γεννήσεις.
- 2.** Οι θάνατοι κατανέμονται εξίσου καθ 'όλη τη διάρκεια του έτους.
- 3.** Η ομάδα των ανθρώπων μειώνεται σταδιακά μόνο από το θάνατο.
- 4.** Η κοορτή είναι κλειστή
- 5.** Το ποσοστό θνησιμότητας σχετίζεται με ένα προκαθορισμένο ποσοστό θνησιμότητας ανά ηλικία.
- 6.** Η ομάδα των ατόμων πεθαίνει σε μια καθορισμένη ηλικία που δεν αλλάζει.
- 7.** Δεν υπάρχει μεταβολή των ποσοστών θανάτων με την πάροδο του χρόνου.
- 8.** Οι πίνακες επιβίωσης γενικά κατασκευάζεται ξεχωριστά για άνδρες και γυναίκες

Το δεύτερο είδος πίνακα επιβίωσης είναι οι κλινικοί πίνακες-αναλογιστική μέθοδος. Για να πραγματοποιηθεί η μέθοδος αυτή απαιτείται ένας μεγάλος αριθμός παρατηρήσεων για να μπορούν να ομαδοποιηθούν οι συναρτήσεις επιβίωσης σε χρονικά διαστήματα. Όπως και ο εκτιμητής Kaplan-Meier, έτσι και η αναλογιστική μέθοδος ενσωματώνει όλες τις πληροφορίες επιβίωσης μέχρι το τέλος της μελέτης. Για παράδειγμα, στον υπολογισμό ενός πενταετούς ποσοστού επιβίωσης των

ασθενών με καρκίνο του μαστού, δεν απαιτείται ο περιορισμός μόνο σε αυτούς τους ασθενείς που έχουν εγγραφεί στη μελέτη για πέντε ή περισσότερα χρόνια. Οι ασθενείς που έχουν εισέλθει για τέσσερα, τρία, δύο, ακόμη και ένα χρόνο συμβάλλουν χρήσιμες πληροφορίες για την αξιολόγηση της πενταετούς επιβίωσης. Συνοπτικά η μέθοδος αυτή χρησιμοποιεί ελλείποντα στοιχεία όπως είναι οι απώλειες κατά την παρακολούθηση ή και πρόσωπα που αποσύρθηκαν ζωντανά από τη μελέτη χωρίς λόγο καθώς και πλήρη στοιχεία του θανάτου.

2.4. Σύγκριση καμπύλων επιβίωσης

Στην ανάλυση επιβίωσης, εκτός από την εκτίμηση των καμπύλων επιβίωσης, μας ενδιαφέρει η σύγκριση του χρόνου επιβίωσης σε δύο ή περισσότερες ομάδες ατόμων που διαφέρουν ως προς ένα χαρακτηριστικό ή που είναι τοποθετημένοι τυχαία σε διαφορετικές θεραπείες. Για παράδειγμα, εάν έχουμε ένα σύνολο ατόμων με την ίδια ασθένεια, μπορεί να ενδιαφερόμαστε να συγκρίνουμε την ικανότητα δύο ή περισσότερων θεραπειών να παρατείνουν τη ζωή των ασθενών ή να διατηρούν την ύφεση της ασθένειας. Στις περισσότερες περιπτώσεις οι χρόνοι επιβίωσης διαφορετικών ομάδων διαφέρουν. Με το σχεδιασμό των εκτιμώμενων συναρτήσεων επιβίωσης μπορούμε να έχουμε μια οπτική εικόνα για το αν υπάρχει διαφορά μεταξύ των συναρτήσεων επιβίωσης διαφορετικών ομάδων. Αυτό όμως μπορεί να μας δώσει μόνο μια γενική ιδέα για τους χρόνους επιβίωσης των ομάδων αφού δεν μπορούμε να συμπεράνουμε αν οι διαφορές αυτές είναι στατιστικά σημαντικές. Ως εκ τούτου, είναι απαραίτητος ένας στατιστικός έλεγχος. Επειδή οι χρόνοι επιβίωσης δεν κατανέμονται κανονικά, πρέπει να εφαρμοστούν μη-παραμετρικοί έλεγχοι που βασίζονται στην ταξινόμηση των χρόνων επιβίωσης. Υπάρχουν διάφορα μη παραμετρικά τεστ, όπως, Gehan's generalized Wilcoxon test, Cox—Mantel test, Logrank test, Peto and Peto's generalized Wilcoxon test κλπ. Στην παρούσα εργασία θα επικεντρωθούμε στο Logrank test.

Η δοκιμή log-rank είναι ουσιαστικά ένα τεστ υποθέσεων για τη σύγκριση των κατανομών επιβίωσης δύο δειγμάτων. Είναι κατάλληλο όταν η κατανομή των δεδομένων είναι λοξή προς τα δεξιά και επίσης όταν υπάρχουν αποκομμένες παρατηρήσεις. Χρησιμοποιείται ευρέως σε κλινικές δοκιμές για να διαπιστωθεί η

αποτελεσματικότητα μιας νέας θεραπείας συγκριτικά με μια ήδη υπάρχουσα. Η δοκιμή log-rank μερικές φορές ονομάζεται Mantel-Cox τεστ, από τους Nathan Mantel και David Cox, μιας και προτάθηκε για πρώτη φορά από αυτούς και ονομάστηκε ως δοκιμή log-rank από τους Richard και τον Julian Peto, γνωστοί ως Peto&Peto (*Peto R & Peto, J, 1972*).

Το τεστ αυτό συγκρίνει τις εκτιμήσεις των συναρτήσεων κινδύνου των δύο ομάδων σε κάθε παρατηρούμενο χρόνο που έχουμε ένα συμβάν. Κατασκευάζεται υπολογίζοντας τον παρατηρούμενο και τον αναμενόμενο αριθμό γεγονότων σε κάθε μία από τις ομάδες σε κάθε παρατηρούμενο χρόνο συμβάντος και στη συνέχεια τα προσθέτουμε για να αποκτήσουμε μια σύνοψη για όλα τα χρονικά σημεία που υπάρχει ένα συμβάν. Έτσι, αν θεωρήσουμε τα O_1 και O_2 ως τους παρατηρούμενους αριθμούς θανάτων, και E_1, E_2 ως τους αναμενόμενους αριθμούς θανάτων στις ομάδες θεραπείας A και B αντίστοιχα, το στατιστικό log-rank δίνεται από τον τύπο:

$$X^2_{logrank} = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}$$

το οποίο ακολουθεί την X^2 κατανομή με 1 βαθμό ελευθερίας. Γενικά, για κ ομάδες θεραπείας, το $X^2_{logrank}$ συγκρίνεται με την X^2 κατανομή με κ-1 βαθμούς ελευθερίας. Ο έλεγχος υποθέσεων που πραγματοποιείται είναι ο εξής:

$$H_0: S_A(t) = S_B(t)$$

$$H_1: S_A(t) \neq S_B(t)$$

Με άλλα λόγια, η μηδενική υπόθεση μιας λέει ότι οι θεραπείες δεν διαφέρουν μεταξύ τους, αφού η επιβίωση των δύο ομάδων είναι ίδια. Απορρίπτουμε την H_0 , εάν η τιμή του $X^2_{logrank}$ είναι πολύ μεγάλη (p-value > 0.05) (*Harrington, 2005*). Σημαντικό είναι το γεγονός ότι, η δοκιμή logrank βασίζεται στις ίδιες υποθέσεις με την καμπύλη επιβίωσης Kaplan-Meier, δηλαδή ότι η αποκοπή δεν σχετίζεται με την πρόγνωση, οι πιθανότητες επιβίωσης είναι οι ίδιες για όλα τα άτομα ανεξαρτήτως του χρόνου εισαγωγής στη μελέτη και τα γεγονότα συμβαίνουν στους καθορισμένους χρόνους. Οι αποκλίσεις από αυτές τις υποθέσεις έχουν μεγαλύτερη σημασία εάν είναι διαφορετικές στις συγκρινόμενες ομάδες, για παράδειγμα εάν η αποκοπή είναι πιθανότερη σε μια ομάδα από την άλλη (*Bland, et al, 2004*).

2.5. Μοντέλο αναλογικών κινδύνων Cox - Proportional Hazards model

Πιο πάνω έγινε αναφορά σε μη παραμετρικό έλεγχο για την ισότητα των συναρτήσεων επιβίωσης διαφόρων ομάδων λαμβάνοντας υπόψη τις τιμές ενός άλλου χαρακτηριστικού που μπορεί να επηρεάζει το χρόνο ζωής στις ομάδες αυτές. Στην περίπτωση όμως που έχουμε περισσότερες από μία μεταβλητές, οι μέθοδοι γίνονται πιο περίπλοκες και η εφαρμογή τους είναι πιο δύσκολη. Το Logrank test, όπως και οι άλλοι έλεγχοι, δε μπορούν να προσαρμοστούν για τη διερεύνηση της επίδρασης διαφόρων μεταβλητών που είναι γνωστό ότι επηρεάζουν την επιβίωση. Η εύρεση της σχέσης μεταξύ μιας μεταβλητής που δηλώνει το χρόνο επιβίωσης ενός ατόμου και άλλων συμμεταβλητών, επιτυγχάνεται συνήθως μέσω ενός μοντέλου παλινδρόμησης. Όταν έχουμε αποκομμένα δεδομένα επιβίωσης, χρησιμοποιείται συνήθως το μοντέλο παλινδρόμησης του Cox (Cox regression model) ή διαφορετικά το μοντέλο αναλογικών κινδύνων του Cox (Cox proportional hazard model). Το μοντέλο Cox παρουσιάστηκε από τον Cox το 1972 (D. R. Cox, 1972) και χρησιμοποιείται για την διερεύνηση της συσχέτισης μεταξύ της επιβίωσης και των επεξηγηματικών μεταβλητών. Μας επιτρέπει να εκτιμήσουμε τον κίνδυνο θανάτου ενός ατόμου, ή άλλου γεγονότος που μας ενδιαφέρει δεδομένου των προγνωστικών τους μεταβλητών.

Έστω $Z = \{Z_1, Z_2, \dots, Z_p\}$ το σύνολο των ανεξάρτητων μεταβλητών (πχ φύλο, ηλικία κλπ). Τότε, η συνάρτηση κινδύνου για το μοντέλο αναλογικών κινδύνων δίνεται από τον τύπο:

$$\lambda(t; Z) = \lambda_0(t) \exp(\beta Z)$$

Ονομάζεται μοντέλων αναλογικών κινδύνων επειδή αν σκεφτούμε το $\lambda_1(t)$ ως τον κίνδυνο για την ομάδα θεραπείας ($Z=1$) και το $\lambda_0(t)$ ως τον κίνδυνο για την ομάδα ελέγχου ($Z=0$) (control group), αντικαθιστώντας στον παραπάνω τύπο έχουμε,

$$\lambda_1(t) = \lambda(t; Z=1) = \lambda_0(t) \exp(\beta)$$

Αυτό σημαίνει ότι ο λόγος κινδύνου (hazard ratio) είναι σταθερός, φ. Με άλλα λόγια, οι κίνδυνοι των ομάδων παραμένουν ανάλογοι με την πάροδο του χρόνου,

$$\varphi = \frac{\lambda_1(t)}{\lambda_0(t)} = \exp(\beta)$$

Το $\lambda_0(t)$ ονομάζεται βασική συνάρτηση κινδύνου (baseline hazard function) και αντικατοπτρίζει τον κίνδυνο για τα άτομα των οποίων όλες οι ανεξάρτητες μεταβλητές παίρνουν την τιμή μηδέν, $Z_i=0$, δηλαδή την ομάδα αναφοράς (Breslow, 1975).

Ένα από τα μεγαλύτερα πλεονεκτήματα του μοντέλου αναλογικών κινδύνων είναι ότι μπορούμε να εκτιμήσουμε τις παραμέτρους β που αντικατοπτρίζουν τις επιδράσεις της θεραπείας και των άλλων μεταβλητών, χωρίς να χρειάζεται να κάνουμε υποθέσεις σχετικά με τη μορφή του $\lambda_0(t)$. Αυτό σημαίνει ότι μπορούμε να εκτιμήσουμε τα β χωρίς να κάνουμε πλήρη μοντελοποίηση, αφού μπορεί να γίνει χωρίς τη γνώση του $\lambda_0(t)$. Το γεγονός αυτό καθιστά το μοντέλο ημιπαραμετρικό.

Ο Cox προτείνει μια συνάρτηση μερικής πιθανοφάνειας (Partial likelihood), υποθέτοντας ότι δεν υπάρχουν ισότιμες παρατηρήσεις (ties) (D. R. Cox, 1972). Στην πράξη όμως παρατηρούνται ισότιμοι χρόνοι επιβίωσης και έτσι η συνάρτηση μερικής πιθανοφάνειας του Cox, έχει τροποποιηθεί (από τους: Kalbfleisch and Prentice – 1980, Breslow – 1974, Efron – 1977) για να χειρίζεται ισότιμες παρατηρήσεις. Πιο κάτω θα περιγραφεί η διαδικασία εκτίμησης των παραμέτρων από ένα μοντέλο Cox, χωρίς ισότιμες παρατηρήσεις.

Έστω ένα σύνολο n ατόμων, k πλήρεις διακεκριμένοι χρόνοι και $n-k$ δεξιά αποκομμένοι χρόνοι επιβίωσης. Έστω επίσης, $t_{(1)} < t_{(2)} < \dots < t_{(k)}$ οι k ταξινομημένοι πλήρεις χρόνοι και $R_{t_{(i)}}$ το σύνολο των ατόμων που βρίσκονται σε κίνδυνο τον χρόνο $t_{(i)}$. Συμβολίζουμε με $Z_{(i)} = (Z_{(i)1}, Z_{(i)2}, \dots, Z_{(i)p})$, το διάνυσμα των συμμεταβλητών που αντιστοιχεί στο άτομο με πλήρη χρόνο ζωής $t_{(i)}$. Τότε η πιθανότητα αποτυχίας ενός ατόμου στο σύνολο $R_{t_{(i)}}$ ορίζεται ως:

$$\frac{\exp(\sum_{j=1}^p \beta_j Z_{(i)j})}{\sum_{l \in R_i} \exp(\sum_{j=1}^p \beta_j Z_{(i)l})} = \frac{\exp(\beta' Z_{(i)})}{\sum_{l \in R_i} \exp(\beta' Z_{(i)l})}$$

Κάθε αποτυχία συμβάλλει κατά παράγοντα και ως εκ τούτου η μερική συνάρτηση πιθανοφάνειας για το σύνολο των δεδομένων είναι:

$$L(\beta) = \prod_{i=1}^k \left\{ \frac{\exp(\beta' Z_{(i)})}{\sum_{l \in R_i} \exp(\beta' Z_l)} \right\}$$

Από την οποία προκύπτει η εκτιμήτρια μέγιστης πιθανοφάνειας $\hat{\beta}$ του β . Έχει αποδειχθεί από τον Cox ότι η πιο πάνω μερική πιθανοφάνεια μπορεί να χρησιμοποιηθεί ως μια συνηθισμένη συνάρτηση πιθανοφάνειας. Ως εκ τούτου, οι συντελεστές παλινδρόμησης β , εκτιμώνται από τις τιμές $\hat{\beta}$ που μεγιστοποιούν την μερική πιθανοφάνεια $L(\beta)$ ή ισοδύναμα τον λογάριθμό της. Ο λογάριθμος της πιθανοφάνειας είναι:

$$l(\beta) = \log L(\beta) = \sum_{i=1}^k \{ \beta' Z_{(i)} - \log [\sum_{l \in R_i} \exp(\beta' Z_l)] \}$$

Οι πρώτες μερικές παράγωγοι είναι,

$$\frac{\partial l}{\partial \beta_r} = \sum_{j=1}^k Z_{(j)r} - \sum_{j=1}^k \left[\frac{\sum_{i \in R_j} Z_{(i)r} e^{\beta' Z_{(i)}}}{\sum_{i \in R_j} e^{\beta' Z_{(i)}}} \right]$$

Η παραπάνω σχέση αποτελεί σύστημα p -εξισώσεων, το οποίο λύνεται με επαναληπτικές μεθόδους όπως την Newton-Raphson.

2.5.1. Έλεγχοι Υποθέσεων

Όταν ο αριθμός των παρατηρήσεων είναι ικανοποιητικός, οι εκτιμητές παλινδρόμησης από ένα μοντέλο Cox κατανέμονται κανονικά. Για τον λόγο αυτό είναι δυνατόν να εκτελεστούν διάφοροι έλεγχοι υποθέσεων βάσει των εκτιμήσεων αυτών. Παρακάτω θα αναλύσουμε δύο από αυτούς του ελέγχους, τον έλεγχο του λόγου πιθανοφανειών και τον έλεγχο του Wald.

2.5.1.1. Έλεγχος του λόγου πιθανοφανειών – Likelihood Ratio Test

Ο έλεγχος αυτός αποτελεί τον πιο σύνηθες τρόπο για έλεγχο υποθέσεων. Μια υπόθεση μπορεί να είναι $\beta_i=0$ το οποίο σημαίνει ότι ο κίνδυνος εξαρτάται από την συμμεταβλητή (covariate) Z_i . Προσαρμόζεται ένα μοντέλο με τις συμμεταβλητές ($\beta_i \neq 0$) και ένα μοντέλο χωρίς τις συμμεταβλητές ($\beta_i=0$). Έστω, \hat{l}_1 η μεγιστοποιημένη

τιμή του λογαρίθμου της μερικής πιθανοφάνειας του μοντέλου για $\beta_i \neq 0$ και \hat{l}_0 για $\beta_i = 0$. Τότε γίνεται σύγκριση κατά τον έλεγχο αυτό της τιμής του $-2(\hat{l}_0 - \hat{l}_1)$ με την X_1^2 κατανομή. Δηλαδή ο έλεγχος υποθέσεων που πραγματοποιείται είναι ο εξής,

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

Κατά τον ίδιο τρόπο γίνεται έλεγχος για όλες τις συμμεταβλητές, όπου για τον έλεγχο της μηδενικής υπόθεσης, συγκρίνεται η μεταβολή της τιμής $-2(\hat{l}_0 - \hat{l}_p)$ με την τιμή της X^2_p κατανομής.

2.5.1.2 Έλεγχος του Wald - Wald test

Εναλλακτικά, μπορούμε να χρησιμοποιήσουμε τον έλεγχο Wald. Η ελεγχοσυνάρτηση για κάθε μεταβλητή j δίνεται από το εξής,

$$W = \left\{ \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \right\}^2$$

Ο έλεγχος υποθέσεων που πραγματοποιείται είναι και πάλι αυτός που αναφέραμε πιο πάνω στον έλεγχο λόγου πιθανοφανειών. Για τον έλεγχο της μηδενικής υπόθεσης $\beta_j = 0$ η τιμή συγκρίνεται με την X_1^2 κατανομή.

Συχνά παρατηρούμε ότι οι τιμές που προκύπτουν από τους παραπάνω ελέγχους είναι πολύ κοντά, ειδικά όταν το μέγεθος του δείγματος είναι μεγάλο. Στα μικρότερα δείγματα οι τιμές αυτές διαφέρουν και θεωρείται πιο αξιόπιστος έλεγχος αυτός του λόγου πιθανοφανειών.

2.5.2. Έλεγχοι της υπόθεσης αναλογικότητας των κινδύνων

Στην παρούσα υποενότητα θα αναπτύξουμε μερικούς ελέγχους της υπόθεσης αναλογικότητας μέσω γραφημάτων και μέσω καταλοίπων (residuals).

2.5.2.1. Γραφικός έλεγχος

Σύμφωνα με την υπόθεση του αναλογικού κινδύνου, για τη συνάρτηση επιβίωσης ενός ατόμου με διάνυσμα συμμεταβλητών $Z_{(i)}=(Z_{(1)}, Z_{(2)}, \dots, Z_{(p)})$, ισχύει

$$S(t;Z) = [S_0(t)]^{e^{\beta'Z}}$$

Έπειτα, λογαριθμίζοντας δύο φορές την πιο πάνω σχέση καταλήγουμε στο εξής,

$$\log(\log(S(t;Z))) = \beta'Z + \log(S_0(t))$$

Τότε, σχεδιάζοντας την γραφική παράσταση $\log(\log(S(t;Z)))$ συναρτήσει του λογαρίθμου χρόνου περιμένουμε ότι θα μας δώσει 2 ευθείες παράλληλες, αν εξετάζουμε για παράδειγμα 2 ομάδες θεραπείας. Έτσι, αν οι καμπύλες που θα προκύψουν είναι παράλληλες ή σχεδόν παράλληλες θεωρούμε ότι ισχύει η υπόθεση αναλογικών κινδύνων. Οι γραφικές παραστάσεις που δημιουργούμε βασίζονται στην εκτίμηση της συνάρτησης της επιβίωσης με μεθόδους που δε χρησιμοποιούν την υπόθεση αναλογικότητας των κινδύνων, όπως για παράδειγμα ο εκτιμητής Kaplan-Meier. Στην περίπτωση όπου έχουμε μια μεταβλητή με πολλά επίπεδα ή όταν έχουμε πολλές μεταβλητές, είναι δύσκολο να ελεγχθεί με γραφικές μεθόδους αν ισχύει η υπόθεση της αναλογικότητας.

2.5.2.2. Έλεγχος μέσω καταλοίπων

Τα πιο γνωστά κατάλοιπα που χρησιμοποιούνται όταν έχουμε ένα μοντέλο Cox είναι Cox – Snell, τα Martingale κατάλοιπα, τα Schoenfeld κατάλοιπα και τα κατάλοιπα απόκλισης (deviance residuals) (*M J Bradburn et al, 2003*). Στα γραμμικά μοντέλα τα σφάλματα κατανέμονται κανονικά και αν το μοντέλο που προσαρμόζουμε στα δεδομένα είναι σωστό, τα κατάλοιπα θα έχουν προσεγγιστικά κανονική κατανομή. Το ανάλογο του μοντέλου του Cox είναι η σύγκριση του καταλοίπου με μία εκθετική κατανομή με παράμετρο 1. Αυτό συμβαίνει επειδή αν $T \sim S(t)$, τότε $S(t) \sim Uniform[0,1]$ και $-\log S(T) \sim Exp(1)$ (από το θεώρημα Inverse CDF). Για παράδειγμα,

- Γενικευμένα Cox-Snell κατάλοιπα (generalized Cox-Snell residuals):

Αν το μοντέλο είναι σωστό, ο εκτιμώμενος αθροιστικός κίνδυνος $\Lambda(t)$ για κάθε άτομο κατά τη στιγμή του θανάτου ή της αποκοπής θα πρέπει να είναι σαν ένα αποκομμένο δείγμα από μια $Exp(1)$. Η γραφική παράσταση του $\log(-\log S(t))$ συναρτήσεως των Cox-Snell καταλοίπων θα αναμένουμε να είναι μια ευθεία γραμμή με κλίση 1 η οποία περνά από την αρχή των αξόνων $(0,0)$.

- Κατάλοιπα Martingale - Martingale residuals:

Για ένα άτομο i ισχύει $r_i = \delta_i - \widehat{\Lambda}(T_i)$, όπου δ_i ο δείκτης αποτυχίας και $\widehat{\Lambda}(T_i)$ ο εκτιμώμενος αθροιστικός κίνδυνος. Τα martingale κατάλοιπα r_i , έχουν τις εξής ιδιότητες: η μέση τιμή τους είναι ίση με μηδέν, το εύρος τιμών τους είναι μεταξύ $-\infty$ και 1, και είναι περίπου ασυσχέτιστα. Για να είναι καλό το μοντέλο, θα πρέπει να κατανέμονται ομοιόμορφα γύρω από το μηδέν. Τα Martingale κατάλοιπα χρησιμοποιούνται κυρίως για την εύρεση ακραίων τιμών, αλλά κυρίως για την εύρεση της συναρτησιακής μορφής μιας μεταβλητής που πρόκειται να εισαχθεί στο μοντέλο του Cox. Αν κάποιο Martingale κατάλοιπο έχει μια υψηλή αρνητική τιμή, τότε η αντίστοιχη παρατήρηση είναι ακραία τιμή και δεν ερμηνεύεται καλά από το μοντέλο.

2.6. Ανταγωνιστικοί Κίνδυνοι-Competing risks

Οι ανταγωνιστικοί κίνδυνοι εμφανίζονται συχνά στην ανάλυση των δεδομένων επιβίωσης. Για παράδειγμα, στην περίπτωση που εξετάζουμε τον χρόνο μέχρι την εμφάνιση ηπατοκυτταρικού καρκίνου σε ασθενείς που πάσχουν από χρόνια ηπατίτιδα Β, ο θάνατος είναι ένας ανταγωνιστικός κίνδυνος. Οι ανταγωνιστικοί κίνδυνοι αφορούν σε καταστάσεις κατά τις οποίες περισσότερες από μία αιτίες αποτυχίας είναι πιθανές. Αν οι αποτυχίες είναι διαφορετικές αιτίες θανάτου τότε μόνο η πρώτη από αυτές καταγράφεται. Σε άλλες περιπτώσεις οι παρατηρήσεις μετά την πρώτη αποτυχία μπορεί να καταγράφονται, αλλά να μην έχουν κάποιο ενδιαφέρον. Μπορούμε, όπως προκύπτει και από τα παραπάνω, να αναπαραστήσουμε ένα μοντέλο ανταγωνιστικών κινδύνων ως ένα μοντέλο με μία αρχική κατάσταση (άτομο εν ζωή ή χωρίς καταγραφή κάποιου γεγονότος) και με πολλά διαφορετικά

καταληκτικά σημεία .Κατά την εκτίμηση της επίπτωσης των εκβάσεων, οι αναλυτές θα πρέπει να χρησιμοποιήσουν τη συνάρτηση αθροιστικής επίπτωσης και όχι την εκτιμώμενη συνάρτηση επιβίωσης Kaplan-Meier. Η χρήση της συνάρτησης επιβίωσης Kaplan-Meier οδηγεί σε εκτιμήσεις της επίπτωσης που είναι μεροληπτικές προς τα πάνω, ανεξάρτητα από το αν τα ανταγωνιστικά γεγονότα είναι ανεξάρτητα το ένα από το άλλο. Όταν εφαρμόζουν μοντέλα παλινδρόμησης με την παρουσία ανταγωνιστικών κινδύνων, οι ερευνητές μπορούν να επιλέξουν από 2 διαφορετικές οικογένειες μοντέλων: μοντελοποίηση της επίδρασης των συμμεταβλητών στον κίνδυνο έκβασης από συγκεκριμένη αιτία ή μοντελοποίηση της επίδρασης των συμμεταβλητών στη συνάρτηση αθροιστικής επίπτωσης. Το πρώτο επιτρέπει σε κάποιον να εκτιμήσει την επίδραση των συμμεταβλητών επί του ποσοστού εμφάνισης της έκβασης σε εκείνα τα άτομα που είναι επί του παρόντος χωρίς περιστατικό. Το τελευταίο επιτρέπει την εκτίμηση της επίδρασης των συμμεταβλητών στον απόλυτο κίνδυνο της έκβασης με την πάροδο του χρόνου. Η προηγούμενη οικογένεια μοντέλων μπορεί να είναι πιο κατάλληλη για την αντιμετώπιση των αιτιωδών ζητημάτων, ενώ το τελευταίο μοντέλο μπορεί να είναι πιο κατάλληλο για την εκτίμηση της κλινικής πρόγνωσης ενός ασθενούς (Austin et al, 2016).

2.6.1. Στατιστική ανάλυση δεδομένων παρουσία ανταγωνιστικών κινδύνων

Οι στατιστικές μέθοδοι εκτίμησης και μοντελοποίησης υπό την παρουσία ανταγωνιστικών κινδύνων, επικεντρώνονται σε παρατηρήσιμες ποσότητες όπως η συνάρτηση κινδύνου για συγκεκριμένη αιτία (cause-specific hazard function) και η αθροιστική συνάρτηση (cumulative incidence).

2.6.1.1. Αθροιστική συνάρτηση κινδύνου (Cumulative Incidence Function – CIF)

Η εκτίμηση της επίπτωσης ενός γεγονότος ως συνάρτηση του χρόνου παρακολούθησης παρέχει σημαντικές πληροφορίες για τον απόλυτο κίνδυνο

εμφάνισης ενός γεγονότος. Όταν δεν έχουμε ανταγωνιστικούς κινδύνους χρησιμοποιείται η εκτιμήτρια Kaplan-Meier για την εκτίμηση της συνάρτησης επιβίωσης. Το πρόβλημα εδώ είναι ότι ο εκτιμητής K-M υπολογίζει την πιθανότητα εμφάνισης του γεγονότος ενδιαφέροντος εν απουσία ανταγωνιστικών κινδύνων, που είναι γενικά μεγαλύτερη από αυτή παρουσία ανταγωνιστικών κινδύνων (Lau et al, 2009). Επιπλέον, ο υποθετικός πληθυσμός στον οποίο δεν υπάρχουν ανταγωνιστικοί κίνδυνοι μπορεί να μην είναι ο πληθυσμός που παρουσιάζει το μεγαλύτερο ενδιαφέρον για την κλινική ιατρική (Pepe et al, 1993), όπως στο καρδιαγγειακό περιβάλλον όπου ο μη καρδιαγγειακός θάνατος μπορεί να είναι ένα σημαντικό ζήτημα.

Η συνάρτηση αθροιστικής επίπτωσης (CIF) επιτρέπει την εκτίμηση της επίπτωσης της εμφάνισης ενός συμβάντος λαμβάνοντας υπόψη τον ανταγωνιστικό κίνδυνο. Αυτό επιτρέπει σε κάποιον να εκτιμήσει την επίπτωση σε έναν πληθυσμό όπου όλα τα ανταγωνιστικά γεγονότα πρέπει να ληφθούν υπόψη στην κλινική λήψη αποφάσεων. Η αθροιστική συνάρτηση επίπτωσης για την κ-οστή αιτία ορίζεται ως εξής:

$$CIF_k = P(T \leq t, D=k)$$

όπου D υποδηλώνει τον τύπο του συμβάντος. Ένα βασικό σημείο είναι ότι, στη περίπτωση των ανταγωνιστικών κινδύνων, μπορεί να συμβεί μόνο ένας τύπος συμβάντος, έτσι ώστε η εμφάνιση ενός συμβάντος να αποκλείει την επακόλουθη εμφάνιση άλλων τύπων συμβάντων. Η συνάρτηση $CIF_k(t)$ δηλώνει την πιθανότητα εμφάνισης του κ-οστού συμβάντος πριν από το χρόνο t και πριν από την εμφάνιση διαφορετικού τύπου συμβάντος. Η CIF έχει την επιθυμητή ιδιότητα ότι το άθροισμα των εκτιμήσεων CIF της επίπτωσης κάθε επιμέρους έκβασης θα είναι ίσο με τις εκτιμήσεις CIF της επίπτωσης του σύνθετης έκβασης που αποτελείται από όλα τα ανταγωνιστικά γεγονότα. Σε αντίθεση με τη συνάρτηση επιβίωσης χωρίς τους ανταγωνιστικούς κινδύνους, το $CIF_k(t)$ δεν θα προσεγγίσει απαραίτητως την μονάδα όσο μεγαλώνει ο χρόνος, λόγω της εμφάνισης ανταγωνιστικών γεγονότων που αποκλείουν την εμφάνιση γεγονότων τύπου k. Η αθροιστική συνάρτηση μπορεί να εκτιμηθεί με πολλούς τρόπους (πχ naïve εκτιμητής).

2.6.1.2. Συνάρτηση κινδύνου συγκεκριμένης αιτίας (cause-specific hazard function) και συνάρτηση κινδύνου υπο-κατανομής (subdistribution hazard function)

Οι ανταγωνιστικοί κίνδυνοι υπονοούν ότι ένα άτομο μπορεί να αντιμετωπίσει ένα σύνολο από διαφορετικές εκβάσεις ή γεγονότα. Στην περίπτωση αυτή παρουσιάζουν ενδιαφέρον δύο διαφορετικοί τύποι συναρτήσεων κινδύνου: η συνάρτηση κινδύνου συγκεκριμένης αιτίας και η συνάρτηση κινδύνου υπο-κατανομής. Η συνάρτηση κινδύνου ειδικής αιτίας (cause-specific hazard function) ορίζεται ως εξής (Putter, et al, 2006):

$$\lambda_k^{cs}(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t, D=k | T \geq t)}{\Delta t}$$

Η συνάρτηση κινδύνου συγκεκριμένης αιτίας υποδηλώνει τον στιγμιαίο ρυθμό εμφάνισης του κ-οστού συμβάντος σε άτομα που είναι επί του παρόντος χωρίς περιστατικά (πχ σε άτομα που δεν έχουν βιώσει κανένα από τα διαφορετικά είδη συμβάντων). Αν κάποιος εξέταζε δύο τύπους συμβάντων, θάνατο που οφείλεται σε καρδιαγγειακά αίτια και θάνατο που οφείλεται σε μη καρδιαγγειακά αίτια, τότε ο κίνδυνος καρδιαγγειακού θανάτου συγκεκριμένης αιτίας υποδηλώνει το στιγμιαίο ρυθμό καρδιαγγειακού θανάτου σε άτομα που δεν έχουν υποστεί ακόμα κανένα γεγονός (δηλαδή άτομα που είναι ακόμα ζωντανά).

Η συνάρτηση υπο-κατανομής κινδύνου, ορίστηκε από τους Fine και Gray (Fine JP & Gray RJ, 1999) ως εξής:

$$\lambda_j^{sub}(t; Z) = \lim_{h \rightarrow 0} \frac{1}{h} P[t \leq T \leq t + h, C = j | T \geq t \cup (T \leq t \cap C \neq j), Z]$$

$$= \frac{[dF_j(t; Z)/dt]}{[1 - F_j(t; Z)]} = \frac{-d}{dt} \log[1 - F_j(t; Z)]$$

Η παραπάνω συνάρτηση υποδηλώνει τον στιγμιαίο κίνδυνο αποτυχίας (θανάτου) από το κ-οστό συμβάν σε άτομα που δεν έχουν βιώσει ακόμα ένα γεγονός τύπου κ.

Πρέπει να λάβουμε υπόψη όμως, ότι αυτό το σύνολο κινδύνων περιλαμβάνει όσους είναι παρόντες χωρίς κανένα συμβάν, καθώς και όσους έχουν αντιμετωπίσει προηγουμένως ένα ανταγωνιστικό γεγονός. Αυτό διαφέρει από τον κίνδυνο που έχει οριστεί για τη συνάρτηση κινδύνου συγκεκριμένων αιτιών, η οποία περιλαμβάνει μόνο εκείνους που είναι επί του παρόντος χωρίς συμβάντα. Χρησιμοποιώντας το ίδιο παράδειγμα με πιο πάνω, ο κίνδυνος υπο-κατανομής του καρδιαγγειακού θανάτου υποδηλώνει τον στιγμιαίο ρυθμό καρδιαγγειακού θανάτου σε άτομα που είναι ακόμα ζωντανά (δηλαδή που δεν έχουν υποστεί ακόμα κανένα συμβάν) ή που έχουν πεθάνει προηγουμένως από μη καρδιαγγειακά αίτια.

2.6.2. Μοντέλα παλινδρόμησης παρουσία ανταγωνιστικών κινδύνων

Στις περιπτώσεις όπου παρουσιάζονται ανταγωνιστικοί κίνδυνοι, υπάρχουν δύο διαφορετικά μοντέλα παλινδρόμησης: μοντελοποίηση κινδύνου που σχετίζεται με συγκεκριμένη αιτία και μοντελοποίηση της αθροιστικής συνάρτησης.

2.6.2.1. Μοντελοποίηση κινδύνου συγκεκριμένης αιτίας

Για τη διερεύνηση των παραγόντων που μπορεί να επηρεάσουν τον στιγμιαίο ρυθμό αποτυχίας από μια συγκεκριμένη αιτία παρουσία όλων των άλλων, ένας ανταγωνιστικός κίνδυνος ανάλογος του μοντέλου αναλογικού κινδύνου Cox φαίνεται να είναι μια εύλογη επιλογή (*Holt , 1978*). Ένα ημιπαραμετρικό μοντέλο αναλογικών κινδύνων Cox για τη συνάρτηση κινδύνου από ειδική αιτία ορίζεται ως (*Bakoyiannis & Touloumi, 2010*):

$$\lambda_j(t; Z) = \lambda_{0j}(t) \exp(\beta_j' Z)$$

Στην πράξη, ένα τέτοιο μοντέλο μπορεί να εκτιμηθεί με την επεξεργασία παρατηρήσεων με αποτυχίες από όλες τις άλλες αιτίες, αλλά με τη *j* παρατήρηση ως αποκομμένη και την τοποθέτηση ενός προτύπου αναλογικού κινδύνου Cox στα προκύπτοντα δεδομένα χρησιμοποιώντας τυποποιημένο στατιστικό πακέτο. Η

διαδικασία δεν απαιτεί ανεξαρτησία των χρόνων λανθάνουσας αποτυχίας (ή οποιαδήποτε άλλη παρόμοια μη αποδεκτή υπόθεση). Ενώ η κοινή προσέγγιση είναι να τοποθετούμε ξεχωριστά μοντέλα, ένα για κάθε αιτία, είναι δυνατόν να τοποθετούνται ταυτόχρονα τα μοντέλα κινδύνου για όλες τις αιτίες και να ελέγχουμε την ισότητα των επιδράσεων των συμμεταβλητών σε διαφορετικούς τρόπους αποτυχίας, μέσω μιας μεθόδου αύξησης των δεδομένων (*Lunn and McNeil, 1995*). Υπάρχουν δύο προσεγγίσεις για την τοποθέτηση του μοντέλου κινδύνων για συγκεκριμένες αιτίες. Η πρώτη είναι να στρωματοποιηθεί ανά τύπο αποτυχίας, η οποία δίνει ίδιες εκτιμήσεις με εκείνες που λαμβάνονται με την τοποθέτηση ξεχωριστών μοντέλων αναλογικών κινδύνων για κάθε αιτία αποτυχίας. Η δεύτερη είναι να συμπεριληφθεί ο τύπος αποτυχίας ως ανεξάρτητη μεταβλητή στο μοντέλο, το οποίο υποθέτει αναλογικούς κινδύνους μεταξύ των διαφόρων τύπων αποτυχίας. Ωστόσο, πρέπει να σημειωθεί ότι η τελευταία υπόθεση μπορεί να μην είναι κατάλληλη σε πολλές περιπτώσεις στη βιοϊατρική έρευνα. Και στις δύο προσεγγίσεις, η επίδραση μιας μεταβλητής σε κάθε τύπο αποτυχίας μπορεί να εκτιμηθεί με τη συμπερίληψη στο μοντέλο των αντίστοιχων συμμεταβλητών. Εναλλακτικά, είναι δυνατό να συμπεριληφθούν στους όρους αλληλεπίδρασης μοντέλου μια συμμεταβλητή με τύπο αποτυχίας, προκειμένου να δοκιμαστεί η υπόθεση μιας κοινής επίδρασης της συμμεταβλητής σε διαφορετικούς τύπους αποτυχίας. Εάν ισχύει η υπόθεση, επιτυγχάνεται αποδοτικότητα.

2.6.2.2. Μοντελοποίηση αθροιστικής επίπτωσης

Οι Fine και Gray πρότειναν ένα ημιπαραμετρικό μοντέλο αναλογικών κινδύνων για τον κίνδυνο υπο-κατανομής του γεγονότος ενδιαφέροντος (*Fine & Gray, 1999*):

$$\lambda_j^{sub}(t; Z) = \lambda_{oj}^{sub}(t) \exp(\beta_j' Z)$$

Με βάση το μοντέλο αυτό, η CIF είναι,

$$F_j(t; Z) = 1 - \exp\left[-\int_0^t \lambda_j^{sub}(u; Z) du\right]$$

Το παραπάνω μοντέλο έχει περιγραφεί και ως μοντέλο παλινδρόμησης CIF. Το όνομα αυτό υποδηλώνει ρητά τη σχέση μεταξύ του κινδύνου υπο-κατανομής και της

επίπτωσης στην επίπτωση του συμβάντος. Δηλαδή, μπορεί κάποιος να προβλέψει άμεσα την αθροιστική επίπτωση για ένα γεγονός ενδιαφέροντος χρησιμοποιώντας τη συνηθισμένη σχέση μεταξύ του κινδύνου και της συνάρτησης επίπτωσης κάτω από το μοντέλο αναλογικών κινδύνων. Έτσι, το μοντέλο κινδύνου υπο-κατανομής επιτρέπει σε κάποιον να εκτιμήσει την επίδραση των ανεξάρτητων μεταβλητών στην αθροιστική συνάρτηση επίπτωσης για το γεγονός που μας ενδιαφέρει.

Οι Lau et al, υποστηρίζουν ότι τα μοντέλα κινδύνου συγκεκριμένης αιτίας είναι καλύτερα για τη μελέτη της αιτιολογίας της ασθένειας, όπου το μοντέλο κινδύνου υπο-κατανομής χρησιμοποιείται για την πρόβλεψη του κινδύνου ενός ατόμου (*Lau et al, 2009*). Επιπλέον, ο κίνδυνος υπο-κατανομής μπορεί να παρουσιάζει μεγαλύτερο ενδιαφέρον αν κάποιος ενδιαφέρεται για τη συνολική επίδραση των μεταβλητών στην επίπτωση της έκβασης ενδιαφέροντος, ακόμη και όταν οι προβλέψεις επίπτωσης δεν παρουσιάζουν άμεσο ενδιαφέρον. Τα επιχειρήματα αυτά υποδηλώνουν ότι τα μοντέλα κινδύνου υπο-κατανομής θα πρέπει να χρησιμοποιούνται για την ανάπτυξη μοντέλων κλινικής πρόβλεψης και κλίμακων εκτίμησης κινδύνου για εκβάσεις επιβίωσης, ενώ τα μοντέλα κινδύνου συγκεκριμένης αιτίας μπορεί να είναι πιο κατάλληλα για την αντιμετώπιση επιδημιολογικών ζητημάτων. Το σκεπτικό πίσω από αυτή την πρόταση είναι ότι η συνάρτηση κινδύνου για συγκεκριμένες αιτίες υποδηλώνει τον στιγμιαίο ρυθμό της πρωταρχικής έκβασης για άτομα που είναι επί του παρόντος ελεύθερα νόσου. Έτσι, ένας συντελεστής παλινδρόμησης από ένα μοντέλο κινδύνου συγκεκριμένης αιτίας μπορεί να ερμηνευτεί ως η σχετική επίπτωση της αντίστοιχης μεταβλητής στην αύξηση του ποσοστού εμφάνισης της πρωτογενούς έκβασης σε άτομα που είναι επί του παρόντος ελεύθερα νόσου. Σε αντίθεση με αυτό, τα κλινικά μοντέλα πρόβλεψης και οι κλίμακες εκτίμησης κινδύνου ενδιαφέρονται για την εκτίμηση της απόλυτης επίπτωσης του γεγονότος ενδιαφέροντος. Καθώς το μοντέλο κινδύνου υπο-κατανομής επιτρέπει σε κάποιον να μοντελοποιήσει άμεσα την επίδραση των μεταβλητών επί της επίπτωσης της πρωτογενούς έκβασης, αφού καταγράψει τα ανταγωνιστικά γεγονότα, προσφέρεται φυσικά για πρόβλεψη κινδύνου.

Πίνακας 2.1. Συστάσεις για την ανάλυση ανταγωνιστικών κινδύνων σε δεδομένα επιβίωσης.

- Οι αθροιστικές συναρτήσεις κινδύνου (CIF) θα πρέπει να χρησιμοποιούνται για την εκτίμηση της επίπτωσης καθενός από τους διαφορετικούς τύπους ανταγωνιστικών κινδύνων. Να μην χρησιμοποιείται η εκτίμηση Kaplan-Meier για τον σκοπό αυτό.
- Οι ερευνητές πρέπει να αποφασίσουν κατά πόσο ο ερευνητικός στόχος είναι η αντιμετώπιση των αιτιολογικών ερωτημάτων ή η εκτίμηση της επίπτωσης ή η πρόγνωση.
- Να γίνεται χρήση του μοντέλου κινδύνου υπο-κατανομής Fine-Gray όταν εστιάζουμε στην εκτίμηση της επίπτωσης ή στην πρόγνωση υπό την παρουσία ανταγωνιστικών κινδύνων.
- Να γίνεται χρήση του μοντέλου κινδύνου συγκεκριμένης αιτίας όταν εστιάζουμε στην αντιμετώπιση αιτιολογικών ζητημάτων.
- Σε ορισμένες περιπτώσεις, και οι 2 τύποι μοντέλων παλινδρόμησης θα πρέπει να εκτιμηθούν για κάθε ένας από τους ανταγωνιστικούς κινδύνους ώστε να καταστεί δυνατή η πλήρης κατανόηση της επίδρασης των μεταβλητών στην επίπτωση και στον ρυθμό εμφάνισης κάθε έκβασης.

Κεφάλαιο 3

Ανάπτυξη και Επικύρωση ενός προγνωστικού μοντέλου

Στο κεφάλαιο αυτό θα αναπτυχθεί η διαδικασία «κτισίματος» ενός μοντέλου πρόβλεψης, ο τρόπος εισαγωγής των μεταβλητών στο μοντέλο και η επικύρωσή του.

3.1. Ανάπτυξη ενός προγνωστικού μοντέλου

Παρά το σημαντικό μέρος της μεθοδολογικής βιβλιογραφίας και των δημοσιευμένων οδηγιών για τον τρόπο διεξαγωγής έρευνας πρόβλεψης, τα περισσότερα μοντέλα υποφέρουν από μεθοδολογικές ελλείψεις (*Mushkudiani et al, 2008 & Mallett et al, 2010*). Ως εκ τούτου οι Steyerberg και Vergouwe (*Steyerberg & Vergouwe, 2014*) πρότειναν κάποια βήματα για την ανάπτυξη ενός μοντέλου πρόβλεψης τα οποία πρέπει να πληρούνται όλα αφού ένα μοντέλο πρόβλεψης έχει ιδιαίτερη βαρύτητα στην εκτίμηση κινδύνου για την παρουσία ασθένειας (διάγνωση) και στην μελλοντική πορεία της νόσου (πρόγνωση) για μεμονωμένους ασθενείς.

3.1.1. Ερευνητικό ζήτημα και έλεγχος δεδομένων

Το πρώτο βήμα είναι η εξέταση του ερευνητικού ζητήματος και ο αρχικός έλεγχος των δεδομένων. Στο βήμα αυτό θα πρέπει να αντιμετωπιστούν κάποια ερωτήματα. Το πρώτο είναι το ποιο είναι το ακριβές ερευνητικό ζήτημα. Είναι πολύ σημαντικό η έκβαση που θα μελετηθεί να ενδιαφέρει τους ασθενείς, τους γιατρούς και γενικά την κλινική κοινότητα. Στην έρευνα πρόβλεψης μας ενδιαφέρει συχνά η κατανόηση των παραγόντων που προβλέπουν το τελικό σημείο (endpoint) και η παροχή εκτιμήσεων για τον απόλυτο κίνδυνο για το τελικό σημείο με βάση ένα συνδυασμό παραγόντων (*Shmueli, 2010*). Ένα άλλο ερώτημα είναι το τι είναι ήδη γνωστό για τους προγνωστικούς παράγοντες. Η βιβλιογραφική ανασκόπηση και οι στενές αλληλεπιδράσεις μεταξύ στατιστικών και κλινικών ερευνητών είναι σημαντικές για την ενσωμάτωση της γνώσης του αντικειμένου στη διαδικασία μοντελοποίησης. Παρόλα αυτά, η μέτρηση της έκβασης θα πρέπει να είναι τυφλή (blinded) και

ανεξάρτητη από οποιαδήποτε γνώση προγνωστικών παραγόντων ούτως ώστε να αποφεύγονται πιθανές μεροληψίες (bias) (Moons et al, 2012).

Επιπλέον, στο βήμα αυτό θα πρέπει να είναι ξεκάθαρο το πώς έγινε η επιλογή των δεδομένων. Ιδανικά τα δεδομένα θα πρέπει να προέρχονται από προοπτικές μελέτες κοορτής και ως εκ τούτου οι τυχαίοποιημένες κλινικές δοκιμές είναι κατάλληλες για την ανάπτυξη προγνωστικού μοντέλου αφού οι μελέτες αυτές είναι η ιδανική μορφή προοπτικής έρευνας (Grobbee et al, 2010). Σημαντικό, επίσης, ζήτημα είναι το αν οι μετρήσεις ήταν αξιόπιστες και πλήρως μετρημένες. Πολλά σύνολα δεδομένων είναι ελλιπή σε σχέση με τις τιμές ορισμένων πιθανών προγνωστικών παραγόντων. Από προεπιλογή, οι ασθενείς με τυχόν ελλείπουσα τιμή αποκλείονται από τις στατιστικές αναλύσεις, όμως αυτό δεν είναι αποτελεσματικό δεδομένου ότι χάνονται διαθέσιμες πληροφορίες άλλων προγνωστικών παραγόντων. Για την επίλυση του προβλήματος αυτού μπορούμε να συμπληρώσουμε τις καλύτερες εκτιμήσεις για τις ελλείπουσες τιμές, αξιοποιώντας τη συσχέτιση μεταξύ των μεταβλητών στο σύνολο δεδομένων (Altman, 2008). Υπάρχουν διάφορες στατιστικές προσεγγίσεις για την αντιμετώπιση των ελλειπουσών τιμών. Η μέθοδος πολλαπλής αντικατάστασης (multiple imputation method) είναι μια διαδικασία η οποία συμπληρώνει τις ελλείπουσες τιμές πολλές φορές (τυπικά τουλάχιστον 5 φορές) για την κατάλληλη αντιμετώπιση της τυχαιότητας της διαδικασίας εκτίμησης. Οι ελλείπουσες τιμές αντικαθίστανται m φορές χρησιμοποιώντας m ανεξάρτητες τυχαίες τιμές από το μοντέλο αντικατάστασης (Rubin, 1987).

3.1.2. Κωδικοποίηση των προγνωστικών μεταβλητών

Ένα από τα πλέον σημαντικότερα ζητήματα για την ανάπτυξη ενός προγνωστικού μοντέλου είναι ο τρόπος εισαγωγής των μεταβλητών στο μοντέλο. Οι κατηγορικές και οι συνεχείς μεταβλητές πρόβλεψης μπορούν να κωδικοποιηθούν με διαφορετικούς τρόπους. Αρχικά γίνεται έλεγχος για το είδος και την κατανομή των μεταβλητών ούτως ώστε να βρεθούν οι κατάλληλοι τρόποι κωδικοποίησης τους (Steyerberg, 2009). Αρκετές φορές κωδικοποιούμε τις μεταβλητές ειδικά αν αυτές είναι κατηγορικές. Πρέπει να είμαστε όμως προσεκτικοί. Ειδικά στην περίπτωση που η μεταβλητή είναι ονομαστική, είναι λάθος να αντικαταστήσουμε τις κατηγορίες με

αριθμητικές τιμές διότι έτσι οι κατηγορίες αποκτούν προσδιορισμένη σχέση και διάταξη. Συνήθως οι κατηγορικές μεταβλητές είναι κωδικοποιημένες ως παράγοντες (factors). Για παράδειγμα, το αλκοόλ μπορούμε να κωδικοποιήσουμε ως εξής, 1: 0 ποτήρια/ημέρα, 2: 1-4 ποτήρια/ημέρα και 3: >4 ποτήρια/ημέρα. Σημαντικό εδώ είναι ο καθορισμός της κατηγορίας αναφοράς (reference category). Συνήθως ως κατηγορία αναφοράς ορίζουμε την μικρότερη ή την μεγαλύτερη κατηγορία ενώ όταν έχουμε δυαδική μεταβλητή (binary) 0/1 ως κατηγορία αναφοράς ορίζουμε το «0». Στην περίπτωση όπου η κατηγορία 1 είναι η κατηγορία αναφοράς στο πιο πάνω παράδειγμα, το μοντέλο Cox θα έχει τη μορφή:

$$\lambda(t;Z)=\lambda_0(t)\exp(\beta_2Z_2+\beta_3Z_3)$$

όπου Z_2 και Z_3 είναι ψευδομεταβλητές (dummies) που αντιπροσωπεύουν τις κατηγορίες 2 και 3 του αλκοόλ αντίστοιχα. Αντίθετα δεν υπάρχει τόσο μεγάλο πρόβλημα αν η μεταβλητή είναι διατάξιμη (ordinal). Το μόνο ερωτηματικό σε τέτοιου είδους κωδικοποιήσεις είναι αν υπάρχει συμφωνία μεταξύ των αποστάσεων των κατηγοριών της διατάξιμης μεταβλητής και της διακριτής μεταβλητής που την αντικαθιστά και χρειάζεται προσοχή διότι αν αγνοηθεί η διάταξη αυτών των μεταβλητών μπορεί να υπάρξει απώλεια προγνωστικής ικανότητας. Ένας άλλος τρόπος αντιμετώπισης των διατάξιμων μεταβλητών είναι να τις εντάξουμε στο μοντέλο ως γραμμικές μεταβλητές. Επίσης, στην περίπτωση κατηγορικής μεταβλητής οι κατηγορίες με σπάνια εμφάνιση μπορούν να συγχωνευθούν με άλλες κατηγορίες οι οποίες εμφανίζονται πιο συχνά.

Στην περίπτωση όπου η μεταβλητή είναι συνεχής, υπάρχουν πολλοί τρόποι κωδικοποίησης. Οι συνεχείς παράγοντες πρόβλεψης μπορούν συχνά να μοντελοποιηθούν ως μια γραμμική συσχέτιση σε ένα μοντέλο παλινδρόμησης (Royston, 2008). Στην περίπτωση αυτή υποθέτουμε ότι η επίδραση της μεταβλητής αυτής είναι η ίδια σε όλο το εύρος του παράγοντα. Αν κρίνουμε ότι η μεταβλητή πρέπει να εισαχθεί στο μοντέλο ως γραμμική, το μοντέλο αναλογικών κινδύνων Cox θα έχει την εξής μορφή:

$$\lambda(t;Z)=\lambda_0(t)\exp(\beta Z)$$

όπου Z είναι η ανεξάρτητη μεταβλητή και e^β η μεταβολή στον λόγο κινδύνου (hazard ratio) εμφάνισης της έκβασης για κάθε μία μονάδα αύξησης της συνεχούς

μεταβλητής. Επίσης, μια συνεχής μεταβλητή μπορεί να διχοτομηθεί. Στην περίπτωση αυτή τα μειονεκτήματα είναι αρκετά, αφού διχοτομώντας την μεταβλητή χάνουμε αρκετή πληροφορία των δεδομένων και η πρόβλεψη δεν είναι ακριβής. Για παράδειγμα, αν διχοτομήσουμε την ηλικία σε άνω των 50 ετών και κάτω των 50 ετών η εκτίμηση του κινδύνου πάνω ή κάτω από ένα όριο θα είναι μη φυσιολογική. Καλό είναι οι συνεχείς παράγοντες πρόβλεψης να μη διχοτομούνται κατά την φάση ανάπτυξης του μοντέλου, αφού έχουν χαθεί πολύτιμες πληροφορίες (Royston, *et al*, 2006). Ένας άλλος τρόπος αντιμετώπισης της συνεχούς μεταβλητής, τον οποίο συναντούμε συχνά σε πολλές μελέτες, είναι η κατηγοριοποίησή της. Δηλαδή εάν η μεταβλητή έχει κατηγοριοποιηθεί σε τρεις κατηγορίες, η μορφή του μοντέλου θα είναι

$$\lambda(t;Z)=\lambda_0(t)\exp(\beta_2Z_2+\beta_3Z_3)$$

όπου Z_2 και Z_3 είναι ψευδομεταβλητές για τις κατηγορίες 2 και 3 αντίστοιχα και η κατηγορία 1 είναι κατηγορία αναφοράς. Στην περίπτωση αυτή τα e^{β_2} και e^{β_3} εκφράζουν την μεταβολή στο hazard ratio της κάθε κατηγορίας αντίστοιχα σε σχέση με την κατηγορία αναφοράς. Σημειώνουμε επίσης εδώ ότι μια τέτοια πράξη δεν ενδείκνυται για ένα μοντέλο πρόβλεψης αφού η πληροφορία που χάνεται κατηγοριοποιώντας μια συνεχή είναι αρκετά μεγάλη.

Οι γραμμικοί όροι προφανώς δεν είναι κατάλληλοι για τους παράγοντες πρόβλεψης με μη γραμμικές συσχετίσεις. Αυτό μπορεί να αντιμετωπιστεί χρησιμοποιώντας splines ή κλασματικά πολυώνυμα. Τα splines είναι πολύ χρήσιμα λόγω της ευελιξίας, της απλότητας, της ευκολίας και της ακρίβειας της αξιολόγησής τους και της ικανότητάς τους να προσεγγίζουν πολύπλοκα σχήματα με τοποθέτηση μιας καμπύλης. Για τα προγνωστικά μοντέλα προτείνονται οι περιορισμένες κυβικές splines (Restricted cubic splines). Κατά τη διαδικασία αυτή γίνεται μια μετατροπή της ανεξάρτητης μεταβλητής ούτως ώστε να μπορούν να χρησιμοποιηθούν όχι μόνο σε συνηθισμένη παλινδρόμηση ελαχίστων τετραγώνων, αλλά και σε ανάλυση επιβίωσης, σε λογιστική παλινδρόμηση και ούτω καθεξής. Το εύρος τιμών της ανεξάρτητης μεταβλητής χωρίζεται, με "κόμβους" που ορίζουν το τέλος ενός τμήματος και την αρχή του επόμενου. Ξεχωριστές καμπύλες είναι κατάλληλες για κάθε τομέα. Συνολικά, τα splines ορίζονται έτσι ώστε η προκύπτουσα προσαρμοσμένη καμπύλη να είναι ομαλή και συνεχής (Ruth, 2016). Δηλαδή οι καμπύλες είναι της μορφής X^3

και έτσι περιορίζουν την μεταβλητή να είναι γραμμική στις ουρές. Για την δημιουργία μιας spline απαιτείται ο καθορισμός κόμβων (knots) μέσα από τους οποίους περνάει η συνάρτηση spline και σχηματίζει μια όσο το δυνατόν ομαλότερη καμπύλη. Πρέπει να γίνουν δύο επιλογές: η πρώτη είναι ο αριθμός και η θέση των κόμβων και η δεύτερη είναι ο βαθμός πολωνύμου που πρέπει να χρησιμοποιηθεί μεταξύ των κόμβων. Απαιτούμε να οριστούν οι μεμονωμένες καμπύλες με τέτοιο τρόπο ώστε να συναντώνται στους κόμβους και στην πραγματικότητα να ενταχθούν "ομαλά". Συνεπώς, μια συνάρτηση spline είναι μια ομάδα ομαλά συνδεδεμένων τετραγωνικών πολωνύμων. "Ομαλά ενωμένα" σημαίνει ότι για τα πολωνύμα του βαθμού n , τόσο η συνάρτηση spline όσο και τα πρώτα παράγωγα $n-1$ είναι συνεχή στους κόμβους. Εάν χρησιμοποιούνται κόμβοι k , η τοποθέτηση ενός πολωνύμου βαθμού n απαιτεί $k+n+1$ παραμέτρους παλινδρόμησης. Εμπειρικά έχει φανεί ότι πέντε κόμβοι είναι αρκετοί για να δημιουργήσουν αρκετά μη-γραμμικά μονοπάτια. Στην πράξη, συνήθως χρησιμοποιούνται κυβικά splines (δηλ. πολωνύμα τρίτου βαθμού), που απαιτούν επιπλέον $k+3$ συντελεστές, εκτός από το intercept. Αυτός είναι ο μικρότερος βαθμός πολωνύμου που επιτρέπει μια καμπυλότητα, παρέχοντας επαρκή ευελιξία για την τοποθέτηση δεδομένων, ενώ δεν απαιτεί πολλούς βαθμούς ελευθερίας όπως τα splines υψηλότερης τάξης.

Ένας άλλος τρόπος αντιμετώπισης μη γραμμικών σχέσεων είναι τα κλασματικά πολωνύμα. Εδώ γίνεται χρήση τετραγωνικών ή κυβικών όρων, δηλ. X^2 , X^3 , όπου X μια συνεχής μεταβλητή, και χρησιμοποιώντας στατιστικές μεθόδους μπορούμε να εξετάσουμε αν τα μοντέλα με τους όρους αυτούς προσαρμόζονται καλύτερα. Ο έλεγχος αυτός μπορεί να γίνει εύκολα με χρήση της LR τεστ, αφού τα μοντέλα είναι εμφωλευμένα (nested). Επίσης μπορεί να γίνει λογαριθμικός μετασχηματισμός $\log(X)$ καθώς επίσης και εκθετικός $\exp(X)$.

3.1.3. Επιλογή τελικού μοντέλου

Μπορούν να ακολουθηθούν διάφορες στρατηγικές για να επιλεγούν οι παράγοντες πρόβλεψης που θα συμπεριληφθούν στο προγνωστικό μοντέλο. Οι μέθοδοι βηματικής επιλογής (stepwise selection) χρησιμοποιούνται ευρέως για τη μείωση των ενός συνόλου υποψηφίων προγνωστικών παραγόντων. Υπάρχουν 3

τρόποι βηματικής επιλογής: η προδρομική μέθοδος επιλογής (forward selection), η οπισθοδρομική μέθοδος επιλογής (backward selection) και ο συνδυασμός αυτών (stepwise). Στην ανάλυση επιβίωσης και συνεπώς στα μοντέλα παλινδρόμησης αναλογικών κινδύνων Cox χρησιμοποιείται το πληροφοριακό κριτήριο Akaike Information Criterion (AIC) για τον καθορισμό των στατιστικά σημαντικών παραγόντων το οποίο βασίζεται στο:

$$\text{minimize AIC} = -2\log(\text{likelihood}) + (k * q)$$

όπου q ο αριθμός των παραμέτρων στο τρέχον μοντέλο και k το πέναλτι που δίνουμε (συνήθως k=3). Γενικά, όσο πιο μικρό το AIC τόσο καλύτερο το μοντέλο, έτσι, βέλτιστο είναι το μοντέλο το οποίο ελαχιστοποιεί το AIC (*Hamparsum B, 1987*).

Κατά την προδρομική μέθοδο ξεκινάμε από ένα άδειο (null) μοντέλο και προσθέτουμε τις στατιστικά σημαντικές μεταβλητές, δηλαδή όποια μεταβλητή βελτιστοποιεί (με την προσθήκη της) το κριτήριο AIC προστίθεται στο μοντέλο.

Κατά την οπισθοδρομική μέθοδο ξεκινάμε από το πλήρες (full) μοντέλο με όλες τις υποψήφιες μεταβλητές και αφαιρούμε τις μη στατιστικά σημαντικές, δηλαδή όποια μεταβλητή βελτιστοποιεί (με την εξάλειψή της) το κριτήριο AIC εξαλείφεται από το μοντέλο. Κατά την stepwise διαδικασία γίνεται ένας συνδυασμός των δύο πιο πάνω διαδικασιών, αφού σε κάθε βήμα εξετάζεται εάν η προσθήκη ή η αφαίρεση κάποιας μεταβλητής βελτιώνει το μοντέλο (*Hocking et al, 1976*).

Γενικά, στη βιβλιογραφία φαίνεται ότι προτιμάται ο οπισθοδρομικός τρόπος διότι μπορούν να ελεγχθούν όλοι οι προγνωστικοί παράγοντες και να παραμείνουν στο μοντέλο όλοι οι συσχετισμένοι παράγοντες κάτι το οποίο μπορεί να μην συμβεί με τον προδρομικό τρόπο επιλογής. Ωστόσο, φαίνεται ότι και οι δύο τρόποι επιλογής έχουν αρκετά μειονεκτήματα. Αυτό συμβαίνει συνήθως όταν το δείγμα δεν είναι αρκετά μεγάλο με αποτέλεσμα οι παράμετροι να υπερεκτιμούνται (*Frank et al, 2001*). Στην πράξη, για την επιλογή του βέλτιστου μοντέλου, συχνά χρησιμοποιούνται τα βήματα του Collett (Collett's approach). Η προσέγγιση αυτή υποθέτει ότι όλες οι μεταβλητές είναι ισότιμες και ότι δεν υπάρχει a-priori γνώση ότι πρέπει να συμπεριληφθεί κάποια συγκεκριμένη μεταβλητή στο μοντέλο. Η προσέγγιση αυτή περιλαμβάνει τα εξής τέσσερα βήματα (*D. Collett, 2003*):

1) Τρέχουμε όλα τα μονοπαραγοντικά μοντέλα και ελέγχουμε κάθε μία μεταβλητή εάν είναι στατιστικά σημαντική ή όχι (συνήθως στο επίπεδο στατιστικής σημαντικότητας (ε.σ.σ) 20%).

2) Τρέχουμε το πολυπαραγοντικό μοντέλο με όλες τις στατιστικά σημαντικές μεταβλητές που βρήκαμε στο βήμα 1 και αρχίζουμε σταδιακά να βγάζουμε μία μία τις μεταβλητές (με ε.σ.σ 10%).

3) Ξεκινάμε με το μοντέλο από το βήμα 2 και προσθέτουμε τις μη στατιστικές μεταβλητές που προέκυψαν από το βήμα 1 και τις ξανά βάζουμε στο μοντέλο για να ελέγξουμε εάν τώρα είναι στατιστικά σημαντικές (ε.σ.σ 10%).

4) Ελέγχουμε το τελευταίο μοντέλο και αφαιρούμε τις μεταβλητές που είναι $p\text{-value} \geq 0.05$. Διερευνούμε για αλληλεπιδράσεις.

Το μοντέλο που προκύπτει από την παραπάνω διαδικασία είναι αυτό με την μικρότερη τιμή AIC.

Υπάρχει ένας περιορισμός ο οποίος λέει ότι αν τα δεδομένα έχουν για παράδειγμα 100 θανάτους, εμείς μπορούμε να εκτιμήσουμε $100/10=10$ μεταβλητές στο μοντέλο. Δηλαδή διαιρούμε με το 10 τον αριθμό των συμβάντων ούτως ώστε να αξιολογήσουμε πόσες μεταβλητές μπορούμε να εκτιμήσουμε. Αν δεν εφαρμόσουμε αυτό τον κανόνα το μοντέλο θα υπερεκτιμά τον κίνδυνο εμφάνισης της έκβασης.

Η πιο πάνω διαδικασία, ειδικά στην περίπτωση όπου έχουμε λογαριθμιστική παλινδρόμηση, γίνεται χρησιμοποιώντας ως κριτήριο το Likelihood Ratio Test (LR test). Κατά τον έλεγχο αυτό χρησιμοποιούνται εμφωλευμένα (nested) μοντέλα, δηλαδή το μικρό μοντέλο είναι υποπερίπτωση του μεγάλου. Το μεγάλο μοντέλο (με τις περισσότερες παραμέτρους) έχει πάντα μεγαλύτερη πιθανοφάνεια από το μικρότερο. Το στατιστικό LR υπολογίζεται ως εξής:

$$D = -2(\ln(\text{πιθανοφάνεια μικρού μοντέλου}) - \ln(\text{πιθανοφάνεια μεγάλου μοντέλου}))$$

Το στατιστικό D ακολουθεί την χ^2 κατανομή με βαθμούς ελευθερίας όσους και η διαφορά των παραμέτρων των 2 μοντέλων ($df_{\text{μεγάλου}} - df_{\text{μικρού}}$). Για παράδειγμα, εάν το μεγάλο μοντέλο έχει 4 παραμέτρους και το μικρό 2 τότε το D θα ακολουθεί χ^2 κατανομή με 2 ($4-2=2$) βαθμούς ελευθερίας. Το καλύτερο μοντέλο κρίνεται από το $p\text{-value}$ της τιμής της διαφοράς D. Γίνεται ο έλεγχος υποθέσεων:

H₀: μεγάλο μοντέλο καλύτερο

H₁: μικρό μοντέλο καλύτερο

αν $p\text{-value} < 0.05$ απορρίπτουμε την H_0 και αν $p\text{-value} > 0.05$ δεχόμαστε την H_0 (J. Royle et al, 2009).

Αφού καταλήξουμε στο τελικό μοντέλο γίνεται γραφικός έλεγχος για να δούμε πως προσαρμόζεται το μοντέλο. Στην περίπτωση όπου έχουμε δεδομένα επιβίωσης συνήθως γίνεται έλεγχος των Cox-Snell καταλοίπων (Generalized Cox-Snell residuals). Αν το μοντέλο είναι σωστό, ο εκτιμώμενος αθροιστικός κίνδυνος $\Lambda(t)$ για κάθε άτομο κατά τη στιγμή του θανάτου ή της αποκοπής πρέπει να είναι σαν ένα δείγμα από $Exp(1)$. Με άλλα λόγια αν κάνουμε ένα γράφημα $\log(-\log S(t))$ Vs $\log(t)$, αν το μοντέλο ταιριάζει επαρκώς στα δεδομένα περιμένουμε να δούμε μια ευθεία γραμμή με κλίση ένα η οποία περνάει από την αρχή των αξόνων. Υπάρχουν, βεβαίως, και άλλοι γραφικοί έλεγχοι όπως τα Martingale residuals, τα Schoenfeld residuals, τα deviance residuals κλπ (Yishu et al, 2017).

3.1.4. Εκτίμηση των συντελεστών παλινδρόμησης του μοντέλου

Αφού καταλήξουμε στο τελικό βέλτιστο μοντέλο, το επόμενο βήμα είναι η εκτίμηση των συντελεστών παλινδρόμησης. Για τα μοντέλα Cox (για δεδομένα επιβίωσης) και λογαριθμιστικής παλινδρόμησης, συνήθως οι συντελεστές παλινδρόμησης εκτιμώνται με τη μέθοδο μέγιστης πιθανοφάνειας. Ορισμένες σύγχρονες τεχνικές έχουν αναπτυχθεί με στόχο τον περιορισμό της υπερπροσαρμογή (overfitting) του μοντέλου στα διαθέσιμα δεδομένα, όπως τεχνικές στατιστικής συρρίκνωσης (shrinkage), ο ελάχιστος απόλυτος συντελεστής συρρίκνωσης και επιλογής (LASSO) και η μέθοδος παλινδρόμησης κορυφογραμμής (ridge regression).

Εδώ θα επικεντρωθούμε σε δεδομένα επιβίωσης έτσι θα αναφερθούμε στην μέθοδο μέγιστης πιθανοφάνειας μέσω της Partial Likelihood. Στην περίπτωση αυτή κάνουμε συμπερασματολογία όπως ακριβώς θα κάναμε και αν είχαμε μια απλή πιθανοφάνεια από ένα μοντέλο λογιστικής παλινδρόμησης (Czepiel, 2002). Το μοντέλο Cox αναλογικών κινδύνων είναι το εξής:

$$\lambda(t;Z)=\lambda_0(t)\exp(\sum_{i=1}^N \beta_i Z_i)$$

$\lambda(t;Z)$: είναι ο κίνδυνος ενός ατόμου με χαρακτηριστικά Z να αναπτύξει την έκβαση που μελετάται, δεδομένης της χρονικής στιγμής t .

Z : είναι όλες οι επεξηγηματικές μεταβλητές

$\lambda_0(t)$: είναι ο κίνδυνος για το άτομο όπου όλες οι επεξηγηματικές του μεταβλητές παίρνουν την τιμή 0 (baseline hazard function).

N : ο αριθμός των επεξηγηματικών μεταβλητών.

Θέλουμε να εκτιμήσουμε τους N συντελεστές παλινδρόμησης, δηλαδή όλα τα β . Αυτό επιτυγχάνεται μεγιστοποιώντας την partial likelihood η οποία είναι η εξής:

$$L(\beta)=\prod_{i=1}^N \left[\frac{\exp(\beta Z_i)}{\sum_{j \in R(X_i)} \exp(\beta Z_j)} \right]^{\delta_i}$$

Όπου δ_i ο δείκτης αποτυχίας: αν $\delta_i=0$ αποκοπή και αν $\delta_i=1$ αποτυχία. Στη συνέχεια λογαριθμίζουμε την πιο πάνω πιθανοφάνεια και υπολογίζουμε την πρώτη παράγωγο. Οι εκτιμητές μέγιστης πιθανοφάνειας μπορούν να βρεθούν μηδενίζοντας την πρώτη παράγωγο.

Η μέθοδος συρρίκνωσης γίνεται μετά την εκτίμηση των συντελεστών παλινδρόμησης με την μέθοδο μέγιστης πιθανοφάνειας. Η ιδέα εδώ είναι να «συρρικνωθεί» η τιμή του συντελεστή παλινδρόμησης ώστε να περιοριστεί η υπερπροσαρμογή του μοντέλου. Οι εκτιμητές υπολογίζονται ως εξής: $s\beta_i$, με

$$s=(X^2 - df)/X^2$$

όπου df οι βαθμοί ελευθερίας του μοντέλου και X^2 η διαφορά των λογαρίθμων πιθανοφάνειας μεταξύ του μοντέλου με και χωρίς τους προγνωστικούς παράγοντες. Η απαιτούμενη συρρίκνωση αυξάνεται όσο μεγαλύτερος είναι ο αριθμός προγνωστικών παραγόντων ή όσο ο αριθμός του δείγματος είναι μικρότερος (*van Houwelingen et al, 1990*). Για παράδειγμα, στην μελέτη των Steyerberg & Vergouwe αναπτύχθηκαν προγνωστικά μοντέλα για την θνησιμότητα εντός 30 ημερών ατόμων με οξύ έμφραγμα του μυοκαρδίου σε ένα δείγμα από την βάση δεδομένων Gusto-I (*Steyerberg et al, 2014*), για το μοντέλο που εκτιμήθηκε σε 259 άτομα βρέθηκε

συντελεστής συρρίκνωσης 0.82. Οι συντελεστές παλινδρόμησης πρέπει να πολλαπλασιάζονται με αυτή την τιμή για να παρέχουν πιο αξιόπιστες προβλέψεις για νέους ασθενείς.

Η μέθοδος συρρίκνωσης μπορεί να χρησιμοποιηθεί και με μεθόδους bootstrap. Αρχικά επιλέγουμε ένα τυχαίο δείγμα bootstrap ίδιου μεγέθους με το αρχικό μας δείγμα, στη συνέχεια επιλέγουμε τους προγνωστικούς παράγοντες στο bootstrap δείγμα σύμφωνα με τη διαδικασία επιλογής και εκτίμησης των συντελεστών παλινδρόμησης. Έπειτα, υπολογίζουμε την αξία των γραμμικών παραγόντων για κάθε ασθενή στο δείγμα και τέλος εκτιμούμε την κλίση του γραμμικού παράγοντα χρησιμοποιώντας την έκβαση των ασθενών στο αρχικό δείγμα. Τα παραπάνω βήματα επαναλαμβάνονται πολλές φορές ούτως ώστε να εκτιμηθεί ο συντελεστής συρρίκνωσης (τιμές από 0 έως 1) ως η μέση τιμή των κλίσεων του τελευταίου βήματος.

Η LASSO είναι μια μέθοδος ανάλυσης παλινδρόμησης που επιλέγει και τακτοποιεί μεταβλητές προκειμένου να ενισχυθεί η ακρίβεια της πρόβλεψης και η ερμηνεία του στατιστικού μοντέλου. Είναι χρήσιμη όταν το μέγεθος του δείγματος είναι μικρό και όταν έχουμε πολλούς υποψήφιους προγνωστικούς παράγοντες. Έστω ένα δείγμα από N άτομα, κάθε άτομο αποτελείται από p μεταβλητές ($X=(X_1, X_2, \dots, X_p)$) και μία μόνο έκβαση (Y_i). Τότε η μέθοδος LASSO εκτιμάει τους συντελεστές παλινδρόμησης ελαχιστοποιώντας τον λογάριθμο πιθανοφάνειας σε $\sum |\beta| < t$, όπου t η συρρίκνωση του μοντέλου (Tibshirani, 1996). Η εκτίμηση των τελικών β γίνεται με την τιμή t , η οποία δίνει το μικρότερο τετραγωνικό σφάλμα σε μια διαδικασία διεπικύρωσης (cross validation).

Τέλος, η μέθοδος κορυφογραμμής (ridge regression) προσαρμόζει το μοντέλο κάτω από την προϋπόθεση ότι το άθροισμα των μέσων τετραγωνικών συντελεστών παλινδρόμησης δεν υπερβαίνουν ένα συγκεκριμένο όριο. Το όριο αυτό επιλέγεται έτσι ώστε να μεγιστοποιείται η προγνωστική ικανότητα του μοντέλου, χρησιμοποιώντας διασταυρωμένη επικύρωση. Στην επικύρωση αυτή, τα δεδομένα χωρίζονται σε k ομάδες. Το μοντέλο προσαρμόζεται στις $k-1$ ομάδες και επικυρώνεται στην ομάδα που παραλείφθηκε. Η διαδικασία αυτή επαναλαμβάνεται k φορές, και κάθε φορά παραλείπεται μία διαφορετική ομάδα (Cessie, 1992). Στα προγνωστικά μοντέλα που εφαρμόζονται σε ιατρικά δεδομένα προτιμάται η μέθοδος

κορυφογραμμής. Η μέθοδος LASSO προτιμάται όταν απαιτείται ένα πιο απλό μοντέλο με λιγότερους προγνωστικούς παράγοντες.

3.2. Αξιολόγηση του μοντέλου

Όπως έχουμε ήδη αναφέρει, τα μοντέλα πρόβλεψης είναι πολύ σημαντικά για την κλινική ιατρική, αφού μπορούν να συνδυάσουν πολλαπλούς προγνωστικούς παράγοντες για να δώσουν απόλυτες εκτιμήσεις κινδύνου για κάθε ασθενή. Ως εκ τούτου, αντιλαμβανόμαστε πόσο σημαντικό είναι να καθοριστεί η ποιότητα των μοντέλων αυτών. Για την αξιολόγηση των μοντέλων χρησιμοποιούνται διάφορα μέτρα επιδόσεων. Τα κύρια κριτήρια για την αξιολόγηση ενός προγνωστικού μοντέλου, στα οποία θα αναφερθούμε εκτενώς πιο κάτω, είναι: η βαθμονόμηση (calibration) του μοντέλου, η ικανότητα διαχωριστικότητας (discrimination) και η κλινική του χρησιμότητα (clinical usefulness) (Moons *et al*, 2012).

3.2.1. Βαθμονόμηση του μοντέλου (Calibration)

Η βαθμονόμηση του μοντέλου έχει να κάνει με τη συμφωνία μεταξύ των εκτιμήσεων του μοντέλου και των παρατηρούμενων αποτελεσμάτων (Steyerberg *et al*, 2010). Για την αξιολόγηση της βαθμονόμησης χρησιμοποιούνται τρεις προσεγγίσεις: γραφικά, βαθμονόμηση in-the-large (κλίση A) και μέσω της κλίσης της βαθμονόμησης (Steyerberg *et al*, 2009).

Κατά τη γραφική αξιολόγηση της βαθμονόμησης, κάνουμε μια γραφική παράσταση με τις προβλέψεις στον χ άξονα και τα παρατηρούμενα αποτελέσματα στον ψ άξονα. Οι παρατηρούμενες τιμές στον ψ άξονα είναι 0 ή 1 (dead/alive), ενώ οι προβλέψεις στον χ άξονα κυμαίνονται μεταξύ 0 και 100%. Όταν οι παρατηρούμενες είναι ίσες με τις προβλεπόμενες (δηλαδή όταν έχουμε καλή πρόβλεψη), οι τιμές αυτές πέφτουν πάνω στις 45 μοίρες της γραμμής που σχηματίζεται στην γραφική παράσταση που αναφέραμε. Επίσης, για την απεικόνιση της συμφωνίας μεταξύ των παρατηρούμενων και των προβλεπόμενων τιμών μπορούν να χρησιμοποιηθούν

τεχνικές εξομάλυνσης (smoothing) για να απεικονιστεί η συσχέτιση (Austin et al, 2014).

Η παλαιότερη ιδέα βαθμονόμησης προτάθηκε από τον Cox το 1958 (Cox DR, 1958), η οποία λέει ότι οι τέλειες προβλέψεις θα πρέπει να είναι στην ιδανική γραμμή, που περιγράφεται με μια σταθερά άλφα (A) στο σημείο μηδέν και κλίση βήτα (B) ίση με 1. Η ατελής βαθμονόμηση μπορεί να χαρακτηριστεί από αποκλίσεις από αυτές τις ιδανικές τιμές. Το σημείο τομής A σχετίζεται με την βαθμονόμηση in-the-large, η οποία συγκρίνει τον μέσο όρο όλων των προβλεπόμενων κινδύνων με τον μέσο παρατηρούμενο κίνδυνο. Επομένως, η παράμετρος υποδεικνύει την έκταση που οι προβλέψεις είναι συστηματικά πολύ χαμηλές (υποεκτίμηση κινδύνου) ή υπερβολικά υψηλές (υπερεκτίμηση κινδύνου). Έτσι, όταν το A παίρνει την τιμή 0 φαίνεται ότι δεν υπάρχει κάποια διαφορά μεταξύ παρατηρούμενων και προβλεπόμενων τιμών, συνεπώς έχουμε καλή βαθμονόμηση. Κατά την ανάπτυξη του μοντέλου, ο παρατηρούμενος κίνδυνος και ο μέσος όρος του προβλεπόμενου κινδύνου είναι ίσα για τα μοντέλα παλινδρόμησης και η αξιολόγηση της βαθμονόμησης in-the-large δεν έχει νόημα. Αντιθέτως, όταν εξετάζουμε τη βαθμονόμηση σε εξωτερικά δεδομένα, υπάρχει διαφορά μεταξύ της μέσης τιμής των εκτιμήσεων και της μέσης τιμής των παρατηρούμενων νέων τιμών, και η διαφορά αυτή ονομάζεται calibration in the large. Κατά την εξωτερική επικύρωση συχνά αντιμετωπίζουμε προβλήματα βαθμονόμησης, όπως για παράδειγμα στο μοντέλο Framingham σε ομάδες διάφορων εθνικοτήτων (D'Agostino et al, 2001). Στην περίπτωση όπου γίνεται χρήση μοντέλου Cox στην εξωτερική επικύρωση δεν υπολογίζεται η σταθερά (intercept), έτσι μπορούμε να συγκρίνουμε τον μέσο όρο των προβλεπόμενων κινδύνων για ένα συγκεκριμένο χρονικό σημείο, πχ 5 ή 10 έτη, με την εκτίμηση από καμπύλη Kaplan-Meier (Royston et al, 2013). Επίσης, άλλοι τύποι μοντέλων όπως τα μοντέλα παλινδρόμησης Weibull, μπορούν να χρησιμοποιηθούν για την εκτίμηση βαθμονόμησης για μοντέλα επιβίωσης (van Houwelingen, et al, 1995).

Κατά την αξιολόγηση της βαθμονόμησης του μοντέλου με την κλίση βαθμονόμησης (calibration slope), εξετάζεται κατά πόσο η κλίση β απέχει από τη μονάδα. Στην περίπτωση όπου έχουμε δεδομένα επιβίωσης, η κλίση βαθμονόμησης είναι απλώς η κλίση των παρατηρούμενων εκβάσεων επιβίωσης στον προβλεπόμενο προγνωστικό δείκτη. Υπολογίζεται με την τοποθέτηση ενός μοντέλου Cox σε νέες

εκβάσεις επιβίωσης με τον προβλεπόμενο προγνωστικό δείκτη, \hat{h} , ως μοναδικός παράγοντας πρόβλεψης στο μοντέλο (Rahman et al, 2017):

$$h(t/x) = h_o(t)exp(\beta\hat{h})$$

Οι τιμές του β κοντά στο 1 δείχνουν ότι το μοντέλο πρόβλεψης είναι καλά βαθμονομημένο, ενώ αν $\beta < 1$ υποδηλώνει υπερπροσαρμογή.

Στην εσωτερική επικύρωση του μοντέλου, το β δείχνει αν πρέπει να γίνει συρρίκνωση στο δείγμα και στην εξωτερική επικύρωση την υπερπροσαρμογή στα δεδομένα και την διαφορά στις επιδράσεις των προγνωστικών παραγόντων (Steyerberg et al, 2014). Η κλίση βαθμονόμησης β είναι συνήθως μικρότερη από 1 εάν ένα μοντέλο έχει αναπτυχθεί σε ένα σχετικά μικρό σύνολο δεδομένων. Ένα τέτοιο εύρημα υποδηλώνει ότι οι προβλέψεις ήταν υπερβολικές, δηλαδή η χαμηλή πρόβλεψη ήταν πολύ χαμηλή και οι υψηλές προβλέψεις ήταν πολύ υψηλές. Στην μελέτη των Steyerberg και Vergouwe, η οποία αναφέρθηκε και πιο πάνω (Steyerberg et al, 2014), αναφέρεται ότι για το μοντέλο που βασίστηκε σε 259 ασθενής (μικρό δείγμα), η κλίση ήταν $0.70 < 1$, ενώ για το μοντέλο που βασίστηκε σε 23034 ασθενής (μεγάλο δείγμα) το β ήταν κοντά στο 1. Για το μοντέλο με το μικρό δείγμα, η τιμή του β θα μπορούσε να σημαίνει ότι οι συντελεστές του μοντέλου πρέπει να συρρικνωθούν (Van Houwelingen et al, 1990).

3.2.2. Ικανότητα διαχωριστικότητας του μοντέλου-Discrimination

Η διαχωριστικότητα αναφέρεται στην ικανότητα του μοντέλου να διαχωρίζει αν το άτομο έχει αναπτύξει την έκβαση ή όχι (Royston et al, 2010). Η διαχωριστικότητα δεν συνδέεται με την βαθμονόμηση του μοντέλου, αφού ένα μοντέλο που προβλέπει τον πραγματικό κίνδυνο για όλα τα άτομα δείχνει τέλεια βαθμονόμηση, αλλά είναι άχρηστο αφού δεν διαχωρίζει τους ασθενείς με έκβαση και τους ασθενείς χωρίς. Με άλλα λόγια, αν ένα μοντέλο έχει τέλεια βαθμονόμηση δεν σημαίνει ότι έχει διαχωριστική ικανότητα. Τα μέτρα αξιολόγησης της διαχωριστικής ικανότητας του μοντέλου είναι η καμπύλη ROC για δίτιμη έκβαση και το c-statistic για δεδομένα

επιβίωσης, η ευαισθησία (sensitivity) και η ειδικότητα (specificity) (Dreiseitla et al, 2002).

3.2.2.1. Ευαισθησία (Sensitivity) και Ειδικότητα (Specificity)

Σε ένα διαγνωστικό τεστ η ευαισθησία είναι η πιθανότητα το τεστ να είναι θετικό δεδομένου ότι το άτομο έχει αναπτύξει την έκβαση. Ειδικότητα είναι η πιθανότητα το τεστ να είναι αρνητικό δεδομένου ότι το άτομο δεν έχει αναπτύξει την έκβαση. Για όλες τις λανθασμένες διαγνώσεις θα ισχύει λοιπόν,

Πιθανότητα ψευδώς αρνητικού (false negative) = 1- ευαισθησία

Πιθανότητα ψευδώς θετικού (false positive) = 1- ειδικότητα

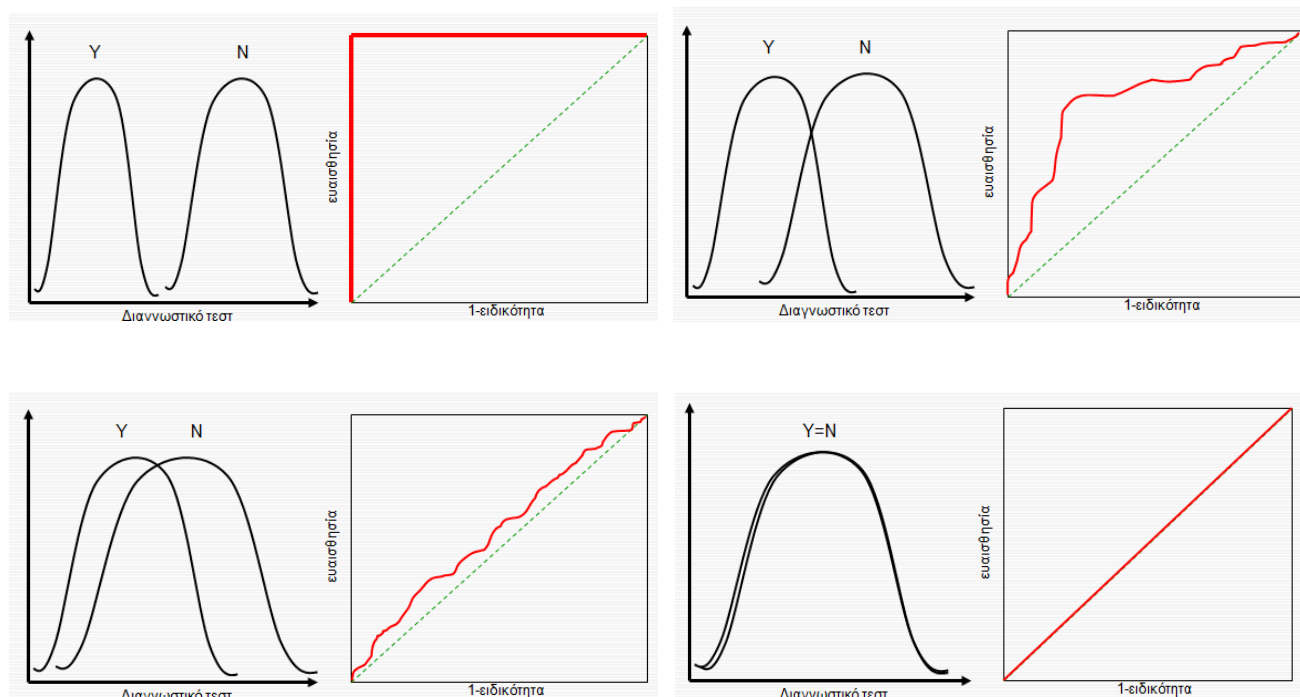
Η ειδικότητα και η ευαισθησία παίρνουν τιμές από 0 μέχρι 100% και μπορούν να αναπαρασταθούν γραφικά μέσω της καμπύλης ROC. Για να αποφασίσουμε εάν όντως ένας ασθενής έχει την ασθένεια υπάρχει ένα χρυσό πρότυπο (gold standard) που θεωρείται η σημερινή προτιμώμενη μέθοδος διάγνωσης μιας ασθένειας. Το πρότυπο αυτό είναι διαφορετικό για κάθε ασθένεια, για παράδειγμα εάν εξετάζουμε το primary angle closure (PAC) το χρυσό πρότυπο είναι η γωνιοσκόπηση (gonioscopy) (Rajul et al, 2008).

3.2.2.2. Καμπύλη ROC

Οι ROC (Receiver Operating Characteristic) καμπύλες (ή καμπύλες λειτουργικού χαρακτηριστικού δέκτη), αποτελούν χρήσιμη τεχνική για την οργάνωση, την επιλογή και την απεικόνιση ταξινομητών με βάση τη γραφική τους παράσταση. Είναι μια οπτική απεικόνιση της διακριτικής ικανότητας μιας μεθόδου και χρησιμοποιούνται ευρέως στην διαγνωστική ιατρική. Στον άξονα χ μιας καμπύλης ROC έχουμε την 1-ειδικότητα ενώ στον άξονα ψ την ευαισθησία. Μια τέτοια καμπύλη δείχνει πως αλλάζει η ευαισθησία όταν αλλάζει η ειδικότητα. Έχουμε μια τέλεια δοκιμασία όταν σχηματίζεται ορθή γωνία δηλ. ευαισθησία 100% χωρίς απώλεια στην ειδικότητα και μια κακή δοκιμασία όταν σχηματίζεται διαγώνιος γραμμή, δηλαδή ευαισθησία 50%

και ψευδώς θετικά 50%. Η περιοχή κάτω από την καμπύλη ROC ονομάζεται AUC (area under the curve) και είναι ένα μέτρο διαγνωστικής ακρίβειας. Αν ένα μοντέλο είναι τέλειο θα έχουμε $AUC=1$, ενώ αν δεν έχει καθόλου διαγνωστική ικανότητα θα έχουμε $AUC=0.5$ (Hoo et al, 2017).

Εικόνα 3.1. Σχέση καμπυλών ROC και διαγνωστικής αξίας



3.2.2.3: Το στατιστικό c (c-statistic ή c-index)

Η πιο συνηθισμένη προσέγγιση για την αξιολόγηση της προγνωστικής απόδοσης του συνόλου είναι ο υπολογισμός της στατιστικής C για τα δεδομένα επιβίωσης, που ονομάζεται «Harrell's C» (Harrell et al., 1982). Μια τιμή $C = 0.5$ αντιστοιχεί σε έναν μη-πληροφοριακό κανόνα πρόβλεψης, ενώ ο $C = 1$ αντιστοιχεί στην τέλεια συσχέτιση, υποδηλώνοντας ότι το C του Harrell είναι ένας εύκολος στην ερμηνεία συντελεστής που αντιπροσωπεύει το σύνολο των παρατηρούμενων χρόνων επιβίωσης

(Schmid και Potapov , 2012). Ονομάζεται c-index διότι είναι ένα μέτρο συμμόρφωσης – concordance measure.

Το C ορίζεται ως εξής, ((Harrell et al., 1982):

$$C = \frac{\sum_{i,j} I(\hat{T}_i > \hat{T}_j) I(n_j > n_i) \Delta_j}{\sum_{i,j} I(\hat{T}_i > \hat{T}_j) \Delta_j}$$

όπου οι δείκτες i και j αναφέρονται σε ζεύγη παρατηρήσεων στο δείγμα. Το στατιστικό C είναι ο αριθμός των αντίστοιχων παρατηρούμενων ζευγών που διαιρούνται με τον αριθμό των συγκρίσιμων ζευγών. Πολλαπλασιάζοντας με τον παράγοντα Δ_j απορρίπτονται ζεύγη παρατηρήσεων που δεν είναι συγκρίσιμα επειδή ο μικρότερος χρόνος επιβίωσης αποκόπτεται, δηλ. $\Delta_j = 0$. Το Harrell C έχει σχεδιαστεί για να εκτιμήσει την πιθανότητα συμμόρφωσης (concordance probability),

$$P(n_j > n_i / T_i > T_j)$$

η οποία συγκρίνει την κατάταξη δύο ανεξάρτητων ζευγών χρόνων επιβίωσης T_i, T_j και προβλέψεων n_i, n_j . Η πιθανότητα συμμόρφωσης αξιολογεί εάν οι μεγάλες τιμές του n_i σχετίζονται με μικρές τιμές T_i και το αντίστροφο. Το Harrell C μπορεί επίσης να ερμηνευτεί ως συνοπτικό μέτρο της περιοχής κάτω από χρονικά εξαρτώμενες καμπύλες ROC (Heagerty and Zheng, 2005, Schmid and Potapov, 2012). Μια τιμή $C = 0,5$ αντιστοιχεί σε έναν μη ενημερωτικό κανόνα πρόβλεψης, ενώ το $C = 1$ αντιστοιχεί στην τέλεια συσχέτιση. Υποστηρίχθηκε λοιπόν ότι το C του Harrell είναι ένας εύκολος-ερμηνευτικός συντελεστής που αντιπροσωπεύει το σύνολο των παρατηρούμενων χρόνων επιβίωσης.

3.2.3. Κλινική χρησιμότητα του μοντέλου (clinical usefulness)

Η ικανότητα διαχωριστικότητας του μοντέλου και η βαθμονόμηση αποτελούν σημαντικές πτυχές ενός μοντέλου πρόβλεψης και εξετάζουν το πλήρες φάσμα των προβλεπόμενων κινδύνων. Ωστόσο, αυτές οι πτυχές δεν αξιολογούν την κλινική χρησιμότητα του μοντέλου, δηλαδή την ικανότητα λήψης καλύτερων αποφάσεων. Σε μια ανάλυση αποφάσεων εναλλακτικών διαγνωστικών ή προγνωστικών μοντέλων το βέλτιστο μοντέλο είναι αυτό που μεγιστοποιεί την έκβαση ενδιαφέροντος. Εάν ένα

μοντέλο στοχεύει στην καθοδήγηση των αποφάσεων θεραπείας, απαιτείται να ταξινομήσει τους ασθενείς με βάση τον κίνδυνό τους, δηλαδή σε ασθενείς χαμηλού κινδύνου και σε ασθενείς υψηλού κινδύνου (όπου ενδείκνυται θεραπεία).

Για να γίνει αυτός ο διαχωρισμός χρειαζόμαστε ένα σημείο απόφασης, το οποίο συχνά είναι δύσκολο να προσδιοριστεί λόγω ελλειπουσών πληροφοριών οφέλους-ζημιάς. Επίσης, ορισμένοι ασθενείς μπορεί να είναι διατεθειμένοι να αναλάβουν υψηλότερο κίνδυνο για πιθανό όφελος από άλλους ασθενείς. Για τους λόγους αυτούς, συνιστάται να εξετάζεται ένα εύρος τιμών κατά την ποσοτικοποίηση της κλινικής χρησιμότητας ενός μοντέλου πρόβλεψης (*Vickers et al, 2006*). Μόλις εφαρμοστεί ένα μέτρο απόφασης για την ταξινόμηση των ασθενών ως χαμηλού και υψηλού κινδύνου, η ευαισθησία και η ειδικότητα χρησιμοποιούνται συχνά ως μέτρα για την κλινική χρησιμότητα. Η βέλτιστη ισορροπία μεταξύ αυτών μπορεί να βρεθεί με την εξέταση ζημιάς-οφέλους της θεραπείας σε συνδυασμό με την συχνότητα εμφάνισης της έκβασης. Το άθροισμα της ευαισθησίας και της ειδικότητας μπορεί να χρησιμοποιηθεί μόνο ως αφηρημένος συνοπτικός δείκτης χρησιμότητας, δεδομένου ότι ένα τέτοιο άθροισμα αγνοεί το σχετικό βάρος των αληθινών θετικών (ευαισθησία) και των ψευδών θετικών (1-ειδικότητα) (*Greenland, 2008*).

Τα πιο πρόσφατα συνοπτικά μέτρα περιλαμβάνουν το καθαρό κέρδος (net benefit, NB). Το μέτρο αυτό συνάδει με την χρήση ενός βέλτιστου σημείου απόφασης για την ταξινόμηση των ασθενών. Το καθαρό κέρδος υπολογίζεται με τον πιο κάτω τύπο:

$$New\ Benefit = \frac{\text{αληθώς θετικά}}{n} - \frac{\text{ψευδώς θετικά}}{n} \left(\frac{pt}{1-pt} \right)$$

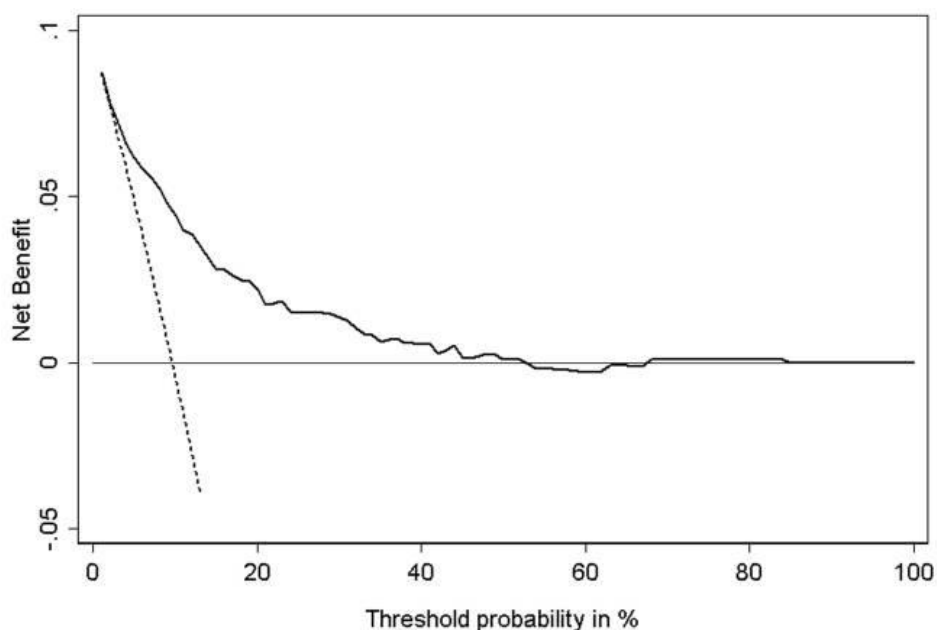
Όπου n ο αριθμός των ασθενών και pt το τελικό σημείο απόφασης (*Vickers et al, 2006*). Εν ολίγοις, αφαιρούμε το ποσοστό όλων των ασθενών που είναι ψευδώς θετικοί από την αναλογία που είναι αληθινά θετικοί, σταθμίζοντας τη σχετική βλάβη ενός ψευδώς θετικού και ενός ψευδώς αρνητικού αποτελέσματος. Για παράδειγμα, στο GUSTO-I η συνολική θνησιμότητα 7% σε 30 ημέρες συνεπάγεται ότι η μέγιστη επιτρεπτή τιμή είναι 7%. Αυτό επιτυγχάνεται με εάν χρησιμοποιήσουμε ένα βέλτιστο όριο 0%. Τότε όλοι οι ασθενείς είναι υποψήφιοι για θεραπεία αφού δεν υπονοούμε

κάποια βλάβη της θεραπείας. Αντίθετα, εάν εκτιμούμε ότι η θεραπεία επιφέρει κάποια βλάβη τότε το βέλτιστο όριο απόφασης είναι $>0\%$. Διαπιστώθηκε ότι η λήψη αποφάσεων με βάση τις προβλέψεις κινδύνου από ένα μοντέλο θα έδινε μια ελαφρώς υψηλότερη τιμή από το να έπαιρναν όλοι θεραπεία. Για παράδειγμα, για ένα όριο 2% , έχουμε 1225 αληθώς θετικές ταξινομήσεις (υποψήφιοι για θεραπεία που πέθαναν), αλλά και 11192 ψευδώς θετικές ταξινομήσεις (υποψήφιοι για θεραπεία που επέζησαν) μεταξύ των 17796 μη αμερικανών ασθενών. Το NB υπολογίζεται ως εξής: $(1225 - 2/98 \times 11192) / 17796 = 5.6\%$. Αυτό είναι ελαφρώς καλύτερο από την θεραπεία όλων, όπου $NB = (1286 - 2/98 \times 16510) / 17796 = 5.3\%$, όπου 1286 πέθαναν και 16510 επιβίωσαν. Αυτή η διαφορά του 0.3% σημαίνει ότι για κάθε 1000 ασθενείς στους οποίους εφαρμόζουμε το μοντέλο πρόβλεψης, εντοπίζονται 3 επιπλέον αληθώς θετικές περιπτώσεις, χωρίς να αυξάνονται οι ψευδώς θετικές περιπτώσεις.

Για τη δημιουργία μιας καμπύλης απόφασης πρέπει να ακολουθηθούν τα εξής 8 βήματα: (Vickers et al, 2006)

- 1) Επιλέγουμε μια τιμή για το pt
- 2) Υπολογίζουμε τον αριθμό των αληθών και ψευδών θετικών αποτελεσμάτων χρησιμοποιώντας pt ως σημείο διακοπής (cut-point) για τον προσδιορισμό ενός θετικού ή αρνητικού αποτελέσματος.
- 3) Υπολογίζουμε το καθαρό όφελος NB του μοντέλου πρόβλεψης
- 4) Επιλέγουμε ένα εύρος τιμών για το pt και επαναλαμβάνουμε τα βήματα 2-3
- 5) Σχεδιάζουμε το καθαρό όφελος στον ψ άξονα έναντι του pt στον χ άξονα
- 6) Επαναλαμβάνουμε τα βήματα 1-5 για κάθε υπό εξέταση μοντέλο
- 7) Επαναλαμβάνουμε τα παραπάνω βήματα για όλες τις στρατηγικές θεραπειών
- 8) Σχεδιάζουμε μια ευθεία γραμμή παράλληλη προς τον άξονα x στο $y = 0$ που αντιπροσωπεύει το καθαρό όφελος που σχετίζεται με την υπόθεση ότι όλοι οι ασθενείς είναι αρνητικοί.

Εικόνα 3.2: καμπύλη απόφασης που προβλέπει τη βιοψία για προστάτη (*Vickers et al, 2006*)



Στο παραπάνω γράφημα η λεπτή γκρι γραμμή είναι το καθαρό κέρδος για λήψη βιοψίας όλων των ατόμων. Η λεπτή μαύρη γραμμή είναι το καθαρό κέρδος λήψης βιοψίας των ατόμων με βάση το στατιστικό μοντέλο. Η παχιά μαύρη γραμμή είναι το καθαρό κέρδος λήψης βιοψίας από κανένα άτομο (*Vickers et al, 2006*). Για να είναι κλινικά χρήσιμο το μοντέλο θα πρέπει οι μισές προβλέψεις να είναι πάνω από το σημείο απόφασης και οι μισές κάτω, διαφορετικά το μοντέλο δεν είναι κλινικά χρήσιμο.

3.3. Επικύρωση του μοντέλου-Validation

Ένα από τα σημαντικότερα ζητήματα στην ανάπτυξη ενός μοντέλου είναι η επικύρωσή του, αφού από αυτό εξαρτάται η πορεία της νόσου ενός ασθενή και πολλές φορές η θεραπεία του. Η επικύρωση ενός μοντέλου διαχωρίζεται σε εσωτερική (*internal*) και εξωτερική (*external*). Οι δύο αυτές επικυρώσεις θα αναπτυχθούν παρακάτω.

3.3.1. Εσωτερική επικύρωση (internal validation)

Η απόδοση ενός προγνωστικού μοντέλου υπερεκτιμάται όταν προσδιορίζεται απλώς στο δείγμα των ατόμων που χρησιμοποιήθηκαν για την κατασκευή του μοντέλου. Υπάρχουν αρκετές μέθοδοι εσωτερικής επικύρωσης που αποσκοπούν στην παροχή ακριβέστερης εκτίμησης της απόδοσης του μοντέλου σε νέα άτομα (Steyerberg *et al*, 2001). Κατά την εσωτερική επικύρωση δεν χρησιμοποιούνται άλλα δεδομένα, εκτός από τα αρχικά δεδομένα που χρησιμοποιήθηκαν για την κατασκευή του μοντέλου. Γενικά, υπάρχουν πολλών ειδών εσωτερικές επικυρώσεις, μερικές από αυτές είναι η μέθοδος bootstrap και η διεπικύρωση (cross-validation).

3.3.1.1. Μέθοδος bootstrap

Η μέθοδος bootstrap βασίζεται στη δημιουργία B νέων δειγμάτων με ίδιο μέγεθος με το αρχικό δείγμα. Τα δείγματα αυτά δημιουργούνται με δειγματοληψία με επανάθεση από το αρχικό δείγμα. Η εκτιμήτρια της παραμέτρου που μας ενδιαφέρει υπολογίζεται για το κάθε ένα από τα B δείγματα bootstrap και παράγουν την κατανομή bootstrap της εκτιμήτριας. Βασική προϋπόθεση είναι ότι οι παρατηρήσεις του αρχικού δείγματος απεικονίζουν όλον τον πληθυσμό. Το bootstrap είναι ένα εργαλείο για τη διεξαγωγή στατιστικών συμπερασμάτων όταν αμφισβητούνται οι τυποποιημένες παραμετρικές υποθέσεις. Για παράδειγμα, είναι γνωστό ότι οι δοκιμές σημαντικότητας για τους συντελεστές παλινδρόμησης μπορεί να είναι παραπλανητικές όταν τα σφάλματα δεν κατανέμονται κανονικά (Efron and Tibshirani, 1993).

Κατά την επικύρωση ενός μοντέλου με τη μέθοδο bootstrap, για κάθε δείγμα bootstrap δημιουργείται ένα προγνωστικό μοντέλο το οποίο εκτιμάται και στο αρχικό δείγμα αλλά και στο bootstrap δείγμα. Η πιο βασική μορφή bootstrap περιέχει τα ακόλουθα βήματα:

- 1) Παίρνουμε ένα δείγμα μεγέθους N από έναν πληθυσμό
- 2) Παίρνουμε B δείγματα μεγέθους N με αντικατάσταση από το αρχικό δείγμα
- 3) Υπολογίζουμε την εκτίμηση θ για κάθε δείγμα B

4) Κατασκευάζουμε μια κατανομή από τις εκτιμήσεις που προέκυψαν από τα B δείγματα.

Η μέθοδος bootstrap έχει αρκετά πλεονεκτήματα, όπως το ότι δίνει σταθερή εκτίμηση για την απόδοση του μοντέλου και το ότι λειτουργεί καλά σε πολυάριθμα δεδομένα όπου ο αριθμός των μεταβλητών είναι μεγαλύτερος από τον αριθμό του δείγματος (*Schumacher et al, 2007*). Για να επιτευχθούν σταθερά αποτελέσματα, η διαδικασία πρέπει να επαναληφθεί, συνήθως, τουλάχιστον 100 φορές. Το κύριο μειονέκτημά της είναι ότι μπορεί να χρησιμοποιηθεί μόνο με αυτοματοποιημένες μεθόδους μοντελοποίησης, όπως οι μέθοδοι βηματικής επιλογής που αναφέρθηκαν πιο πάνω.

3.3.1.2. Μέθοδος διεπικύρωσης (cross-validation)

Η μέθοδος διεπικύρωσης συχνά ονομάζεται εκτίμηση περιστροφής (rotation estimation). Χρησιμοποιείται κυρίως σε περιπτώσεις όπου ο στόχος είναι πρόβλεψη και κάποιος θέλει να υπολογίσει πόσο ακριβές είναι ένα μοντέλο πρόβλεψης. Η μέθοδος διεπικύρωσης είναι να ελέγξει την ικανότητα του μοντέλου να προβλέψει νέα δεδομένα που δεν χρησιμοποιήθηκαν για την εκτίμησή του, προκειμένου να επισημάνει προβλήματα όπως η υπερπροσαρμογή (overfitting) ή η μεροληψία (bias) και να δώσει μια εικόνα για το πώς το μοντέλο θα γενικευθεί σε ένα ανεξάρτητο σύνολο δεδομένων. Η επικύρωση γίνεται σε ένα μέρος του δείγματος το οποίο δεν χρησιμοποιήθηκε για την ανάπτυξη του μοντέλου. Το δείγμα χωρίζεται τυχαία σε δύο ομάδες. Η διεπικύρωση ακολουθεί τον τρόπο επικύρωσης με τη μέθοδο διαχωρισμού δείγματος (split sample method) όπου ο προσδιορισμός του μοντέλου γίνεται με την μία ομάδα δεδομένων και η επικύρωση με την άλλη ομάδα δεδομένων. Τις περισσότερες φορές οι διαχωρισμοί του δείγματος γίνονται 50% - 50%, ωστόσο εάν τα δείγματα διαχωρίζονται εντελώς τυχαία είναι πιθανόν να υπάρξουν ανισοροπίες στην κατανομή της έκβασης και των προγνωστικών παραγόντων. Η διαφορά στην μέθοδο διεπικύρωσης είναι ότι η διαδικασία αυτή επαναλαμβάνεται για διαδοχικά μέρη των ασθενών. Για παράδειγμα, τα δεδομένα μπορεί να διαχωριστούν σε 10 ομάδες όπου η κάθε ομάδα περιέχει το 1/10 των ασθενών. Τα 9/10 χρησιμοποιούνται για το σχεδιασμό του μοντέλου και το 1/10 για την επικύρωση του μοντέλου και η διαδικασία αυτή να επαναληφθεί 10 φορές. Με αυτόν τον τρόπο, όλοι οι ασθενείς

έχουν χρησιμοποιηθεί μία φορά για να ελεγχθεί το μοντέλο και απόδοση του μοντέλου εκτιμάται ως η μέση τιμή όλων των εκτιμήσεων (*Rubin, 1987*)

3.3.2. Εξωτερική επικύρωση (external validation)

Η εξωτερική επικύρωση είναι η επικύρωση των αποτελεσμάτων-συμπερασμάτων μιας επιστημονικής μελέτης εκτός του πλαισίου αυτής της μελέτης. Δηλαδή, είναι ο βαθμός στον οποίο τα αποτελέσματα μιας μελέτης μπορούν να γενικευθούν σε άλλες καταστάσεις, ανθρώπους, χρόνους κλπ (*Aronson et al, 2007*). Αντίθετα, η εσωτερική επικύρωση είναι η επικύρωση των συμπερασμάτων που συνάγονται στο πλαίσιο μιας συγκεκριμένης μελέτης. Η γενίκευση των αποτελεσμάτων είναι σχεδόν πάντα ένας στόχος στην έρευνα, έτσι η εξωτερική επικύρωση αποτελεί σημαντική ιδιότητα οποιασδήποτε μελέτης. Η μαθηματική ανάλυση της εξωτερικής επικύρωσης αφορά τον προσδιορισμό του κατά πόσο είναι εφικτή η γενίκευση των αποτελεσμάτων μιας μελέτης σε ετερογενείς πληθυσμούς και την εκπόνηση στατιστικών μεθόδων που παράγουν έγκυρες γενικεύσεις (*Pearl et al, 2014*). Η εξωτερική επικύρωση μπορεί να γίνει με διάφορες προσεγγίσεις. Εδώ θα επικεντρωθούμε σε δύο από αυτές, την χρονική επικύρωση (temporal) και την γεωγραφική επικύρωση (geographical).

3.3.2.1. Χρονική επικύρωση (temporal validation)

Η χρονική επικύρωση πραγματοποιείται σε δεδομένα που προέρχονται από το ίδιο μέρος όμως σε μια άλλη χρονική στιγμή, δηλαδή γίνεται διαχωρισμός ενός συνόλου δεδομένων που υπάρχει ήδη (*Moons et al, 2012*). Για παράδειγμα, τα δεδομένα αυτά μπορεί να προέρχονται από ένα αρχείο νοσοκομείου και ο διαχωρισμός να γίνει με βάση τη θεραπεία. Δηλαδή, το ένα δείγμα να είναι τα άτομα που έλαβαν νωρίτερα θεραπεία και να αναπτυχθεί το μοντέλο με βάση αυτό το δείγμα και το άλλο δείγμα να είναι τα άτομα που έλαβαν πιο πρόσφατα θεραπεία και να γίνει η επικύρωση του μοντέλου με βάση το δεύτερο δείγμα. Με άλλα λόγια, η χρονική επικύρωση αναφέρεται στην εγκυρότητα των αποτελεσμάτων σε σχέση με την εξέλιξη του χρόνου.

3.3.2.2. Γεωγραφική επικύρωση (geographic validation)

Με τη γεωγραφική επικύρωση μπορούμε να γενικεύσουμε τα ευρήματα του προγνωστικού μοντέλου σε άλλες χώρες ή τοποθεσίες, κάτι το οποίο δεν είναι εφικτό με την χρονική επικύρωση. Η γεωγραφική επικύρωση μπορεί να γίνει διαχωρίζοντας ένα υπάρχον δείγμα από πχ άλλη χώρα, το οποίο δεν χρησιμοποιήθηκε κατά την ανάπτυξη του μοντέλου πρόβλεψης. Γενικά, η γεωγραφική επικύρωση θεωρείται πιο αξιόπιστη από την χρονική, αφού τα δεδομένα σε μια γεωγραφική επικύρωση διαφέρουν αρκετά μεταξύ τους, κυρίως σε εθνικότητα, τρόπο ζωής, διατροφή κλπ. Επίσης, στη χρονική επικύρωση έχουμε ίδια κριτήρια εισόδου/αποκλεισμού, ίδιο ορισμός έκβασης και ίδιες μετρήσεις, ενώ στην γεωγραφική επικύρωση όλα αυτά είναι διαφορετικά συγκριτικά με την έρευνα σχεδιασμού του μοντέλου.

Κεφάλαιο 4

Δημιουργία κλίμακας (risk score) για την εκτίμηση του κινδύνου από ένα προγνωστικό μοντέλο παρουσία ανταγωνιστικών κινδύνων

Η πρόβλεψη της εμφάνισης ενός ανεπιθύμητου γεγονότος με την πάροδο του χρόνου είναι ένα σημαντικό ζήτημα στην κλινική ιατρική. Τα μοντέλα πρόβλεψης και τα συναφή συστήματα βαθμολόγησης κινδύνου (points-based risk-scoring systems) ή κλίμακες εκτίμησης κινδύνου είναι δημοφιλείς στατιστικές μέθοδοι για την περιγραφή της σχέσης μεταξύ ενός πολυμεταβλητού συνόλου παραγόντων κινδύνου των ασθενών και του κινδύνου εμφάνισης ενός ανεπιθύμητου γεγονότος (outcome). Οι κλίμακες εκτίμησης κινδύνου είναι δημοφιλείς μεταξύ των κλινικών ιατρών λόγω της εύκολης χρήσης τους καθώς επιτρέπουν την ταχεία αξιολόγηση του κινδύνου του ασθενούς χωρίς την χρήση ηλεκτρονικού υπολογιστή ή άλλων ηλεκτρονικών συσκευών. Η χρήση τέτοιων κλιμάκων επιτρέπει την λήψη κλινικών αποφάσεων βάσει τεκμηριωμένων στοιχείων και υπάρχει αυξανόμενο ενδιαφέρον για την θνησιμότητα που οφείλεται σε ειδικά αίτια και τις μη θανατηφόρες εκβάσεις. Ωστόσο όταν εξετάζουμε αυτού του είδους εκβάσεων πρέπει να λάβουμε υπόψη τους ανταγωνιστικούς κινδύνους των οποίων η εμφάνιση δεν επιτρέπει την εμφάνιση του γεγονότος που μας ενδιαφέρει. Στο παρόν κεφάλαιο θα περιγραφεί η διαδικασία ανάπτυξης κλίμακας για την εκτίμηση κινδύνου απουσία και παρουσία ανταγωνιστικών κινδύνων.

4.1. Κλίμακα εκτίμησης κινδύνου απουσία ανταγωνιστικών γεγονότων

Εν απουσία ανταγωνιστικών γεγονότων, υπάρχει μια άμεση σχέση μεταξύ των συντελεστών παλινδρόμησης από ένα μοντέλο αναλογικών κινδύνων Cox και της συχνότητας εμφάνισης του γεγονότος ενδιαφέροντος:

$$S(t) = S_0(t) \exp(\beta Z)$$

όπου $S(t)$ είναι η συνάρτηση επιβίωσης για ένα άτομο με διάνυσμα ανεξάρτητων μεταβλητών Z , $S_0(t)$ υποδηλώνει την συνάρτηση επιβίωσης για ένα άτομο με όλα τα $Z=0$ (baseline survival function) και β είναι το διάνυσμα των συντελεστών παλινδρόμησης από το μοντέλο Cox (Putter et al, 2017). Πιο κάτω θα περιγραφεί το πλαίσιο για την ανάπτυξη κλίμακας εκτίμησης κινδύνου απουσία ανταγωνιστικών κινδύνων το οποίο βασίζεται στο πλαίσιο που χρησιμοποιείται για την ανάπτυξη των Framingham risk scores και επιτρέπει την ενσωμάτωση τόσο των συνεχών όσο και των κατηγορηματικών παραγόντων κινδύνου (Sullivan et al, 2014).

Βήμα 1: Εκτίμηση των συντελεστών παλινδρόμησης

Έστω το μοντέλο Cox αναλογικών κινδύνων

$$\lambda(t; Z) = \lambda_0(t) \exp(\beta_1 Z_1 + \beta_2 Z_2 + \dots + \beta_p Z_p).$$

Τότε τα $\beta_1, \beta_2, \dots, \beta_p$ είναι οι εκτιμήσεις των συντελεστών παλινδρόμησης ανάλογα με το μοντέλο που χρησιμοποιείται κάθε φορά (πχ μοντέλο Cox, μοντέλο λογιστικής παλινδρόμησης κλπ). Δηλαδή εκτιμάται ο συντελεστής παλινδρόμησης για κάθε μεταβλητή πρόβλεψης.

Βήμα 2: Οργάνωση των παραγόντων κινδύνου σε κατηγορίες και καθορισμός των τιμών αναφοράς

Όταν έχουμε συνεχή μεταβλητή πρέπει να την χωρίσουμε σε κατηγορίες και να καθορίσουμε την κατηγορία αναφοράς. Για παράδειγμα, εάν το εύρος της ηλικίας είναι 30-79, χρησιμοποιούμε τις ηλικιακές κατηγορίες: 30-39, 40-49, 50-59, 60-69 και 70-79. Δηλαδή, προσδιορίζουμε τις κλινικά σημαντικές κατηγορίες κάθε μιας από τις συνεχείς μεταβλητές. Έπειτα, θέτουμε μια τιμή αναφοράς ώστε να δώσουμε βαθμό σε κάθε κατηγορία. Γενικά, οι μέσες τιμές είναι αποδεκτές τιμές αναφοράς, πχ χρησιμοποιώντας τις παραπάνω ηλικιακές κατηγορίες έχουμε 34.5, 44.5, 54.5, 64.5 και 74.5 αντίστοιχα. Στην περίπτωση όπου έχουμε δίτιμη μεταβλητή (1=ναι/0=όχι), το 0 αντιπροσωπεύει την κατηγορία αναφοράς. Το W_{ij} υποδηλώνει την τιμή αναφοράς (πχ μέσες τιμές για συνεχείς μεταβλητές που ταξινομούνται σε κατηγορίες

και 0 ή 1 για κατηγορικούς παράγοντες κινδύνου) για την j-οστή κατηγορία του i-οστού παράγοντα κινδύνου, με $i=1,2,\dots,p$ $j=1,2,\dots,c_i$ όπου c_i ο συνολικός αριθμός κατηγοριών του παράγοντα κινδύνου i.

Βήμα 3: Προσδιορισμός του προφίλ του παράγοντα κινδύνου αναφοράς

Στο τρίτο βήμα, καθορίζουμε την κατάλληλη κατηγορία για κάθε παράγοντα κινδύνου που χρησιμεύει ως βασική κατηγορία. Η βασική κατηγορία για κάθε παράγοντα κινδύνου είναι η κατηγορία που αποδίδεται σε 0 βαθμούς στην κλίμακα εκτίμησης κινδύνου. Στις κατηγορίες που αντικατοπτρίζουν τις χειρότερες (λιγότερο υγιείς) καταστάσεις του παράγοντα κινδύνου θα δοθούν θετικά σημεία, ενώ στις κατηγορίες που αντικατοπτρίζουν τις υγιέστερες καταστάσεις θα αποδοθούν αρνητικά σημεία. Η τιμή αναφοράς της βασικής κατηγορίας δηλώνεται με το W_{iREF} , για καθέναν από τους παράγοντες κινδύνου i, $i=1,2,\dots,p$.

Βήμα 4: Προσδιορισμός της απόστασης της κάθε κατηγορίας από την βασική κατηγορία σε μονάδες παλινδρόμησης

Στη συνέχεια, προσδιορίζουμε πόσο μακριά είναι η κάθε κατηγορία από τη βασική κατηγορία W_{iREF} σε μονάδες παλινδρόμησης με χρήση του τύπου:

$$\beta_i(W_{ij} - W_{iREF})$$

για κάθε κατηγορία j του κάθε παράγοντα κινδύνου i.

Βήμα 5: Καθορισμός του προσαρμοσμένου πολλαπλασιαστή B

Ο πολλαπλασιαστής ή σταθερά B είναι ο αριθμός των μονάδων παλινδρόμησης που αντανακλούν σε έναν βαθμό στην τελική κλίμακα. Για παράδειγμα, στην κλίμακα που δημιουργήθηκε στη μελέτη Framingham για τον κίνδυνο εμφάνισης καρδιακής πάθησης, η σταθερά αυτή βασίστηκε στην ηλικία διότι φάνηκε ότι είναι σημαντικός παράγοντας κινδύνου. Έστω X_1 =ηλικία και $B_1=0.05$, τότε η σταθερά που συνδέεται με τον κίνδυνο για αύξηση της ηλικίας κατά 5 έτη είναι $B=0.05 \times 5=0.25$.

Βήμα 6: Προσδιορισμός των βαθμών κάθε κατηγορίας κάθε παράγοντα

Οι βαθμοί που συνδέονται με μια δεδομένη κατηγορία ορίζονται ως η απόσταση αυτής της κατηγορίας από την βασική κατηγορία για τον συντελεστή κινδύνου (που υπολογίζεται στο βήμα 4) διαιρεμένο με B (από το 5^ο βήμα). Ο τύπος είναι ο εξής:

$$points_{ij} = \beta_j \cdot (W_{ij} - W_{iREF}) / B$$

Χρησιμοποιώντας τον τύπο αυτό, η κατηγορία αναφοράς κάθε παράγοντα κινδύνου παίρνει 0 βαθμούς στην κλίμακα.

Βήμα 7: Προσδιορισμός της σχέσης κινδύνου και συνολικών βαθμών

Στο τελευταίο βήμα υπολογίζεται για κάθε παράγοντα κινδύνου ένας συνολικός βαθμός. Δηλαδή, καθορίζεται το σύνολο των βαθμών που σχετίζονται με κάθε προφίλ κινδύνου (κάθε πιθανός συνδυασμός) των διάφορων κατηγοριών κινδύνου. Στο μοντέλο αναλογικών κινδύνων Cox ο τύπος για τον υπολογισμό της εκτίμησης του κινδύνου είναι:

$$\hat{p} = 1 - So(t) \exp(\sum \beta_i X_i - \sum \beta_i \bar{X}_i)$$

4.2. Κλίμακα εκτίμησης κινδύνου παρουσία ανταγωνιστικών γεγονότων

Όπως έχουμε ήδη αναφέρει, η χρήση τέτοιων κλιμάκων εκτίμησης κινδύνου διευκολύνει την λήψη κλινικών αποφάσεων βάσει τεκμηριωμένων στοιχείων. Υπάρχει όμως αυξανόμενο ενδιαφέρον για την θνησιμότητα από συγκεκριμένη αιτία και τις μη θανατηφόρες εκβάσεις. Ωστόσο, όταν εξετάζουμε τέτοιου είδους εκβάσεων πρέπει να λαμβάνουμε υπόψη τους ανταγωνιστικούς κινδύνους των οποίων η εμφάνιση αποκλείει την εμφάνιση του γεγονότος ενδιαφέροντος. Στη συνέχεια θα περιγραφεί το πλαίσιο για την ανάπτυξη κλίμακας εκτίμησης κινδύνου παρουσία ανταγωνιστικών κινδύνων το οποίο βασίζεται στο πλαίσιο που χρησιμοποιείται για την ανάπτυξη των Framingham Risk Scores για την πρόβλεψη στεφανιαίας νόσου, όπου τα σημεία που σχετίζονται με κάθε επίπεδο του παράγοντα κινδύνου είναι

σχετικά με τα σημεία που συνδέονται με μια 10-ετή αύξηση της ηλικίας. Αν και δεν απαιτείται η ηλικία να περιλαμβάνεται σε κάθε κλίμακα εκτίμησης κινδύνου, η συχνότητα εμφάνισης πολλών εκβάσεων αυξάνεται με την αύξηση της ηλικίας. Ωστόσο, σε συγκεκριμένα πλαίσια, οι ερευνητές μπορούν να επιλέξουν να τυποποιήσουν τις επιδράσεις άλλων παραγόντων κινδύνου σε σχέση με μια διαφορετική συνεχή μεταβλητή που είναι μέρος της κλίμακα εκτίμησης κινδύνου. Αναπτύσσουμε μια κλίμακα κινδύνου για τον κ-οστό τύπο συμβάντος παρουσία ανταγωνιστικών κινδύνων (*Austin et al, 2016*).

Βήμα 1: Ο κίνδυνος υπο-κατανομής του γεγονότος ενδιαφέροντος παραπέμπει στο σύνολο των παραγόντων κινδύνου και των προγνωστικών της έκβασης ενδιαφέροντος. Εκτιμάται ο συντελεστής παλινδρόμησης (beta coefficient ή log-hazard ratio) για κάθε μεταβλητή πρόβλεψης.

Βήμα 2: Το μοντέλο μπορεί να περιέχει τόσο συνεχείς όσο και κατηγορικές μεταβλητές, ωστόσο η κλίμακα εκτίμησης κινδύνου είναι οργανωμένη γύρω από τις κατηγορίες. Έτσι, πρέπει να προσδιοριστούν κλινικά σημαντικές κατηγορίες για κάθε μία από τις συνεχείς μεταβλητές.

Βήμα 3: Η δημιουργία μιας κλίμακας εκτίμησης κινδύνου απαιτεί να προσδιορίζεται μια κατηγορία για κάθε παράγοντα κινδύνου που θα χρησιμεύει ως επίπεδο αναφοράς ή βασική κατηγορία. Στην κατηγορία αναφοράς θα αποδοθούν μηδενικά σημεία στην κλίμακα εκτίμησης κινδύνου. Έπειτα προσδιορίζεται το μέσο σημείο της κάθε κατηγορίας που προέκυψε από μια συνεχή μεταβλητή. Η χρήση των ενδιάμεσων σημείων (midpoints) για κάθε κατηγορία επιτρέπει να καθοριστεί η απόσταση μεταξύ κάθε κατηγορίας του παράγοντα κινδύνου και του επιπέδου αναφοράς αυτής της κατηγορίας. Για τον προσδιορισμό του μέσου σημείου της κατώτερης κατηγορίας, μπορεί κανείς να επιλέξει να χρησιμοποιήσει το πρώτο εκατοστημόριο παρά την ελάχιστη τιμή της μεταβλητής. Παρομοίως, για τον προσδιορισμό του μέσου σημείου της υψηλότερης κατηγορίας, μπορεί κανείς να επιλέξει να χρησιμοποιήσει το 99οστο εκατοστημόριο, αντί της μέγιστης τιμής αυτής της μεταβλητής. Για τις μεταβλητές με ένα μικρό ποσοστό ακραίων τιμών, μια τέτοια προσέγγιση μπορεί να οδηγήσει σε πιο ευαίσθητα μεσαία σημεία για τις ακραίες κατηγορίες.

Βήμα 4: Χρησιμοποιώντας τους εκτιμώμενους συντελεστές παλινδρόμησης από το μοντέλο κινδύνου υπο-κατανομής, καθορίζεται πόσο μακριά είναι κάθε κατηγορία του παράγοντα κινδύνου από την κατηγορία αναφοράς του κάθε παράγοντα, σε μονάδες παλινδρόμησης.

Βήμα 5: Στο βήμα αυτό πρέπει να οριστεί η σταθερά B που υποδηλώνει τον αριθμό των μονάδων παλινδρόμησης που αντανακλούν ένα σημείο στην τελική κλίμακα κινδύνου. Θα ακολουθήσουμε την σύμβαση της κλίμακας κινδύνου Framingham, όπου το B υποδηλώνει την αύξηση του κινδύνου που συνδέεται με μια 5-ετή αύξηση της ηλικίας. Επομένως, εάν η ηλικία μετριέται σε έτη, το B θα οριστεί ίσο με το 5πλάσιο του συντελεστή παλινδρόμησης για την ηλικία στο μοντέλο κινδύνου υπο-κατανομής (που εκτιμάται στο βήμα 1).

Βήμα 6: Στη συνέχεια υπολογίζεται ο αριθμός των σημείων που σχετίζονται με κάθε κατηγορία κάθε παράγοντα κινδύνου. Αυτό γίνεται με την πρόσθεση των σημείων που σχετίζονται με καθεμιά από τις κατηγορίες κινδύνου που περιλαμβάνει αυτό το προφίλ κινδύνου.

Βήμα 7: Το τελευταίο βήμα είναι να προσδιοριστεί η επίπτωση της έκβασης εντός συγκεκριμένης χρονικής διάρκειας που συνδέεται με κάθε μία από τις πιθανές τιμές κλίμακας εκτίμησης κινδύνου. Αυτό γίνεται με την χρήση τριών δευτερευόντων βημάτων:

A) Η συνάρτηση αθροιστικής επίπτωσης για ένα δεδομένο διάλυσμα ανεξάρτητων μεταβλητών είναι ίση με

$$I_k(t/x) = 1 - \exp(-\exp(\beta k X) \int_0^t \lambda_{o,k}(s) ds)$$

όπου $\lambda_{o,k}(t)$ η συνάρτηση κινδύνου για τον κ-οστό τύπο συμβάντος για ένα άτομο που έχει όλα τα $X=0$ (baseline hazard function) (Wolbers et al, 2009).

B) Από το μοντέλο που πήραμε από το βήμα 1 μπορούμε να πάρουμε μια εκτίμηση τύπου Breslow για την βασική συνάρτηση κινδύνου για ένα άτομο αναφοράς. Για

δεδομένο χρονικό διάστημα παρακολούθησης, το, μπορούμε να παρατηρήσουμε μια αριθμητική εκτίμηση:

$$\Lambda_{o,k}(t) = \int_0^{t_0} \lambda_{o,k}(s) ds$$

Το $1 - \exp(-\Lambda_{o,k})$ είναι η baseline αθροιστική συνάρτηση επίπτωσης (δηλαδή, το CIF για ένα άτομο του οποίου όλες οι ανεξάρτητες μεταβλητές παίρνουν την τιμή 0) για το κ-στο τύπο συμβάντος.

Γ) Μια κλίμακα εκτίμησης κινδύνου επιδιώκει να απλοποιήσει την εφαρμογή του μοντέλου πρόβλεψης προσεγγίζοντας το

$$\beta_k X = \sum_{j=1}^p \beta_j X_j$$

από το άθροισμα των βαθμών της κλίμακας που σχετίζονται με τον συγκεκριμένο παράγοντα κινδύνου που πολλαπλασιάζεται με τη σταθερά B. Έτσι, στο τύπο από το βήμα A) αντικαθιστούμε το $\beta_k X$ με το άθροισμα των βαθμών που σχετίζονται με το δεδομένο προφίλ παράγοντα κινδύνου πολλαπλασιασμένο με τη σταθερά B. Η τροποποίηση είναι απαραίτητη επειδή κάθε κατηγορία που βασίζεται σε μια συνεχή μεταβλητή (πχ ηλικία) αντιπροσωπεύεται από την απόσταση (σε μονάδες παλινδρόμησης) από το επίπεδο αναφοράς για την κατηγορία αυτή. Έτσι, το άθροισμα των βαθμών πρέπει να τροποποιηθεί με την προσθήκη της αριθμητικής τιμής που σχετίζεται με το επίπεδο αναφοράς (πχ μέση τιμή) πολλαπλασιασμένο με τον συντελεστή παλινδρόμησης για τη συνεχή μεταβλητή πρόβλεψης. Αυτό επιτρέπει την εκτίμηση της επίπτωσης της έκβασης για κάθε τιμή του αθροίσματος των βαθμών της κλίμακας.

4.3. Αξιολόγηση της κλίμακας εκτίμησης κινδύνου

Γνωρίζουμε ότι υπάρχουν περιορισμοί στις κλίμακες εκτίμησης κινδύνου αφού προκειμένου να επιτευχθεί η απλότητα στη χρήση της χάνουμε μερικές από τις πληροφορίες. Η απώλεια πληροφοριών είναι ελαφρώς πιο έντονη όταν οι παράγοντες κινδύνου μοντελοποιούνται ως συνεχείς μεταβλητές επειδή η κλίμακα κινδύνου βασίζεται σε κατηγορίες. Ένας τρόπος αξιολόγησης της κλίμακας κινδύνου είναι η αξιολόγηση του βαθμού συμφωνίας μεταξύ των εκτιμήσεων κινδύνου βάσει της κλίμακας κινδύνου και εκείνων που βασίζονται στο πολυπαραγοντικό μοντέλο που χρησιμοποιήσαμε.

Η αξιολόγηση αυτής της συμφωνίας μπορεί να γίνει με τη χρήση του σταθμισμένου κάππα (weighted kappa) ή αλλιώς Cohen's weighted kappa. Ο συντελεστής kappa του Cohen (κ) είναι ένα στατιστικό στοιχείο που μετρά τη συμφωνία μεταξύ δύο ή περισσότερων ατόμων. Θεωρείται γενικά ότι είναι ένα πιο ισχυρό μέτρο από τον υπολογισμό του απλού ποσοστού συμφωνίας, καθώς το κ λαμβάνει υπόψη τη δυνατότητα τυχαίας συμφωνίας. Υπολογίζεται ως εξής:

$$k = \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e}$$

όπου p_o είναι η παρατηρούμενη συμφωνία μεταξύ των ατόμων και p_e είναι η υποθετική πιθανότητα τυχαίας συμφωνίας. Εάν $\kappa=1$ έχουμε απόλυτη συμφωνία, ενώ αν $\kappa=0$ δεν υπάρχει συμφωνία (Smeeton 1985).

Για παράδειγμα, οι Sullivan et al. (2004), αξιολόγησαν τον βαθμό συμφωνίας μεταξύ των εκτιμήσεων κινδύνου βάσει της κλίμακας κινδύνου και εκείνων που βασίζονται στο μοντέλο λογιστικής παλινδρόμησης και στο μοντέλο Cox χρησιμοποιώντας στο σταθμισμένο κάππα το οποίο υπολογίζεται ως εξής:

$$k = \frac{1 - \sum_{i=1}^k \sum_{j=1}^k W_{ij} X_{ij}}{\sum_{i=1}^k \sum_{j=1}^k W_{ij} m_{ij}}$$

όπου k ο αριθμός των κωδικοποιήσεων, W_{ij} τα βάρη, X_{ij} και m_{ij} οι παρατηρούμενοι και οι αναμενόμενοι πίνακες αντίστοιχα (Cohen, 1968). Βρέθηκε ότι υπάρχει πολύ καλή συμφωνία μεταξύ των εκτιμήσεων κινδύνου που βασίζονται στην κλίμακα

κινδύνου και εκείνων που βασίζονται στο πολυπαραγοντικό μοντέλο λογιστικής παλινδρόμησης και στα μοντέλα αναλογικού κινδύνου Cox αντίστοιχα. Συγκεκριμένα, το σταθμισμένο κάππα για το παράδειγμα της λογιστικής παλινδρόμησης ήταν 0.85 (95% ΔΕ: 0.83-0.86) και για το παράδειγμα με το μοντέλο αναλογικών κινδύνων ήταν 0.87 (95% ΔΕ: 0.85-0.88).

Η αξιολόγηση της κλίμακας κινδύνου είναι πολύ σημαντική αφού η κλίμακα αυτή είναι ένα εργαλείο για τους κλινικούς ιατρούς ώστε να βοηθήσουν στη διαδικασία λήψης κλινικών αποφάσεων σχετικά με τη θεραπεία και να τους βοηθήσουν να παρακινήσουν τους ασθενείς σε υγιείς συμπεριφορές.

Κεφάλαιο 5

Εφαρμογή σε δεδομένα

Πιο πάνω αναπτύχθηκαν οι μέθοδοι ανάπτυξης και επικύρωσης προγνωστικών μοντέλων. Στο παρόν κεφάλαιο θα γίνει εφαρμογή των παραπάνω μεθόδων σε δεδομένα από εννέα νοσοκομειακά κέντρα με ασθενείς με χρόνια ηπατίτιδα Β για την εκτίμηση του κινδύνου εμφάνισης ηπατοκυτταρικού καρκίνου λαμβάνοντας υπόψη τα ανταγωνιστικά γεγονότα.

5.1. Ανάλυση Δεδομένων

Για την ανάλυση χρησιμοποιήθηκαν τα δεδομένα από την δημιουργία της PAGE-B κλίμακας (*Paratheodoridis et al, 2016*). Το σύνολο δεδομένων που περιλάμβανε ασθενείς από εννέα κέντρα χρησιμοποιήθηκε ως σύνολο δεδομένων ανάλυσης για να αποκομιστεί κλίμακα πρόβλεψης ΗΚΚ λαμβάνοντας υπόψη το θάνατο. Όλοι οι ασθενείς με χρόνια ηπατίτιδα Β των εννέα συμμετεχόντων κέντρων συμπεριλήφθηκαν αν ήταν ενήλικες (ηλικίας ≥ 16 ετών), Καυκάσιοι και είχαν λάβει θεραπεία με ETV ή TDF για ≥ 12 μήνες. Τα συμμετέχοντα κέντρα ήταν στην Ελλάδα (Αθήνα [2 κέντρα], Λάρισα, Θεσσαλονίκη), Ιταλία (Μιλάνο), Ισπανία (Βαρκελώνη, Μαδρίτη), Ολλανδία (Ρότερνταμ) και Τουρκία (Άγκυρα). Έχουν συμπεριληφθεί ασθενείς που δεν έλαβαν ή είχαν προηγουμένως υποβληθεί σε θεραπεία με άλλα φάρμακα. Ασθενείς με μη αντιρροπούμενη κίρρωση ήπατος, ΗΚΚ που διαγνώστηκαν πριν από την εμφάνιση του ETV / TDF, ασθενείς με μόλυνση από ηπατίτιδα D, ηπατίτιδα C ή ιό ανθρώπινης ανοσοανεπάρκειας και ασθενείς με μεταμόσχευση ήπατος αποκλείστηκαν. Συνολικά υπήρχαν 1815 ασθενείς και αφαιρέθηκαν οι 490 από το κέντρο του Μιλάνου, οι οποίοι είχαν χρησιμοποιηθούν ως σύνολο επικύρωσης. Έτσι, τα άτομα που αναλύθηκαν ήταν τελικά 1325.

5.1.1. Παρακολούθηση (follow-up) και ορισμοί

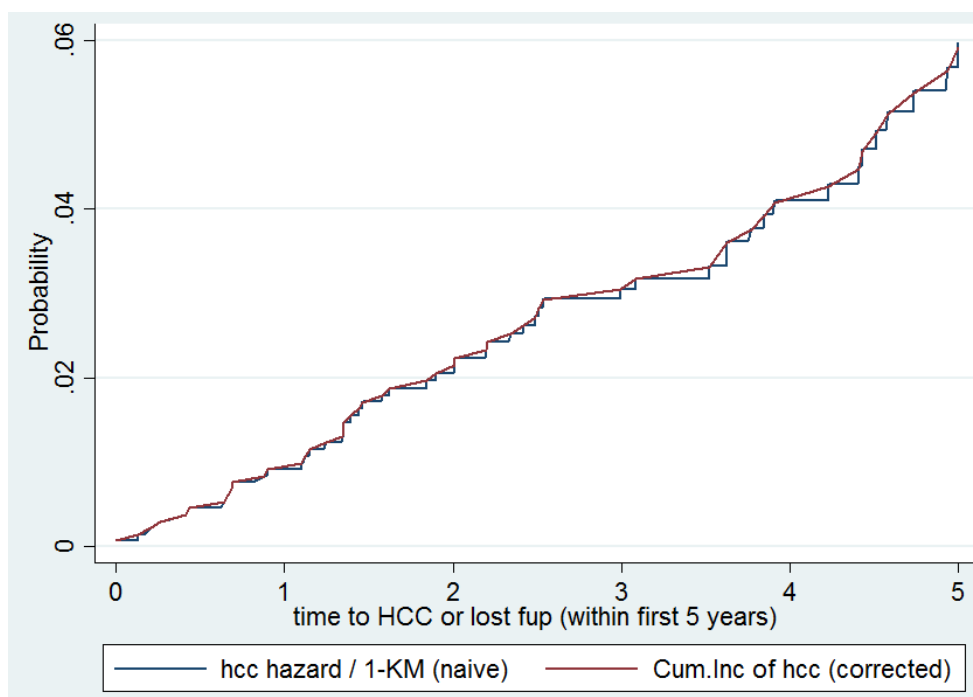
Η χρόνια ηπατίτιδα Β (ΧΗΒ) διαγνώστηκε σε ασθενείς με θετικό HBsAg για ≥ 6 μήνες, αυξημένη αμινοτρανσφεράση αλανίνης (ALT) και ορό DNA HBV > 2000 IU / ml. Οι ασθενείς ταξινομήθηκαν ανάλογα με τη σοβαρότητα της ηπατικής τους νόσου σε: α) ασθενείς με ΧΗΒ μόνο (χωρίς κίρρωση) εάν είχαν υποβληθεί σε βιοψία ήπατος χωρίς να ανευρεθεί κίρρωση και β) ασθενείς με κίρρωση. Ασθενείς χωρίς βιοψία ήπατος πριν από την αγωγή και χωρίς άλλα σημάδια κίρρωσης θεωρήθηκαν ως περιπτώσεις. Όλοι οι ασθενείς υποβλήθηκαν σε θεραπεία με ETV και / ή TDF και παρακολούθηθηκαν σε κάθε συμμετέχον κέντρο σύμφωνα με διεθνείς ή / και εθνικές οδηγίες κλινικής πρακτικής. Η κλινική εξέταση και οι συνήθεις εργαστηριακές εξετάσεις πραγματοποιούνταν τουλάχιστον κάθε 6 μήνες. Τα επίπεδα HBV DNA προσδιορίζονταν κάθε 6-12 μήνες στο εργαστήριο κάθε κέντρου με διάφορες δοκιμασίες αλυσιδωτής αντίδρασης πολυμεράσης (ευαισθησίες: 10-80 IU / ml). Η ιολογική ύφεση θεωρήθηκε ότι υπήρχε σε ασθενείς που πέτυχαν HBV DNA < 80 IU / ml που διατηρήθηκε σε όλη τη θεραπεία με ETV / TDF. Τα επίπεδα υπερηχογραφίας και / ή άλφα-εμβρυοπρωτεΐνης διεξάγονταν κάθε 6 μήνες σε ασθενείς με κίρρωση και κάθε 12 μήνες σε ασθενείς χωρίς κίρρωση. Η διάγνωση της ΗΚΚ βασίστηκε σε τυποποιημένα ιστολογικά και / ή συμβατά ραδιολογικά ευρήματα (European Association for the Study of the Liver, 2012). Η είσοδος στη μελέτη αυτή ορίστηκε ως η ημερομηνία έναρξης του ETV / TDF. Η παρακολούθηση θεωρήθηκε ως το χρονικό διάστημα μεταξύ της καταχώρησης της μελέτης και των τελευταίων διαθέσιμων κλινικών πληροφοριών μέχρι τον Μάιο του 2014, ενώ η διάρκεια της θεραπείας θεωρήθηκε ως το χρονικό διάστημα μεταξύ της εγγραφής της μελέτης μέχρι το τέλος της θεραπείας ή της τελευταίας παρακολούθησης της θεραπείας. Ο χρόνος ανάλυσης ήταν το χρονικό διάστημα μεταξύ της εγγραφής της μελέτης και της διάγνωσης της ΗΚΚ ή το τέλος της παρακολούθησης χωρίς την ανάπτυξη ΗΚΚ. Στον πίνακα 5.1 παρουσιάζονται τα γενικά χαρακτηριστικά του συνόλου δεδομένων που χρησιμοποιήθηκε στην ανάλυση και στο γράφημα 5.1 απεικονίζεται η αθροιστική πιθανότητα ηπατοκυτταρικού καρκίνου μέσα στην 5ετία.

Πίνακας 5.1. Κύρια χαρακτηριστικά καυκάσιων ασθενών με χρόνια ηπατίτιδα Β (ΧΗΒ) που έλαβαν θεραπεία με ETV ή TDF, (N=1325).

Μεταβλητές	N(%) ή mean(SD)
Φύλο	
Γυναίκες	402 (30.3%)
Άνδρες	923 (69.7%)
Ηλικία (έτη)	51.86 (14.24)
Κίρρωση	
Όχι	1033 (79.3%)
Ναι	269 (20.7%)
Αιμοπετάλια ($\times 10^3/mm^3$)	196 (67)
Ανάπτυξη ΗΚΚ κατά την παρακολούθηση	
Όχι	1274 (96.2%)
Ναι	51 (3.8%)

Ως ανταγωνιστικά γεγονότα, θεωρήθηκαν τα άτομα που είχαν πεθάνει αλλά δεν είχαν εμφανίσει προηγουμένως ΗΚΚ και όσα άτομα είχαν κάνει μεταμόσχευση ήπατος (OLT) αλλά δεν έκαναν ΗΚΚ. Συνολικά είχαμε 28 ανταγωνιστικά γεγονότα, 51 γεγονότα αποτυχίας (εμφάνιση ΗΚΚ) και 1246 αποκομμένες παρατηρήσεις. Στη συνέχεια κατασκευάστηκαν οι καμπύλες Kaplan-Meier (1-KM) και οι καμπύλες αθροιστικής επίπτωσης CIF. Στο γράφημα 5.1 απεικονίζονται οι δύο αυτές καμπύλες. Οι εκτιμήσεις KM λαμβάνουν υπόψη τον κάθε χρόνο αποτυχίας αγνοώντας τον θάνατο, δηλαδή, ο υπολογισμός KM δεν λαμβάνει υπόψη την εξάρτηση μεταξύ των ανταγωνιστικών κινδύνων. Συγκεκριμένα, η χρήση του KM απαιτεί ανεξάρτητους κινδύνους και έναν κόσμο όπου δεν συμβαίνουν ανταγωνιστικοί κίνδυνοι. Αντίθετα, βλέπουμε ότι η συνάρτηση CIF λαμβάνει υπόψη και τους χρόνους όπου συμβαίνει το ανταγωνιστικό γεγονός. Αναμένουμε ότι η καμπύλη KM θα είναι πιο ψηλά από την CIF, αφού υπερεκτιμά, όμως αυτό δεν διακρίνεται στο γράφημα λόγω του μικρού αριθμού ανταγωνιστικών γεγονότων που υπήρχε στο σύνολο δεδομένων. Ωστόσο, μπορεί στο σχήμα να μην φαίνεται η διαφορά αυτή, όμως βλέποντας τις εκτιμήσεις KM και CIF, παρατηρούμε ότι στην 5ετία η Kaplan-Meier δίνει λίγο μεγαλύτερη

εκτίμηση από την CI. Ο λόγος που δεν διακρίνουμε μεγάλη διαφορά ανάμεσα στις 2 καμπύλες είναι ο μικρός αριθμός των ανταγωνιστικών γεγονότων.



Γράφημα 5.1. Kaplan-Meier (1-KM) εκτιμήσεις και CIF καμπύλη

5.1.2. Μονοπαραγοντική ανάλυση

Για τις ανάγκες της ανάλυσης δημιουργήθηκε μια μεταβλητή έκβασης («εμφάνιση ΗΚΚ στην 5ετία») με 3 επίπεδα. Το επίπεδο 0 περιέχει τις αποκομμένες παρατηρήσεις (censored), το επίπεδο 1 αν έκανε καρκίνο στην 5ετία και το επίπεδο 2 περιέχει τα ανταγωνιστικά γεγονότα. Δηλαδή, το επίπεδο 1 είναι το γεγονός αποτυχίας (failure event) και το επίπεδο 2 το ανταγωνιστικό γεγονός (competing event). Όλη η ανάλυση έγινε λαμβάνοντας υπόψη τα ανταγωνιστικά γεγονότα, δηλαδή τον θάνατο και το OLT εάν δεν έκανε ΗΚΚ.

Πραγματοποιήθηκε παλινδρόμηση ανταγωνιστικών κινδύνων σύμφωνα με το μοντέλο Fine and Gray. Κατά τη διαδικασία αυτή, τίθεται ένα μοντέλο για τον κίνδυνο υπο-κατανομής για το γεγονός αποτυχίας, γνωστός ως υπο-κίνδυνος (subhazard). Επίσης, σε αντίθεση με τους κινδύνους που οφείλονται σε συγκεκριμένες αιτίες, υπάρχει 1-1 αντιστοιχία μεταξύ των υπο-κινδύνων και των

CIFs για τους αντίστοιχους τύπους συμβάντων. Οι μεταβλητές επηρεάζουν τον υποκίνδυνο αναλογικά, παρόμοια με την παλινδρόμηση του Cox. Με βάση το μοντέλο αυτό, έγινε μονοπαραγοντική ανάλυση όλων των μεταβλητών που είχαν συλλεχθεί με έκβαση την εμφάνιση ΗΚΚ στην 5ετία υπό την παρουσία του θανάτου. Οι μεταβλητές που ελέγχθηκαν είναι η ηλικία (συνεχής μεταβλητή), το φύλο (κατηγορική 0: γυναίκες/ 1:άνδρες), τα αιμοπετάλια (συνεχής μεταβλητή) και η εμφάνιση ή όχι κίρρωσης (κατηγορική 0: CHB-όχι κίρρωση/ 1: κίρρωση). Στον πίνακα 5.2 παρουσιάζονται τα αποτελέσματα από όλες τις μονοπαραγοντικές αναλύσεις, δίνονται οι εκτιμήσεις του λόγου κινδύνου υπο-κατανομής SHR (Subdistribution hazard ratio), το τυπικό σφάλμα (standard error), το p-value και τα αντίστοιχα 95% διαστήματα εμπιστοσύνης (95% Confidence Interval).

Πίνακας 5.2. Μονοπαραγοντική ανάλυση όλων των ανεξάρτητων μεταβλητών με έκβαση την ανάπτυξη ΗΚΚ στην 5ετία υπό την παρουσία του θανάτου (N=1325)

Μεταβλητή	SHR**	St.error	p-value	95% ΔΕ
Φύλο				
Γυναίκες*	1			
Άνδρες	5.12	2.67	0.002	[1.84 , 14.24]
Ηλικία (έτη)	1.06	0.008	<0.001	[1.04 , 1.07]
Αιμοπετάλια ($\times 10^3 / mm^3$)	0.999	2.27e-06	<0.001	[0.998,0.999]
Κίρρωση				
Όχι*	1			
Ναι	6.21	1.79	<0.001	[3.52 , 10.94]

*κατηγορία αναφοράς

** Subdistribution hazard ratio

5.1.3. Επιλογή τελικού μοντέλου

Για να επιλέξουμε το τελικό μοντέλο εφαρμόστηκε προδρομική βηματική (forward stepwise) μέθοδος επιλογής προγνωστικών παραγόντων. Το επίπεδο στατιστικής σημαντικότητας ορίστηκε στο 5% και η εκτίμηση των συντελεστών παλινδρόμησης έγινε με τη μέθοδο μέγιστης πιθανοφάνειας (συγκεκριμένα εδώ, μεγιστοποίηση της μερικής πιθανοφάνειας).

1^ο βήμα: Όπως βλέπουμε από τον παραπάνω πίνακα σχεδόν όλες οι μεταβλητές είναι το ίδιο στατιστικά σημαντικές (<0.001) εκτός από το φύλο το οποίο έχει $p\text{-value}=0.002$. Στην περίπτωση όπου τα $p\text{-value}$ διαφέρουν μεταξύ τους, στο βήμα αυτό επιλέγουμε την μεταβλητή με το μικρότερο $p\text{-value}$ και την προσθέτουμε στο μοντέλο. Εδώ ξεκινάμε με το μοντέλο το οποίο περιέχει την ηλικία.

2^ο βήμα: Στη συνέχεια πραγματοποιήθηκαν όλα τα πολυπαραγοντικά μοντέλα με δύο ανεξάρτητες μεταβλητές («ηλικία» και ο συνδυασμός όλων των υπολοίπων). Όλοι οι πιθανοί συνδυασμοί των παραγόντων που έγιναν ήταν στατιστικά σημαντικοί. Επιλέχθηκε ως δεύτερη μεταβλητή τα αιμοπετάλια με $p\text{-value}<0.001$.

3^ο βήμα: Έπειτα πραγματοποιήθηκαν όλα τα πολυπαραγοντικά μοντέλα με τρεις ανεξάρτητες μεταβλητές («ηλικία», «αιμοπετάλια» και ο συνδυασμός όλων των υπολοίπων). Στο βήμα αυτό προστέθηκε το φύλο με $p\text{-value}=0.002$. Έτσι συνεχίστηκε η διαδικασία με ανεξάρτητες μεταβλητές τις: «ηλικία», «αιμοπετάλια» και «φύλο».

4^ο βήμα: Στο τελευταίο βήμα πραγματοποιήθηκε το πολυπαραγοντικό μοντέλο με τις 3 ανεξάρτητες μεταβλητές που προστέθηκαν στο 3^ο βήμα και την τελευταία ανεξάρτητη μεταβλητή που απέμεινε. Δηλαδή εκτιμήθηκε το μοντέλο με ανεξάρτητες μεταβλητές τις εξής: «ηλικία», «αιμοπετάλια», «φύλο» και «κίρρωση ήπατος». Όλες οι μεταβλητές βρέθηκαν στατιστικά σημαντικές έτσι το τελικό μοντέλο αποτελείται από τις πιο πάνω τέσσερις μεταβλητές.

5.1.4. Τελικό μοντέλο

Το τελικό μοντέλο που προέκυψε μέσα από τη διαδικασία προδρομικής βηματικής επιλογής περιλαμβάνει τους εξής τέσσερις παράγοντες: «ηλικία», «αιμοπετάλια», «φύλο» και «κίρρωση ήπατος». Τα αποτελέσματα από το πολυπαραγοντικό μοντέλο παρουσιάζονται στον πίνακα 5.3.

Πίνακας 5.3. Τελικό μοντέλο πολυπαραγοντικής ανάλυσης με εξαρτημένη μεταβλητή την εμφάνιση ΗΚΚ στην 5ετία λαμβάνοντας υπόψη την παρουσία θανάτου. Το επίπεδο στατιστικής σημαντικότητας ορίστηκε στο 5% και το μέγεθος δείγματος είναι N=1325.

Μεταβλητή	SHR**	St. Error	p-value	95% ΔΕ
Ηλικία (έτη)	1.05	0.01	<0.001	[1.03 , 1.07]
Αιμοπετάλια, $\times 10^3/mm^3(K/\mu l)$	0.99	3.50e-06	0.010	[0.998 , 0.999]
Φύλο				
Γυναίκες*	1			
Άνδρες	4.59	2.43	0.004	[1.62 , 12.96]
Κίρρωση				
Όχι*	1			
Ναι	2.80	1.12	0.010	[1.27 , 6.15]

*κατηγορία αναφοράς

** Subdistribution hazard ratio

5.1.5. Σύγκριση τελικού μοντέλου με το μοντέλο από PAGE-B

Όπως έχουμε ήδη αναφέρει, το τελικό μοντέλο «κτίστηκε» με βάσει τα δεδομένα με τα οποία δημιουργήθηκε η κλίμακα εκτίμησης κινδύνου PAGE-B (*Paratheodoridis et al, 2016*). Έτσι, μπορεί να γίνει η σύγκριση του τελικού μοντέλου (το οποίο λαμβάνει υπόψη τα ανταγωνιστικά γεγονότα) με το μοντέλο από

την PAGE-B (το οποίο δεν λαμβάνει υπόψη τα ανταγωνιστικά γεγονότα). Σημειώνουμε ότι και στα 2 μοντέλα οι μεταβλητές οι οποίες συμπεριλήφθησαν στο μοντέλο ήταν οι ίδιες με επίπεδο στατιστικής σημαντικότητας 5%. Τα αποτελέσματα από τα δύο μοντέλα παρουσιάζονται στον πίνακα 5.4.

Πίνακας 5.4. Αποτελέσματα από τελικό μοντέλο και μοντέλο από PAGE-B (N=1325).

Μεταβλητή	Τελικό μοντέλο (με ανταγωνιστικούς κινδύνους)		Μοντέλο από PAGE-B (χωρίς να λαμβάνονται υπόψη οι ανταγωνιστικοί κίνδυνοι)	
	SHR (95%ΔΕ)	p-value	HR(95%ΔΕ)	p-value
Ηλικία (ανά έτος αύξησης)	1.05(1.03-1.07)	<0.001	1.05(1.02-1.08)	<0.001
Φύλο (άνδρες Vs γυναίκες)	4.59(1.62-12.96)	0.004	4.63(1.66-12.90)	0.003
Αιμοπετάλια, $\times 10^3/mm^3$ (K/ μ l)	0.99(0.998-0.999)	0.010	0.99(0.984-0.996)	0.001
Κίρρωση (Ναι Vs Όχι)	2.80(1.27-6.15)	0.011	2.68(1.39-5.18)	0.003

Τα αποτελέσματα από τα δύο μοντέλα είναι αρκετά κοντά. Παρατηρούμε ότι κάποιες εκτιμήσεις από το μοντέλο με τα ανταγωνιστικά γεγονότα είναι χαμηλότερες. Η διαφορά αυτή ίσως να ήταν μεγαλύτερη εάν είχαμε περισσότερα ανταγωνιστικά γεγονότα (εδώ είχαμε μόλις 28).

5.2. Δημιουργία κλίμακας εκτίμησης κινδύνου εμφάνισης ΗΚΚ σε ασθενείς με χρόνια ηπατίτιδα Β

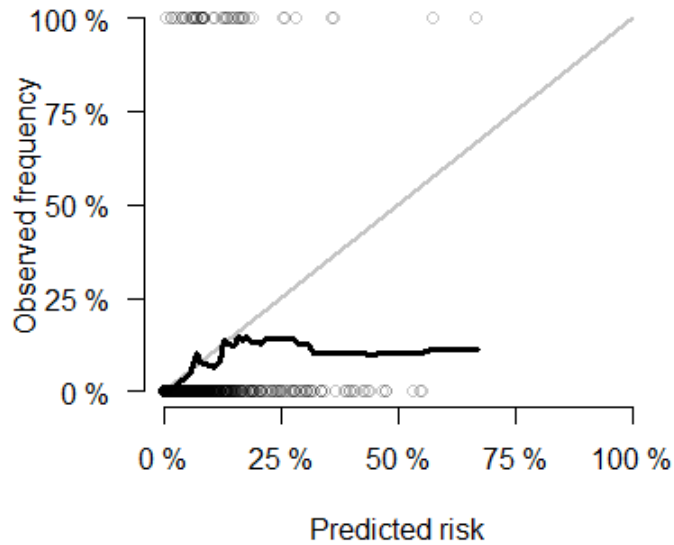
Για τις ανάγκες της δημιουργίας κλίμακας εκτίμησης κινδύνου εμφάνισης ΗΚΚ σε ασθενείς με χρόνια ηπατίτιδα Β έγιναν κάποιες τροποποιήσεις στο πιο πάνω τελικό μοντέλο. Η μεταβλητή της κίρρωσης ήπατος χρειάστηκε να αφαιρεθεί από το μοντέλο, επειδή σε πολλά άτομα δεν υπήρχε πληροφορία για την μεταβλητή αυτή. Αυτό οφείλεται στο ότι χρειάζονται πολλές διαδικασίες ώστε να μπορέσει κάποιος να απαντήσει εάν έχει κίρρωση ή όχι (πχ βιοψία), με αποτέλεσμα να υπάρχουν ελλείπουσες τιμές σε πολλά άτομα. Επίσης, η συνεχής μεταβλητή για τα αιμοπετάλια έγινε κατηγορική με 3 κατηγορίες (1: ≥ 200.000 , 2: 100.000-199.999, 3: < 100.000). Οι εκτιμήσεις του τελικού μοντέλου, με βάση το οποίο κατασκευάστηκε η κλίμακα εκτίμησης κινδύνου, παρουσιάζονται στον πίνακα 5.5.

Πίνακας 5.5. Τελικό μοντέλο με βάση το οποίο δημιουργήθηκε η κλίμακα εκτίμησης κινδύνου (N=1325).

Μεταβλητή	β (coef.)	St. Error	p-value	95% ΔΕ
Ηλικία	0.05	0.009	<0.001	(0.03 , 0.06)
Αιμοπετάλια				
>200*	1			
100-199	1.46	1.93	0.001	(0.58 , 2.32)
<100	2.19	4.59	<0.001	(1.19 , 3.19)
Φύλο				
Γυναίκες *	1			
Άνδρες	1.56	2.50	0.003	(0.53 , 2.59)

*κατηγορία αναφοράς

Η ικανότητα διαχωριστικότητας (discrimination) του παραπάνω μοντέλου εξετάστηκε μέσω του δείκτη c που αναφέρθηκε πιο πάνω, και βρέθηκε ότι ισούται με 0.81. Επίσης, έγινε εσωτερική επικύρωση με bootstrap όπου επίσης βρέθηκε ότι $c=0.81$. Η βαθμονόμηση του μοντέλου εκτιμήθηκε γραφικά συγκρίνοντας την προβλεπόμενη πιθανότητα εμφάνισης ηπατοκυτταρικού καρκίνου εντός πέντε ετών με την παρατηρούμενη πιθανότητα εμφάνισης ΗΚΚ εντός πέντε ετών (γράφημα 5.2).



Γράφημα 5.2. Βαθμονόμηση του προγνωστικού μοντέλου

5.2.1. Κλίμακα εκτίμησης κινδύνου – Risk Score

Στον πίνακα 5.6 δίνονται οι παράγοντες κινδύνου που χρησιμοποιήθηκαν για τη δημιουργία της κλίμακας κινδύνου, οι αντίστοιχες κατηγορίες αναφοράς W_{ij} , οι συντελεστές παλινδρόμησης β_i , οι αποστάσεις κάθε κατηγορίας από τη βασική κατηγορία $\beta_i(W_{ij} - W_{iREF})$, και οι τελικοί βαθμοί που προκύπτουν μετά από στρωγγυλοποίηση των αποστάσεων $\beta_i(W_{ij} - W_{iREF})/B$.

Πίνακας 5.6. Κλίμακα εκτίμησης κινδύνου

Μεταβλητές	W_{ij}	β_i	$\beta_i(W_{ij} - W_{iREF})$	Βαθμοί _{ij} = $\beta_i(W_{ij} - W_{iREF})/B$
Ηλικία		0.05		
<30	24.5= W_{1REF}		0	0
30-39	34.5		0.5	2
40-49	44.5		1	4
50-59	54.5		1.5	6
60-69	64.5		2	8

≥ 70	74.5		2.5	10
Φύλο		1.59		
Γυναίκες	$0=W_{2REF}$		0	0
Άνδρες	1		1.59	6
Αιμοπετάλια		-0.01		
≥ 200.000	$300.000=W_{3REF}$		0	0
100.000-199.999	150.000		1.5	6
<100.000	70.000		2.3	9

Το θεωρητικό εύρος των βαθμών είναι από 0 έως 25. Στη συνέχεια εκτιμήθηκαν οι αντίστοιχες πιθανότητες αθροιστικής επίπτωσης για κάθε βαθμολογία, σύμφωνα με το 7^ο βήμα για την ανάπτυξη κλίμακας κινδύνου παρουσία ανταγωνιστικών κινδύνων όπως περιγράφηκε στην ενότητα 4.2. Στον πίνακα 5.7 παρουσιάζεται ο πενταετής κίνδυνος εμφάνισης ΗΚΚ σύμφωνα με την βαθμολογία κινδύνου, παρουσία ανταγωνιστικών κινδύνων.

Πίνακας 5.7. Συνολικοί βαθμοί κινδύνου και οι αντίστοιχες εκτιμήσεις κινδύνου στην 5ετία.

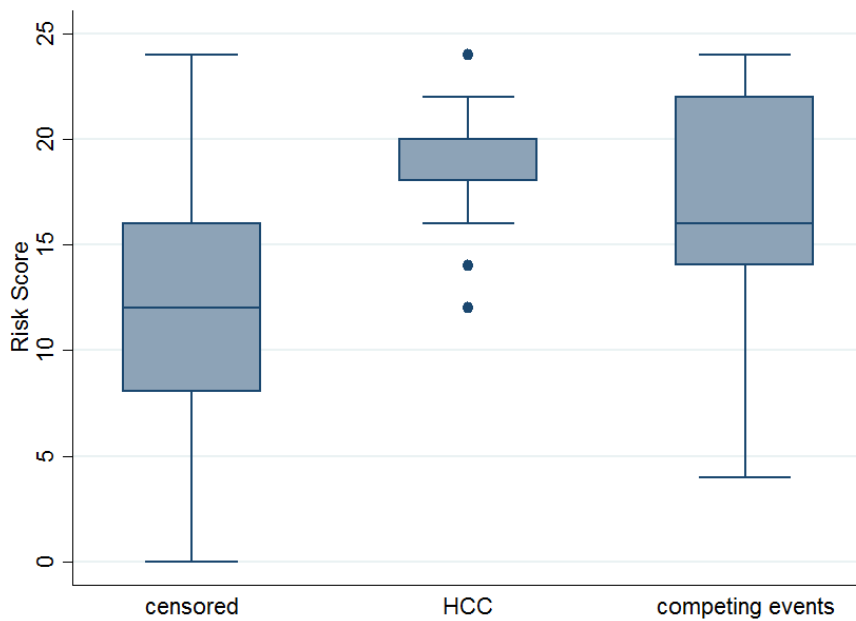
Βαθμοί κινδύνου	CIF (στην 5ετία)	%CIF
0	0.0008	0.08
2	0.0014	0.14
4	0.0024	0.24
6	0.0040	0.40
8	0.0068	0.68
9	0.0088	0.88
10	0.0113	1.13
12	0.0189	1.89
13	0.0243	2.43
14	0.0313	3.13
15	0.0403	4.03
16	0.0518	5.18

17	0.0665	6.65
18	0.0851	8.51
19	0.1086	10.86
20	0.1380	13.80
21	0.1747	17.47
22	0.2197	21.97
23	0.2797	27.97
25	0.4146	41.46

Πίνακας 5.8. Σύγκριση των εκτιμήσεων κινδύνου της τελικής κλίμακας σε σχέση με τις εκτιμήσεις κινδύνου της κλίμακας PAGE-B

Βαθμοί κινδύνου	%CIF (τελική κλίμακα)	%risk (PAGE-B)
0	0.08	0.09
2	0.14	0.15
4	0.24	0.25
6	0.40	0.43
8	0.68	0.73
9	0.88	0.94
10	1.13	1.23
12	1.89	2.07
13	2.43	2.68
14	3.13	3.48
15	4.03	4.50
16	5.18	5.82
17	6.65	7.50
18	8.51	9.64
19	10.86	12.37
20	13.80	15.79
21	17.47	20.04
22	21.97	25.25
23	27.97	31.53
25	41.46	47.36

Η βαθμολογία ≤ 9 υποδηλώνει καθόλου ή ελάχιστο 5ετή κίνδυνο εμφάνισης ΗΚΚ, ενώ η βαθμολογία ≥ 10 και ιδιαίτερα ≥ 18 υποδηλώνουν αυξημένο 5ετή κίνδυνο εμφάνισης ΗΚΚ που απαιτεί συνεχή και προσεκτική παρακολούθηση. Στη συνέχεια αξιολογήθηκε η κλίμακα κινδύνου υπολογίζοντας τον δείκτη c, ο οποίος βρέθηκε ότι ισούται με 0.81. Κατασκευάστηκε επίσης, θηκόγραμμα (γράφημα 5.3) το οποίο δείχνει την κατανομή της κλίμακας κινδύνου ανά τιμή έκβασης για το σύνολο δεδομένων.



Γράφημα 5.3. Θηκόγραμμα για την κατανομή της κλίμακας ανά τιμή της έκβασης.

Κεφάλαιο 6

Συζήτηση

Στην παρούσα διπλωματική εργασία πραγματοποιήθηκε μια ανασκόπηση και περιγραφή των μεθόδων που χρησιμοποιούνται για την ανάπτυξη και επικύρωση προγνωστικών μοντέλων σε δεδομένα επιβίωσης, λαμβάνοντας υπόψη τους ανταγωνιστικούς κινδύνους. Ένα ανταγωνιστικό γεγονός είναι οτιδήποτε μπορεί να παρεμποδίσει την παρατήρηση του γεγονότος που μας ενδιαφέρει, όπως για παράδειγμα ο θάνατος. Έγινε αναλυτική περιγραφή της μεθόδου ανάπτυξης κλίμακας εκτίμησης κινδύνου απουσία και παρουσία των ανταγωνιστικών κινδύνων.

Το προγνωστικό μοντέλο, καθώς και η κλίμακα εκτίμησης κινδύνου εμφάνισης ηπατοκυτταρικού καρκίνου στην 5ετία, αναπτύχθηκαν για ασθενείς με χρόνια ηπατίτιδα Β οι οποίοι υποβάλλονται σε θεραπεία με ETV ή TDF. Τα δεδομένα, είναι αυτά που χρησιμοποιήθηκαν για την ανάπτυξη της κλίμακας εκτίμησης PAGE-B, (*Paratheodoridis et al, 2016*), αλλά εδώ λήφθηκαν υπόψη και οι ανταγωνιστικοί κίνδυνοι, δηλαδή ο θάνατος από κάθε αιτία και αν είχαν κάνει μεταμόσχευση ήπατος αυτοί που δεν είχαν εμφανίσει ΗΚΚ. Ένα σημαντικό πλεονέκτημα της κλίμακας κινδύνου που αναπτύχθηκε είναι η πολύ απλή και εύκολη χρήση της στην κλινική πράξη, καθώς βασίζεται μόνο στην ηλικία, το φύλο και τον αριθμό των αιμοπεταλίων του κάθε ασθενούς, χωρίς να χρειάζεται κάποιος περίπλοκος μαθηματικός υπολογισμός.

Η κίρρωση είναι ένας γνωστός παράγοντας κινδύνου για εμφάνιση ηπατοκυτταρικού καρκίνου σε ασθενείς με χρόνια ηπατίτιδα Β, κάτι που επιβεβαιώθηκε και στην εργασία αυτή, αφού ήταν ένας στατιστικά σημαντικός παράγοντας. Ωστόσο, η διάγνωση της κίρρωσης μπορεί να μην είναι πάντα απλή, καθώς απαιτεί βιοψία ήπατος ή τουλάχιστον μια αξιόπιστη μη επεμβατική μέθοδο εκτίμησης της ίνωσης. Για τον λόγο αυτό, σε πολλούς ασθενείς υπήρχαν ελλείψεις τιμές στην μεταβλητή της κίρρωσης και έτσι δε λήφθηκε υπόψη η μεταβλητή αυτή στο μοντέλο πριν την ανάπτυξη της κλίμακας.

Εφαρμόστηκε εσωτερική επικύρωση με $c\text{-index}=0.81$, όμως δεν εφαρμόστηκε εξωτερική επικύρωση, αλλά σύγκριση των αποτελεσμάτων μας με αυτά της PAGE-B. Τα αποτελέσματα ήταν πολύ κοντά με αυτά της PAGE-B, πράγμα που δείχνει την εγκυρότητα της PAGE-B, αφού το γεγονός ότι δεν έλαβαν υπόψη τους ανταγωνιστικούς κινδύνους δεν επηρέασε τις εκτιμήσεις. Το γεγονός αυτό οφείλεται στο ότι τα ανταγωνιστικά γεγονότα ήταν πολύ λίγα με αποτέλεσμα να μη φαίνεται ξεκάθαρα η υπερεκτίμηση της Kaplan-Meier, ωστόσο σε περίπτωση που ήταν περισσότερα οι εκτιμήσεις των δύο μοντέλων θα ήταν διαφορετικές. Σε κάθε περίπτωση, τα ανταγωνιστικά γεγονότα θα πρέπει να λαμβάνονται υπόψη, αφού η παρουσία τους δεν μας επιτρέπει να παρατηρήσουμε το γεγονός ενδιαφέροντος.

Σύμφωνα με την κλίμακα εκτίμησης κινδύνου εμφάνισης ΗΚΚ, οι βαθμολογίες ≤ 9 υποδηλώνουν καθόλου ή ελάχιστο 5ετή κίνδυνο ΗΚΚ ενώ οι βαθμολογίες ≥ 10 και ιδιαίτερα ≥ 18 δείχνουν αυξημένο 5ετή κίνδυνο για εμφάνιση ΗΚΚ που απαιτεί συνεχή και προσεκτική παρακολούθηση. Το ποσοστό των ασθενών που ταξινομήθηκαν στην ομάδα χαμηλού κινδύνου, και ίσως αποφύγουν την παρακολούθηση για ΗΚΚ, ήταν περίπου 25%, ενώ το ποσοστό των ατόμων που ταξινομήθηκαν στην ομάδα υψηλού κινδύνου, και απαιτείται στενή παρακολούθηση, ήταν περίπου 27%.

Τα προγνωστικά μοντέλα χρησιμοποιούνται ευρέως στην ιατρική με σκοπό να βοηθήσουν τους γιατρούς καθώς και τους ασθενείς στην λήψη αποφάσεων. Επίσης, η ανάπτυξη κλιμάκων που βασίζονται σε αυτά τα μοντέλα, κάνει πολύ εύκολη την εκτίμηση του κινδύνου κάθε ασθενούς, αφού απαιτούν την γνώση μόνο ορισμένων μετρήσεων των ασθενών. Συμπερασματικά, η παρούσα εργασία επικεντρώθηκε στα βήματα ανάπτυξης και επικύρωσης προγνωστικών μοντέλων και της δημιουργίας κλίμακας εκτίμησης κινδύνου, δίνοντας έμφαση στα ανταγωνιστικά γεγονότα και στον τρόπο με τον οποίο μπορούμε να τα λάβουμε υπόψη.

Περίληψη

Ο σκοπός της παρούσας εργασία ήταν να πραγματοποιηθεί μια ανασκόπηση και μια περιγραφή των μεθόδων που χρησιμοποιούνται για την ανάπτυξη και την επικύρωση προγνωστικών μοντέλων, να γίνει μια ανασκόπηση της μεθοδολογίας για την ανάπτυξη κλίμακας εκτίμησης κινδύνου υπό την παρουσία ανταγωνιστικών κινδύνων και τέλος να γίνει μια εφαρμογή σε διαθέσιμα δεδομένα με ασθενείς με χρόνια ηπατίτιδα Β για την εκτίμηση του κινδύνου εμφάνισης ηπατοκυτταρικού καρκίνου λαμβάνοντας υπόψη τα ανταγωνιστικά γεγονότα.

Γίνεται μια αναφορά στις βασικές έννοιες της Ανάλυσης Επιβίωσης (δεδομένα επιβίωσης, συνάρτηση κινδύνου, αποκομμένες παρατηρήσεις κλπ), στο μοντέλο αναλογικών κινδύνων Cox, καθώς και στους διάφορους έλεγχους της υπόθεσης αναλογικότητας (γραφικοί έλεγχοι, έλεγχος μέσω καταλοίπων κλπ). Οι ανταγωνιστικοί κίνδυνοι εμφανίζονται συχνά στην ανάλυση των δεδομένων επιβίωσης. Αναφέρονται μέθοδοι στατιστικής ανάλυσης παρουσία ανταγωνιστικών κινδύνων (αθροιστική συνάρτηση κινδύνου, συνάρτηση κινδύνου συγκεκριμένης αιτίας και συνάρτηση κινδύνου υπό-κατανομής), καθώς και μοντέλα παλινδρόμησης παρουσία ανταγωνιστικών κινδύνων, όπως το ημιπαραμετρικό μοντέλο αναλογικών κινδύνων των Fine&Gray για την εκτίμηση του κινδύνου υπό-κατανομής.

Για τη διαμόρφωση ενός προγνωστικού μοντέλου, το οποίο λαμβάνει υπόψη τους ανταγωνιστικούς κινδύνους, είναι απαραίτητο να καθοριστεί η έκβαση, όλοι οι προγνωστικοί παράγοντες και το ανταγωνιστικό γεγονός. Σημαντικό είναι, επίσης, να επιλεγεί ο σωστός τύπος μοντέλου ανάλογα με τα δεδομένα, για παράδειγμα, εδώ είχαμε δεδομένα επιβίωσης και έτσι επιλέχθηκε το μοντέλο Fine&Gray το οποίο λαμβάνει υπόψη τα ανταγωνιστικά συμβάντα. Στη συνέχεια γίνεται διαχείριση των ελλειπουσών τιμών και η κωδικοποίηση των συνεχών μεταβλητών (διχοτόμηση, κατηγοριοποίηση, splines κλπ). Το επόμενο βήμα είναι η μέθοδος επιλογής των προγνωστικών παραγόντων κινδύνου (βηματική μέθοδος) και ακολούθως γίνεται η εκτίμηση των συντελεστών παλινδρόμησης, συνήθως με τη μέθοδο μέγιστης πιθανοφάνειας. Αναφέρονται επίσης, νέες μέθοδοι εκτίμησης των συντελεστών (LASSO, μέθοδος κορυφογραμμής), οι οποίες έχουν σκοπό να περιορίσουν την

υπερπροσαρμογή του μοντέλου. Μετά την επιλογή του τελικού μοντέλου, είναι απαραίτητο να γίνει η επικύρωσή του (εσωτερική ή εξωτερική) και η αξιολόγηση της βαθμονόμησής του, της ικανότητας διαχωριστικότητας του και της κλινικής του χρησιμότητας. Τέλος, έγινε αναφορά στα βήματα που πρέπει να γίνουν για την ανάπτυξη μιας κλίμακας κινδύνου απουσία και παρουσία ανταγωνιστικών κινδύνων.

Χρησιμοποιήθηκαν τα δεδομένα από την ανάπτυξη της κλίμακας κινδύνου PAGE-B (ασθενείς από εννέα κέντρα με χρόνια ηπατίτιδα Β). Η έκβαση ήταν η εμφάνιση ΗΚΚ στην 5ετία και ως ανταγωνιστικά γεγονότα θεωρήθηκαν ο θάνατος από όλες τις αιτίες και η μεταμόσχευση ήπατος πριν από την εμφάνιση ΗΚΚ. Πραγματοποιήθηκε προδρομική βηματική μέθοδος επιλογής παραγόντων σε επίπεδο στατιστικής σημαντικότητας 5% από την οποία προέκυψαν τέσσερις στατιστικά σημαντικοί παράγοντες κινδύνου για την εμφάνιση ΗΚΚ στην 5ετία: φύλο, ηλικία, αιμοπετάλια και κίρρωση του ήπατος. Το μοντέλο αυτό αξιολογήθηκε με εσωτερική επικύρωση και βρέθηκε ότι $c\text{-index}=0.81$. Για την ανάπτυξη της κλίμακας χρησιμοποιήθηκε το μοντέλο με τις μεταβλητές: ηλικία, φύλο και αιμοπετάλια, επειδή στην μεταβλητή της κίρρωσης υπήρχαν πολλές ελλείπουσες τιμές. Οι ασθενείς με βαθμολογία ≤ 9 είχαν καθόλου ή ελάχιστο κίνδυνο εμφάνισης ΗΚΚ στην 5ετία, και οι ασθενείς με βαθμολογία ≥ 18 είχαν υψηλό κίνδυνο εμφάνισης ΗΚΚ. Περίπου το 25% των ασθενών ταξινομήθηκαν στην ομάδα χαμηλού κινδύνου, και ίσως αποφύγουν την παρακολούθηση για ΗΚΚ, ενώ το 27% των ατόμων ταξινομήθηκαν στην ομάδα υψηλού κινδύνου και απαιτείται στενή παρακολούθηση.

Σήμερα, η ιατρική βασίζεται πολύ στα προγνωστικά μοντέλα, αφού μέσα από αυτά ενημερώνει τους γιατρούς αλλά και τους ίδιους τους ασθενείς για τους παράγοντες κινδύνου που προκαλούν την εμφάνιση μιας ανεπιθύμητης έκβασης και τους καθοδηγούν ώστε να λάβουν τις σωστές αποφάσεις για τον κάθε ασθενή.

Summary

The aim of this thesis is the review and the description of methods to develop and validate predictive models. Moreover, the methodology for the development of the risk score under the presence of competing risks was reviewed. After this review a application to available data from patients with chronic hepatitis B was performed in order to predict the risk of developing hepatocellular carcinoma taking in account the competing events.

We make a reference to basic definitions of survival analysis (survival data, hazard function, censored observations, etc), to Cox proportional hazards model and to various checks of proportional hazards assumptions (graphical checks, through residuals, etc). The competing risks often appear in survival data analysis. We refer to methods of statistical analysis under competing risks (cumulative incidence function, cause-specific hazard function and sub-distribution hazard function) and to regression models with the presence of competing risks, like the semi-parametric model of proportional risks by Fine&Gray for the prediction of sub-distribution hazard.

In order to develop a predictive model, which takes in account the competing risks, it is necessary to define the outcome, all the predictive factors and the competing event. In the presence of competing risks, the appropriate model should be chosen Fine&Gray model. The next steps include evaluation of missing values , coding of predictors, model specification and model estimation. There are methods to estimate the coefficients (LASSO, ridge regression) to reduce overfitting of the model. After the choice of the final mode, a validation (internal or external) and, an evaluation of the calibration, discrimination and its clinical usefulness are performed. In the thesis, the methodology for the development of a risk score with or without competing risks was described.

The application of these methods was based on data from PAGE-B risk score (patients from nine centers with chronic hepatitis B). The outcome was the the development of hepatocellular carcinoma (HCC) in five years after treatment initiation. The competing events were death and liver transplantation. A forward stepwise method was and there were four variables included in the model: age, sex, platelet count and cirrhosis. Internal validation was performed using bootstrap and the resulting c-index was 0.081 based on this score, a points-based risk score was

developed to predict the 5-year risk of HCC development. Patients with score ≤ 9 had low risk of HCC, whereas patients with score ≥ 18 had high 5-year risk. The difference between the risk score taking into account competing events and PAGE-B were negligible due the low number of competing events in this dataset.

Prediction models are now used widely and they provide an important tool for clinicians. As these models may guide decisions related to the follow-up and treatment if patients it is important to strengthen methodological rigor. Competing risks should be consider in the development of prediction models when analyzing survival data.

Ευρετήριο αγγλικών-ελληνικών

Akaike's Information Criterion

Πληροφοριακό κριτήριο Ακάικε

Backward selection

Οπισθοδρομική μέθοδος

Baseline hazard function

Βασική συνάρτηση κινδύνου

Bias

Μεροληψία

Binary

Διαδική

Blinded

Τυφλή

Calibration

Βαθμονόμηση

Calibration slope

Κλίση βαθμονόμησης

Cause-specific hazard function

Συνάρτηση κινδύνου συγκεκριμένης αιτίας

XHB

Χρόνια ηπατίτιδα Β

Censoring

Αποκοπή

CIF-Cumulative Incidence Function

Αθροιστική συνάρτηση επίπτωσης

Clinical life tables

Κλινικοί πίνακες επιβίωσης

Clinical usefulness

Κλινική χρησιμότητα

Competing risks

Ανταγωνιστικοί κίνδυνοι

Concordance measure

Μέτρο συμμόρφωσης

Concordance probability

Πιθανότητα συμμόρφωσης

Confidence interval

Διάστημα εμπιστοσύνης

Controls

Ομάδα ελέγχου

Covariate

Συμμεταβλητή

Competing risk regression

Παλινδρόμηση ανταγωνιστικών κινδύνων

Cox's Proportional Hazard model

Μοντέλο αναλογικών κινδύνων του Cox

Cross validation

Μέθοδος διεπικύρωσης

Deviance residuals

Κατάλοιπα απόκλισης

Discrimination

Διαχωριστικότητα

Dummies

Ψευδομεταβλητές

Endpoint

Τελικό σημείο

External validation

Εξωτερική επικύρωση

Factors

Παράγοντες

Failure indicator

Δείκτης αποτυχίας

Forward selection

Προδρομική μέθοδος

Frequency tables

Πίνακες συχνοτήτων

Geographical validation

Γεωγραφική επικύρωση

Gold standard

Χρυσό πρότυπο

Hazard ratio

Λόγος κινδύνου

HBV

Ιός ηπατίτιδας Β

Interval censoring

Αποκοπή κατά διαστήματα

Interval validation

Εσωτερική επικύρωση

Knots

Κόμβοι

Left censoring

Αριστερή αποκοπή

Life tables

Πίνακες επιβίωσης

Likelihood Ratio test

Έλεγχος του λόγου πιθανοφανειών

Logistic regression

Λογαριθμιστική παλινδρόμηση

Median survival time

Διάμεσος χρόνος επιβίωσης

Missing data

Ελλείπουσες τιμές

Multiple imputation method

Μέθοδος πολλαπλής αντικατάστασης

Nested models

Εμφωλευμένα μοντέλα

Net Benefit

Καθαρό κέρδος

Ordinal

Διατάξιμη

Outcome

Έκβαση

Overfitting

Υπερπροσαρμογή

Partial likelihood

Μερική πιθανοφάνεια

Points based risk score

Κλίμακα εκτίμησης κινδύνου

Population life table

Πίνακας επιβίωσης πληθυσμού

Receiver Operating Characteristic

Καμπύλες λειτουργικού χαρακτηριστικού δείκτη

Relapse

Υποτροπή της νόσου

Residuals

Κατάλοιπα

Ridge regression

Μέθοδος παλινδρόμησης κρυφογραμμής

Right censoring

Δεξιά αποκοπή

Rotation estimation

Εκτίμηση περιστροφής

Sensitivity

Ευαισθησία

Shrinkage

Συρρίκνωση

Smoothing

Εξομάλυνση

Specifity

Ειδικότητα

Split sample method

Μέθοδος διαχωρισμού δείγματος

Staggered entry

Σταδιακή είσοδος

Standard error

Τυπικό σφάλμα

Stepwise selection

Βηματική επιλογή

Subdistribution hazard model

Μοντέλο κινδύνου υπο-κατανομής

Subdistribution hazard ratio

Λόγος κινδύνου υπο-κατανομής

Subhazard

Υπο-κίνδυνος

Survival function

Συνάρτηση επιβίωσης

Temporal validation

Χρονική επικύρωση

Ties

Ισότιμες παρατηρήσεις

Time scale

Κλίμακα χρόνου

Weighted kappa

Σταθμισμένο κάππα

Βιβλιογραφία

- Adams F, ed. The genuine works of Hippocrates. Baltimore: Wilkins and Wilkins, 1939
- Altman DG, Bland JM. Missing data. *BMJ* 2007;334:424.
- Aronson, E., Wilson, T. D., Akert, R. M., & Fehr, B. (2007). *Social psychology*. (4 ed.). Toronto, ON: Pearson Education
- Assmann G, Culldn P, Schulte H. Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the prospective cardiovascular Münster (PROCAM) study. *Circulation* 2002;105(3):310-5.
- Austin P, Douglas S. Lee, Ralph B. D'Agostino and Jason P. Fine, Developing points-based risk-scoring systems in the presence of competing risks, 2016, *Statistics in Medicine*
- Austin PC, Lee DS, D'Agostino RB, Fine JP, Developing points-based risk-scoring systems in the presence of competing risks. *Stat Med*. 2016;35(22):4056-72.
- Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Stat Med* 2014;33:517–535.
- Bakoyiannis G, Touloumi G, Practical methods for competing risks data: A review, 2010, *Statistical Methods in Medical Research* 21(3) 257–272.
- Bland, J. M.; Altman, D. G. (2004). "The logrank test". *BMJ*. 328 (7447): 1073.
- Breslow, N. E. (1975). "Analysis of Survival Data under the Proportional Hazards Model". *International Statistical Review / Revue Internationale de Statistique*. 43 (1): 45–57
- Cai, T., Pepe, M. S., Zheng, Y., Lumley, T. & Jenny, N. S. (2006), "The sensitivity and specificity of markers for event times," *Biostatistics*, 72, 182-97.
- Cessie S. Le and Houwelingen J. C. Van. Ridge Estimators in Logistic Regression. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*. 1992;41(1):191-201.
- Chapman and Hall; 1992. *Analysis of Survival times*. In: *Practical statistics for Medical research*; pp. 365–93
- Cockburn F CR, Gamsu HR, Greenough A, Hopkins A, International Neonatal Network. The CRIB (clinical risk index for babies) score: a tool assessing initial

neonatal risk and comparing performance of neonatal intensive care units. *Lancet* 1993;342:1938

- Cohen, J. (1968). "Weighed kappa: Nominal scale agreement with provision for scaled disagreement or partial credit". *Psychological Bulletin*. 70 (4): 213–220.
- Collet David: (2003) *Modelling survival data in Medical research*, second edition, Chapman & Hall/CRC, 2nd edition
- Conroy RM, Pyörälä K, Fitzgerald AP, Sans S, Menotti A, De Backer G, De Bacquer D, Ducimetière P, Jousilahti P, Keil U, Njølstad I, Oganov RG, Thomsen T, Tunstall-Pedoe H, Tverdal A, Wedel H, Whincup P, Wilhelmsen L, Graham IM; SCORE project group. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J*. 2003;24(11):987-1003.
- Cox D, Oakes D. *Analysis of Survival Data*. Chapman & Hall: London, 1984
- Cox D –Regression models with life tables, 1972, *JRSS Series B*, 34: 187-220
- Cox DR. Two further applications of a model for binary regression. *Biometrika* 1958;45:562–565.
- Czepiel S.A. *Maximum Likelihood Estimation of Logistic Regression Models: Theory and Implementation*. 2002.
- D'Agostino RB Sr, Grundy S, Sullivan LM, Wilson P. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. *JAMA* 2001;286:180–187.
- Darity, William A., Jr., ed. (2008). "Censoring, Left and Right". *International Encyclopedia of the Social Sciences*. (2nd ed.). Macmillan. pp. 473–474.
- Efron 1977, *The Efficiency of Cox's Likelihood Function for Censored Data*
- Efron and Tibshirani 1993. *An introduction to the bootstrap*, CHAPMAN & HALL/CRC Boca Raton London New York Washington, D.C
- European Association for the Study of the Liver. European Organisation for Research and Treatment of Cancer. *EASL-EORTC Clinical Practice Guidelines: management of hepatocellular carcinoma*. *J Hepatol* 2012;56:908–943
- Ewout W. Steyerberg and Yvonne Vergouwe, 2014: *Towards better clinical prediction models: seven steps for development and an ABCD for validation*
- Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94:496–509

- Fisher B, Costantino JP, Wickerham DL, Redmond CK, Kavanah M, Cronin WM, et al. Tamoxifen for prevention of breast cancer: report of the national surgical adjuvant breast and bowel project P1 study. *J Natl Cancer Inst* 1998;90:137188
- Frank E. Harrell Jr. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. 2001.
- Galea MH, Blamey RW, Elston CE, Ellis IO. The Nottingham prognostic index in primary breast cancer. *Breast Cancer Res Treat* 1992;22:20719.
- Gray RJ. A class of k-sample tests for comparing the cumulative incidence of a competing risk. *Ann Stat* 1988; 16: 1141–1154.
- Greenland S. The need for reorientation toward cost-effective prediction. *Stat Med* 2008;27:199–206.
- Grobbee D E and Hoes A W. *Clinical Epidemiology: Principles, Methods and Applications for Clinical Research*. *International Journal of Epidemiology* 2010;39:318–319.
- Hamparsum Bozdogan: Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions, 1987, pp 345–370
- Harrell, F., Califf, R., Pryor, D., Lee, K., Rosati, R., 1982. Evaluating the yield of medical tests. *Journal of the American Medical Association* 247, 2543–2546.
- Harrington, David (2005). "Linear Rank Tests in Survival Analysis". *Encyclopedia of Biostatistics*.
- Heagerty, P. J. & Zheng, Y. (2005), "Survival Model Predictive Accuracy and ROC Curves," *Biometrics*, 61, 92–105.
- Hocking R. R.. A Biometrics Invited Paper. The Analysis and Selection of Variables in Linear. *Biometrics* 1976;32(1):1-49
- Holt JD. Competing risk analyses with special reference to matched pair experiments. *Biometrika* 1978; 65: 159–165.
- Hoo ZH, Candlish J, Teare D: What is an ROC curve? *Emerg Med J*. 2017 Jun;34(6):357-359.
- J. Andrew Royle, Robert M. Dorazio, in *Hierarchical Modeling and Inference in Ecology*, 2009
- Kalbfleisch and Prentice, 1980, *The Statistical Analysis of Failure Time Data*, Second Edition

- Karel G M Moons, Patrick Royston, Yvonne Vergouwe, Diederick E Grobbee, Douglas G Altman: Prognosis and prognostic research: what, why, and how? *BMJ* 2009;338:b375.
- Kannel WB, McGee D, Gordon T. A general cardiovascular risk profile: the Framingham study. *Am J Cardiol* 1976;38:4651.
- Koller MT, Raatz H, Steyerberg EW, Wolbers M. Competing risks and the clinical community: irrelevance or ignorance? *Stat Med.* 2012;31:1089– 1097.
- Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *Am J Epidemiol.* 2009;170:244–256.
- Lawless JF. *Statistical Models and Methods for Lifetime Data.* New York, NY: John Wiley & Sons; 1982.
- Lee E. (2003). *Statistical Methods for Survival Data Analysis.*
- Lunn M and McNeil D. Applying Cox regression to competing risks. *Biometrics* 1995; 51: 524–532.
- M J Bradburn, T G Clark, S B Love, and D G Altman , *Survival Analysis Part III: Multivariate data analysis – choosing a model and assessing its adequacy and fit,* *Br J Cancer.* 2003; 89(4): 605–611.
- Mahmood Syed S., Levy Daniel, Vasan Ramachandran S., Wang Thomas J.. *The Framingham Heart Study and the Epidimiology of Cardiovascular Diseases: A Historical Prespective.* *Lancet* 2014;383(9921):998-1008.
- Mallett S, Royston P, Dutton S, Waters R, Altman DG. Reporting methods in studies developing prognostic models in cancer: a review. *BMC Med* 2010.
- Manish Kumar Goel, Pardeep Khanna, and Jugal Kishore, *Understanding survival analysis: Kaplan-Meier estimate,* *Int J Ayurveda Res.* 2010;1(4): 274–278.
- Miller, Rupert G. (1997), *Survival analysis,* John Wiley & Sons, ISBN 0-471-25218-2
- Moons KGM, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart* 2012;98:683-690.
- Mushkudiani NA, Hukkelhoven CW, Hernandez AV, Murray GD, Choi SC, Maas AI, Steyerberg EW, 2008. A systematic review finds methodological improvements necessary for prognostic models in determining traumatic brain injury outcomes. *J Clin Epidemiol,* 61(4):331-43

- Nashef SA, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European system for cardiac operative risk evaluation (EuroSCORE). *Eur J Cardiothorac Surg.* 1999;16(1):9-13.
- Papatheodoridis G, Dalekos G, Sypsa V, Yurdaydin C, Buti M, Goulis J, Calleja JL, Chi H, Manolakopoulos S, Mangia G, Gatselis N, Keskin O, Savvidou S, de la Revilla J, Hansen BE, Vlachogiannakos I, Galanis K, Idilman R, Colombo M, Esteban R, Janssen HL, Lampertico P. PAGE-B predicts the risk of developing hepatocellular carcinoma in Caucasians with chronic hepatitis B on 5-year antiviral therapy. *J Hepatol* 2016;64(4):800-6.
- Pearl, Judea; Bareinboim, Elias (2014). "External validity: From do-calculus to transportability across populations". *Statistical Science.* 29 (4): 579–595.
- Pencina, M. J. & D'Agostino, R. B. (2004), "Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation," *Statistics in Medicine*, 23, 2109–23.
- Pepe MS, Mori M. Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Stat Med.* 1993;12:737–751.
- Peto Richard; Peto Julian (1972). "Asymptotically Efficient Rank Invariant Test Procedures". *Journal of the Royal Statistical Society, Series A.* Blackwell Publishing. 135 (2): 185–207.
- Pokhrel A, Dyba T, Hakulinen T., A Greenwood formula for standard error of the age-standardised relative survival ratio, *Eur J Cancer.* 2008 Feb;44(3):441-7.
- Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med.* 2007;26:2389–2430.
- Rahman M, Gareth Ambler, Babak Choodari-Oskoei, and Rumana Z. Omar , Review and evaluation of performance measures for survival prediction models in external validation settings. *BMC Med Res Methodol.* 2017; 17: 60.
- Rajul Parikh, Annie Mathai, Shefali Parikh, G Chandra Sekhar and Ravi Thomas: Understanding and using sensitivity, specificity and predictive values, *Indian J Ophthalmol.* 2008 Jan-Feb; 56(1): 45–50.
- Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol* 2013;13:33.
- Royston P, Altman DG. Visualizing and assessing discrimination in the logistic regression model. *Stat Med.* 2010;29(24):2508-20.

- Royston P, Sauerbrei W. *Multivariable Model-building: A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Modelling Continuous Variables*. Chichester, England, Hoboken, NJ: John Wiley, 2008.
- Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med* 2006;25:127–141.
- Rubin Donald B. . *Multiple Imputation for Nonresponse in Surveys*. Book Series: Wiley Series in Probability and Statistics. 1987.
- Ruth Croxford, Institute for Clinical Evaluative Sciences: *Restricted Cubic Spline Regression: A Brief Introduction*, Paper 5621-2016
- Schmid, M., Potapov, S., 2012. A comparison of estimators to evaluate the discriminatory power of time-to-event models. *Statistics in Medicine* 31, 2588–2609.
- Schober P and Thomas R. Vetter, *Survival Analysis and Interpretation of Time-to-Event Data: The Tortoise and the Hare*, *Anesth Analg*. 2018 Sep; 127(3): 792–798.
- Schumacher M, Binder H, Gerds T. Assessment of survival prediction models based on microarray data. *Bioinformatics*. 2007;23(14):1768-74.
- Shmueli G. To explain or to predict? *Statist Sci* 2010 *Statistical Science* 2010, Vol. 25, No. 3, 289–310
- Smeeton, N.C. (1985). "Early History of the Kappa Statistic". *Biometrics*. 41 (3): 795.
- Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Rotterdam: Springer, 2009.
- Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014;35(29):1925-31.
- Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21:128–138.
- Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD: Internal validation of predictive models: efficiency of some procedures for logistic regression analysis, *J Clin Epidemiol*. 2001 Aug;54(8):774-81
- Stuart A., Ord J.K, Arnold S. (1999), *Kendall's Advanced Theory of Statistics: Volume 2A—Classical Inference and the Linear Model*, sixth edition, §20.2–20.3.

- Sullivan LM, Massaro JM, D'Agostino RB Sr. Presentation of multivariate data for clinical use: The Framingham Study risk score functions. *Stat Med*. 2004;23(10):1631-60.
- Tibshirani R. Regression and shrinkage via the Lasso. *J R Stat Soc Ser B* 1996;58:267–288.
- van Houwelingen HC, Thorogood J. Construction, validation and updating of a prognostic model for kidney graft survival. *Stat Med* 1995;14:1999–2008.
- van Houwelingen J. C. , Le Cessie S. . Predictive value of statistical models. *Stat Med*. 1990;9(11):1303-25.
- Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;26:565–574.
- Wolbers M, Koller MT, Wittman JC, Steyerberg EW. Prognostic models with competing risks: methods and application to coronary risk prediction. *Epidemiology* 2009; 20(4):555–561.
- Wong VW, Chan SL, Mo F, Chan TC, Loong HH, Wong GL, et al. Clinical scoring system to predict hepatocellular carcinoma in chronic hepatitis B carriers. *J Clin Oncol* 2010;28:1660–1665.
- Woodward M, Brindle P, Tunstall-Pedoe H; SIGN group on risk estimation. Adding social deprivation and family history to cardiovascular risk assessment: the ASSIGN score from the Scottish Heart Health Extended Cohort (SHHEC). *Heart* 2007;93(2):172-6.
- Yang HI, Yuen MF, Chan HL, Han KH, Chen PJ, Kim DY, Ahn SH, Chen CJ, Wong VW, Seto WK, Risk estimation for hepatocellular carcinoma in chronic hepatitis B (REACH-B): development and validation of a predictive score. 2011 Jun;12(6):568-74.
- Yishu Xue, Elizabeth D. Schifano: Diagnostics for the Cox model, *Communications for Statistical Applications and Methods* 2017; 24(6): 583-604.
- Yuen MF, Tanaka Y, Fong DY, Fung J, Wong DK, Yuen JC, et al. Independent risk factors and predictive score for the development of hepatocellular carcinoma in chronic hepatitis B. *J Hepatol* 2009;50:80–88.
- Ε. Σιβρίδης: Παθολογική Ανατομική, Α΄ τόμος 2007, Ιατρικές εκδόσεις Π.Χ Πασχαλίδης