Evidence synthesis methods for continuous outcomes

Papadimitropoulou, A.

ΠΕΡΙΛΗΨΗ ΣΤΑ ΕΛΛΗΝΙΚΑ

Μέτα-ανάλυση είναι η στατιστική μεθοδολογία για την ποσοτική σύνθεση δεδομένων από διαφορετικές πηγές με σκοπό: α) την ανίχνευση μεγεθών επίδρασης, β) την εκτίμηση των μεγεθών και της μεταβλητότητάς τους και γ) την ανάλυση παραγόντων (συν-μεταβλητές) που έχουν αντίκτυπο σε αυτά. Παρόλο που ο όρος επινοήθηκε από κοινωνικούς επιστήμονες τη δεκαετία του 1970, ο ευρύτερος όρος «σύνθεση έρευνας» προέρχεται από την προεδρική ομιλία του Λόρδου φυσικού Rayleigh το 1885 τονίζοντας την ανάγκη όχι μόνο για «νέο υλικό» αλλά και για την «πέψη και αφομοίωση του παλιού» (Rayleigh, 1885). Πάνω από έναν αιώνα αργότερα, μια γενική αναζήτηση του όρου «μέτα-ανάλυση» στο Google αποφέρει πάνω από 600 εκατομμύρια αποτελέσματα, ενώ η αντίστοιχη αναζήτηση στο PubMed αναφέρει τον όρο σε περισσότερες από 120000 δημοσιεύσεις. Είναι, λοιπόν, δίκαιο να πούμε ότι η σύσταση του Rayleigh έχει ευρέως υιοθετηθεί.

Η εισαγωγή επίσημων μεθοδολογιών σύνθεσης δεδομένων έχει οδηγήσει σε αλλαγή της προοπτικής των επιστημόνων σχετικά με την σημαντικότητα του αποτελέσματος της έρευνάς τους. Το αποτέλεσμα μιας μοναδικής μελέτης μπορεί πλέον να θεωρηθεί ως ένα στοιχείο συσσωρευμένης επιστημονικής έρευνας, παρά η αδιαμφισβήτη αλήθεια σε ένα επιστημονικό ερώτημα.

Επιπλέον, η εισαγωγή τεχνικών σύνθεσης δεδομένων έχει αναμφισβήτητα συμβάλει στην μετακίνηση της επιστημονικής κοινότητας προς τις αρχές της ανοικτής επιστήμης. Το γνωστό «πρόβλημα του συρταριού» αντιμετωπίζεται επίσης, καθώς οι επιστήμονες διαπίστωσαν ότι οι μελέτες με μη-στατιστικά σημαντικά αποτελέσματα είναι δυνητικά εξίσου πολύτιμες με εκείνες που παρουσιάζουν χαμηλές τιμές σημαντικότητας (Button et al., 2013).

Η ΔΙΑΤΡΙΒΗ

Αυτή η διατριβή περιγράφει νέα στατιστική μεθοδολογία και καινοτόμες εφαρμογές στον τομέα της σύνθεσης δεδομένων από συγκριτικές κλινικές μελέτες εστιάζοντας το ενδιαφέρον στα συνεχή δεδομένα. Ιδιαίτερη προσοχή εφιστάται στην ανάλυση συνεχών συγκεντρωτικών δεδομένων (δηλαδή στο επίπεδο της μελέτης) δημιουργώντας ψεύδο ατομικά (προσωπικά) δεδομένα συμμετεχόντων στις μελέτες, έτσι ώστε η εκ νέου ανάλυσή τους να μπορεί να χρησιμοποιήσει το πλαίσιο των γραμμικών μικτών μοντέλων. Αυτή η νέα προσέγγιση αξιοποιεί την ιεραρχική φύση των μέτα-αναλυτικών δεδομένων και την πληθώρα στατιστικών μεθόδων και επιλογών λογισμικού που διατίθενται για τα γραμμικά μικτά μοντέλα. Επιπλέον, βασιζόμενοι σε εμπειρικά δεδομένα παρέχουμε συστάσεις σχετικά με την προτιμώμενη προσέγγιση για τη μέτα-ανάλυση συνεχών δεδομένων που μετρήθηκαν πριν και μετά την ιατρική παρέμβαση. Αυτή η διατριβή ασχολείται επίσης με το ζήτημα των χαμένων δεδομένων στη μέτα-ανάλυση δικτύου (network meta-analysis), όπου προτείνουμε το πλαίσιο των μοντέλων μίξης υποδείγματος/μοτίβου (pattern mixture model). Παρέχουμε λεπτομερή συζήτηση για κάθε κεφάλαιο στις επόμενες παραγράφους.

Το Κεφάλαιο 1 παρουσιάζει μια γενική εισαγωγή στο πλαίσιο της σύνθεσης δεδομένων, που ορίζεται ως ένα σύνολο στατιστικών τεχνικών με σκοπό τη σύνθεση (συνδυασμό και σύνοψη) πολλαπλών ποσοτικών αποτελεσμάτων. Ξεκινάμε επισημαίνοντας την ανάγκη για το ευρύτερο αυτό πλαίσιο και την ευρεία υιοθέτησή του από διαφορετικούς κλάδους. Το υπόλοιπο μέρος της εισαγωγής χωρίζεται σε δύο μέρη: τις μεθοδολογίες μέτα-ανάλυσης και μέτα-ανάλυσης δικτύου.

Εισάγουμε την έννοια της μέτα-ανάλυσης και τους διαφορετικούς τύπους της με βάση το ερευνητικό ζήτημα που μας ενδιαφέρει. Επιπλέον, συζητάμε εν συντομία τις κύριες δομές δεδομένων για τη μέτα-ανάλυση: τα συγκεντρωτικά δεδομένα, τα οποία είναι διαθέσιμα ως σύνοψη δεδομένων στο επίπεδο της μελέτης και τα ατομικά (προσωπικά) δεδομένα στο επίπεδο των συμμετεχόντων από πολλαπλές μελέτες. Εισάγουμε τη σχετική ορολογία και τα δύο θεμελιώδη στατιστικά μοντέλα: τα μοντέλα σταθερών και τυχαίων επιδράσεων, των οποίων η διαφορά έγκειται στην υπόθεση ότι τα πραγματικά υποκείμενα μεγέθη επίδρασης είναι πανομοιότυπα σε όλες τις μελέτες (συνεπώς κοινά) ή προέρχονται από μια κατανομή πραγματικών μεγεθών επίδρασης, αντίστοιχα. Ένα βασικό στοιχείο που επισημαινεται είναι ότι ο στατιστικός έλεγχος για ετερογένεια δε θα πρέπει να καθοδηγεί την απόφαση επιλογής μεταξύ των δύο προαναφερθέντων μοντέλων. Η απόφαση θα πρέπει να βασίζεται αποκλειστικά στο ερώτημα ποιο μοντέλο ταιριάζει στην κατανομή των παρατηρούμενων μεγεθών επίδρασης. Τις περισσότερες φορές το μοντέλο τυχαίων επιδράσεων είναι καταλληλότερο. Συζητάμε συνοπτικά τα πλεονεκτήματα της μέτα-ανάλυσης που χρησιμοποιεί τα ατομικά δεδομένα στο επίπεδο των συμμετεχόντων σε σύγκρση με τη μέτα-ανάλυση συγκεντρωτικών δεδομένων. Ενώ η πρώτη προσέγγιση είναι αρκετά απαιτητική, προτιμάται από την ανάλυση των συγκεντρωτικών δεδομένων, με την προϋπόθεση ότι είναι δυνατή η πρόσβαση στα ατομικά δεδομένα στο επίπεδο του ασθενών.

Στο δεύτερο μέρος της εισαγωγής περιγράφουμε την επέκταση της μέτα-ανάλυσης σε τεχνικές που επιτρέπουν τη σύγκριση πολλαπλών θεραπείων, οι οποίες δεν έχουν μελετηθεί ταυτόχρονα σε συγκριτικές κλινικές μελέτες, δηλαδή τις έμμεσες συγκρίσεις θεραπείας και τη μέτα-ανάλυση δικτύου. Αυτές οι μέθοδοι σύνθεσης δεδομένων έχουν χαρακτηριστεί ως «νέας γενιάς εργαλεία για τη σύνθεση πληροφορίας» και έχουν αναπτυχθεί ευρέως τα τελευταία 15 χρόνια. Εξηγούμε τις υποθέσεις αυτών των μεθόδων, τη μεταβατικότητα (transitivity), και την στατιστική παρουσίαση της μεταβατικότητας (τη συμφωνία της άμεσης και έμμεση πληροφορίας (consistency)) και περιγράφουμε μια μπεϋζιανή (Bayesian) προσέγγιση στη μέτα-ανάλυση δικτύου.

Το Κεφάλαιο 2 παρουσιάζει μια καινούργια μέθοδο μέτα-ανάλυσης συνεχών δεδομένων από μελέτες που συγκρίνουν δύο ιατρικές παρεμβάσεις (θεραπείες). Η μέθοδος που προτείνουμε στηρίζεται στις ιδιότητες των επαρκών στατιστικών, που είναι στατιστικά στοιχεία που μεταφέρουν όλη τη βασική πληροφορία από ένα δείγμα για μια παράμετρο του πληθυσμού. Στην περίπτωση της μέτα-ανάλυσης δεδομένων που ακολουθούν την κανονική κατανομή, τα επαρκή στατιστικά είναι οι μέσες τιμές, οι τυπικές αποκλίσεις και τα μεγέθη των δύο δειγμάτων, τα οποία είναι συχνά διαθέσιμα ως τα συγκεντρωτικά δεδομένα από δημοσιευμένες μελέτες. Μπορούμε, τότε, να κατασκευάσουμε ψεύδο ατομικά δεδομένα των συμμετεχόντων με τις ίδιες μέσες τιμές, τυπικές αποκλίσεις και μεγέθη. Όταν αναλύσουμε τα ψεύδο ατομικά δεδομένα συμμετεχόντων με ένα γραμμικό μικτό μοντέλο, η πιθανοφάνεια τους θα είναι ίδια με την πιθανοφάνεια των πραγματικών ατομικών δεδομένων των συμμετεχόντων. Παρέχουμε τον αλγόριθμο για την δημιουργία

αυτών των ψεύδο ατομικών δεδομένων μαζί με κώδικα σε διάφορα στατιστικά πακέτα και γλώσσες προγραμματισμού.

Όταν έχουμε κατασκευάσει τα ψεύδο ατομικά δεδομένα, τότε αυτά μπορούν να αναλυθούν με τον ίδιο τρόπο που θα αναλύαμε τα πραγματικά ατομικά δεδομένα, χρησιμοποιώντας γραμμικά μικτά μοντέλα. Υπάρχει πληθώρα διαθέσιμων μοντέλων που περιγράφονται εκτενώς στο Κεφάλαιο 2. Η χρήση των γραμμικών μικτών μοντέλων προσφέρει σημαντικά πλεονεκτήματα σε σχέση με τις παραδοσιακά χρησιμοποιούμενες προσεγγίσεις που βασίζονται σε συγκεντρωτικά δεδομένα στο επίπεδο των μελετών· πιο συγκεκριμένα, επιτρέπει πιο σύνθετες/ρεαλιστικές υποθέσεις για τις εντός-των-μελετών διακύμανσεις των καταλοίπων. Η μεθολογία που αναπτύσσουμε σε αυτό το κεφάλαιο παρουσιάζεται μέσα από δύο παραδείγματα πραγματικών δεδομένων που προέρχονται από μια μέτα-ανάλυση που συγκρίνει τα επίπεδα πλάσματος μεταξύ ασθενών με νόσο Αλτσχάιμερ και υγιών συμμετεχόντων. Συνολικά, 12 μοντέλα ψεύδο ατομικών δεδομένων σε αύξουσα πολυπλοκότητα προσαρμόζονται στα δεδομένα και συγκρίνονται με τις παραδοσιακές μεθόδους συγκεντρωτικών δεδομένων. Παρουσιάζεται επίσης μια μελέτη προσομοίωσης, στην οποία προσομοιώνουμε και κατόπιν αναλύουμε πραγματικά και ψεύδο ατομικά δεδομένα συμμετεχόντων. Μελετάμε τις επιδράσεις του αριθμού των μελετών και των μεγεθών δείγματος ανά ομάδα συμμετεχόντων σε κάθε μελέτη, στις πιθανότητες κάλυψης, τη μεροληψία και το μέσο τετραγωνικό σφάλμα (MSE) για τη μέθοδο των ψεύδο ατομικών δεδομένων και τη μέθοδο τυχαίων επιδράσεων με/χωρίς την τροποποίηση των Hartung-Knapp. Τα αποτελέσματα της προσέγγισης των πραγματικών ατομικών δεδομένων είναι σε μεγάλο βαθμό πανομοιότυτα με τα αποτελέσματα των ψεύδο ατομικών δεδομένων, εκτός από τις λίγες περιπτώσεις προβλημάτων απόκλισης των μοντέλων. Γενικά, η προσέγγιση των ψεύδο ατομικών δεδομένων έχεις καλές ιδιότητες, με πιθανότητες κάλυψης κοντά στο ονομαστικό επίπεδο, σε σύγκριση με τη μέθοδο DL, η οποία παράγει πιθανότητες κάλυψης κάτω από το ονομαστικό επίπεδο. Επιπλέον, οι προσομοιώσείς μας έδειξαν μικρότερες τιμές MSE σε σύγκριση με την προσέγγιση DL και παρόμοιες τιμές με τη διόρθωση REML HK κάτω από διάφορα σενάρια.

Στο Κεφάλαιο 3 προτείνουμε μια επέκταση της μεθόδου των ψεύδο ατομικών δεδομένων στην περίπτωση όπου τα συνεχή δεδομένα μετρώνται πριν και μετά την ιατρική παρέμβαση σε μια κλινική μελέτη. Τυπικά η μέτα-ανάλυση αυτών των δεδομένων περιλαμβάνει συνήθως τον υπολογισμό και την σύνοψη μέσων διαφορών των τελικών τιμών (μετρήσεις μετά την ιατρική παρέμβαση) ή των μεταβολών των τιμών πριν (αρχικές) και μετά την παρέμβαση (τελικές). Υποστηρίζουμε ότι μια πιο κατάλληλη προσέγγιση, η οποία λαμβάνει υπόψη τις πιθανές ανισορροπίες στις αρχικές τιμές και τη συσχέτιση των αρχικών και τελικών τιμών, είναι η ανάλυση συνδιακύμανσης (ANCOVA) σε ψεύδο ατομικά δεδομένα, με βάση τα κατάλληλα επαρκή στατιστικά στοιχεία. Τα επαρκή στατιστικά στοιχεία σε αυτήν την περίπτωση είναι: οι μέσες τιμές των αρχικών και τελικών τιμών με τις τυπικές τους αποκλίσεις σε κάθε ομάδα συμμετεχόντων, και η συσχέτιση μεταξύ αρχικών και τελικών τιμών. Το ψεύδο ατομικά δεδομένα, λοιπόν, μπορούν να αναλυθούν χρησιμοποιώντας είτε μοντέλα σε ένα στάδιο στρωματοποιημένα κατά μελέτη ή μοντέλα τυχαίων επιδράσεων ανά μελέτη σε δύο στάδια.

Η μέθοδος των ψεύδο ατομικών δεδομένων επιτρέπει επιπρόσθετα την ενσωμάτωση του αποτελέσματος της αλληλεπίδρασης της θεραπείας με την αρχική τιμή (τιμή βάσης), δίνοντας έτσι τη δυνατότητα εξερεύνησης διαφορετικών εκβάσεων στη θεραπεία. Τα πλεονεκτήματα της προσέγγισης που αναπτύσσουμε παρουσιάζονται μέσα από δύο παραδείγ-

ματα. Στο πρώτο παράδειγμα, χρησιμοποιούμε τα συγκεντρωτικά δεδομένα και δείχνουμε ότι ANCOVA ψεύδο ατομικών δεδομένων προσφέρει πανομοιότυπα αποτελέσματα με την ANCOVA πραγματικών ατομικών δεδομένων όπως εφαρμόστηκε από Riley et al., 2013. Στο δεύτερο παράδειγμα, παρατηρείται ανισορροπία στις τιμές βάσεις και διαφορετική απόκριση στη θεραπεία. Η προσέγγιση της ανάλυσης συνδιακύμανσης ψεύδο ατομικών δεδομένων λαμβάνει υπόψη την ανισορροπία στις τιμές βάσεις και παρέχει ακριβέστερες εκτιμήσεις του μεγέθους επίδρασης της θεραπείας. Επιπλέον, διαπιστώνεται στατιστικά σημαντική αλληλεπίδραση μεταξύ των τιμών βάσης (αρχικές) και του μεγέθους επίδρασης της θεραπείας, η οποία μπορεί να γίνει αντιληπτή σε μια ανάλυση μέτα-παλινδρόμησης.

Το Κεφάλαιο 4 ασχολείται με τις κατάλληλες και λιγότερο κατάλληλες μέτα-αναλυτικές μεθόδους για συγκεντρωτικά συνεχή δεδομένα που μετρώνται πριν και μετά την ιατρική παρέμβαση. Το βασικό παράδειγμα είναι ένα σύνολο δεδομένων από εννέα μελέτες που καταγράφουν μετρήσεις βάρους, πριν και μετά την πρόσληψη συμπληρωμάτων ασβεστίου, όπου παρατηρήθηκε σημαντική ανισορροπία στις τιμές βάσης (αρχικές μετρήσεις βάρους). Αυτό το σύνολο δεδομένων αναλύθηκε αρχικά χρησιμοποιώντας μια προσέγγιση ανάλυσης συνδιακύμανσης σε συγκεντρωτικά δεδομένα, η οποία είναι επιρρεπής σε οικολογική μεροληψία (ecological bias).

Εισάγουμε διάφορες μεθόδους μέτα-ανάλυσης συγκεντρωτικών συνεχών δεδομένων: α) μια γνωστή, αλλά παραβλεπόμενη στη βιβλιογραφία, προσέγγιση ανάκτησης αποτελεσμάτων ANCOVA από δημοσιευμένα συγκεντρωτικά δεδομένα υποθέτοντας ίσες διακυμάνσεις μεταξύ των δύο ομάδων πριν και μετά τη θεραπεία και β) την προσέγγιση των ψεύδο ατομικών δεδομένων, όπου αρχικά κατασκευάζουμε τα ψεύδο δεδομένα και στη συνέχεια τα αναλύουμε χρησιμοποιώντας ανάλυση συνδιακύμανσης σε ένα ή δύο στάδια. Συνολικά, οι μέθοδοι που χρησιμοποιούν την ανάλυση συνδιακύμανσης προτιμώνται έναντι των συνηθισμένων μεθόδων μέτα-ανάλυσης που χρησιμοποιούν τελικές τιμές ή μεταβολές τιμών. Η μέθοδος των ψεύδο ατομικών δεδομένων ενός σταδίου προσφέρει πρόσθετα οφέλη έναντι της ανάκτησης αποτελεσμάτων ανάλυσης συνδιακύμανσης, καθώς είναι δυνατό να διαχωριστούν οι επιδράσεις εντός των μελετών και μεταξύ των μελετών και μπορεί να διαχειριστεί την αλληλεπίδραση μεταξύ θεραπείας και αρχικών τιμών. Συνιστάται επίσης η προσέγγιση των ψεύδο ατομικών δεδομένων σε δύο στάδια καθώς ουσιαστικά προσφέρει παρόμοια πλεονεκτήματα με την προσέγγιση σε ένα στάδιο και απαιτεί λιγότερη στατιστική τεχνογνωσία. Χρησιμοποιούμε ένα παράδειγμα μέτα-ανάλυσης όπου λείπει ένα μεγάλο μέρος των συγκεντρωτικών στατιστικών στοιχείων και καθοδηγούμε τον ενδιαφερόμενο ερευνητή βήμα προς βήμα σε μια σταδιακή διαδικασία προεπεξεργασίας δεδομένων ακολουθούμενη από την κατάλληλη ανάλυση.

Στο Κεφάλαιο 5, παρουσιάζουμε ένα διαδικτυακό εργαλείο για τη μέτα-ανάλυση μελετών με συνεχή δεδομένα (λαμβάνοντας υπόψη τις αρχικές τους τιμές, τιμές βάσης), γεγονός που καθιστά ευκολότερη στο ευρύτερο κοινό την πρόσβαση στις αναλύσεις. Αυτό το διαδραστικό λογισμικό προορίζεται να συνδυάσει την ευελιξία του πλαισίου των γραμμικών μικτών μοντέλων με παραδοσιακές προσεγγίσεις για τη μέτα-ανάλυση συνεχών δεδομένων, προσφέροντας πολλές επιλογές μοντελοποίησης. Σε προηγούμενα κεφάλαια, αναφέραμε ότι οι μέθοδοι που χρησιμοποιούν τις τελικές τιμές και τις μεταβολές τιμών, θα αποφέρουν γενικά λιγότερο ακριβείς εκτιμήσεις για το συγκεντρωτικό αποτέλεσμα της θεραπείας και ότι οι μέθοδοι που στηρίζονται στην ανάλυση συνδιακύμανσης προτιμώνται με ή χωρίς την ύπαρξη ανισορροπίας στις αρχικές τιμές.

Η εφαρμογή `MA-cont:pre/post effect size` είναι εύκολη στη χρήση τόσο σε γνώστες του αντικειμένου αλλά και στο ευρύ κοινό. Όπως συμβαίνει με κάθε λογισμικό, ο χρήστης θα πρέπει να αντιμετωπίζει την εφαρμογή με κριτικό μυαλό και επιπροσθέτως ενθαρρύνουμε τους μη τεχνικούς χρήστες να ζητούν συμβουλές από στατιστικούς, όταν χρειάζεται για την ερμηνεία των ευρημάτων.

Τα Κεφάλαια 6 και 7 ασχολούνται με μεθόδους μέτα-ανάλυσης δικτύου.

Το Κεφάλαιο 6 παρουσιάζει μια εφαρμογή συνθέσης δεδομένων, όπου διεξάγουμε μια συστηματική ανασκόπηση και μπεϋζιανή μέτα-ανάλυση δικτύου για την αξιολόγηση της σχετικής αποτελεσματικότητας φαρμακολογικών και μη φαρμακολογικών παρεμβάσεων σε καταθλιπτικούς ασθενείς, ανθεκτικούς στη θεραπεία. Αυτή η εφαρμογή είναι η πρώτη σύνθεση διαφορετικών τύπων παρεμβάσεων (φαρμακολογικές έναντι σωματικών), υπό την υπόθεση των «μεταβατικών» εικονικών θεραπειών/πλασέμπο, η οποία διερευνάται διεξοδικά στην παράγραφο 6.3.3.

Η κύρια συμβολή αυτής της εργασίας είναι ότι παρέχει ιεράρχηση ιατρικών παρεμβάσεων, οι οποίες είναι απίθανο να συγκριθούν μια-προς-μια σε μια τυχαιοποιημένη μελέτη. Οι ανθεκτικοί στη θεραπεία καταθλιπτικοί ασθενείς είναι μια ομάδα ασθενών με σοβαρή κατάθλιψη, για τους οποίους μια πληθώρα διαθέσιμων επιλογών μπορεί να μην είναι χρήσιμες και είναι επιτακτική ανάγκη οι επαγγελματίες υγείας να ενημερώνονται για την ιεράρχηση των θεραπειών με βάση τη σχετική αποτελεσματικότητα.

Στο Κεφάλαιο 7 εισάγουμε μια μέθοδο που βασίζεται στο μοντέλο μίξης υποδείγματος σε ένα στάδιο για να αντιμετωπίσουμε το ζήτημα των χαμένων (ελλιπών) συνεχών δεδομένων στη μέτα-ανάλυση δικτύου. Τα χαμένα δεδομένα είναι συχνό φαινόμενο στο επίπεδο της μελέτης και ως εκ τούτου έχουν αντίκτυπο στις εκτιμήσεις σύνθεσης δεδομένων, συχνά οδηγώντας σε μεροληπτικά συμπεράσματα όταν οι ερευνητές αδυνατούν να τα συνυπολογίσουν στην ανάλυση τους. Εισάγουμε το μοντέλο μίξης υποδείγματος σε ένα στάδιο, το οποίο μας επιτρέπει να μελετήσουμε πόσο «απέχουν» οι χαμένες παρατηρήσεις από την υπόθεση ότι λείπουν τυχαία (missing-at-random MAR). Οι «αποστάσεις» ποσοτικοποιούνται από τις ενημερωτικές παραμέτρους έλλειψης (IMDoM IMRoM), οι οποίες είναι η διαφορά και ο λόγος, αντίστοιχα, της μη παρατηρούμενης μέσης τιμής στα χαμένα δεδομένα και της μέσης τιμής στα παρατηρούμενα δεδομένα.

Αντί να υποθέτουμε σταθερές τιμές για τις ενημερωτικές παραμέτρους έλλειψης (δεδομένων), μπορούμε να υποθέσουμε ότι ακολουθούν μια εκ-των-προτέρων κατανομή και να εξετάσουμε μια πληθώρα επιλογών, για παράδειγμα οι παράμετροι είναι σχετικές με την ιατρική παρέμβαση, την κλινική δοκιμή ή είναι κοινές εντός του δικτύου των παρεμβάσεων. Μεταξύ των πλεονεκτημάτων του μοντέλου μίξης υποδείγματος σε ένα στάδιο είναι η δυνατότητα να μάθουμε για τα χαμένα δεδομένα συνδυάζοντας δεδομένα από μελέτες με διαφορετικά κλάσματα έλλειψης δεδομένων σε ένα μόνο βήμα. Η προτεινόμενη μέθοδός μας παρουσιάζεται μέσα από δύο δίκτυα που περιλαμβάνουν σημαντικό αριθμό χαμένων δεδομένων στις κλινικές μελέτες τους. Σε αυτές τις εφαρμογές, μαθαίνουμε ελάχιστα για τις ενημερωτικές παραμέτρους έλλειψης, με εξαίρεση έναν συνδυασμό θεραπείας, όπου θα μπορούσαμε να υποστηρίξουμε ότι οι συμμετέχοντες που υποβλήθηκαν σε αυτή τη θεραπεία αλλά δεν ολοκλήρωσαν τη μελέτη (και συνεπώς τα δεδομένα τους λείπουν), τείνουν να έχουν μικρότερη μείωση χρόνου με συμπτώματα σε σχέση με τους συμμετέχοντες που ολοκλήρωσαν τη μελέτη.

ΚΥΡΙΑ ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΟΠΤΙΚΕΣ ΓΙΑ ΜΕΛΛΟΝΤΙΚΗ ΕΡΕΥΝΑ

Οι μέθοδοι σύνθεσης συνεχών δεδομένων έχουν λάβει λιγότερη μεθοδολογική προσοχή σε σύγκριση με αυτές των κατηγορικών δυαδικών δεδομένων, καθώς τα συνεχή δεδομένα ερευνώνται λιγότερο συχνά σε συστηματικές ανασκοπήσεις. Σε αυτή τη διατριβή, προτείνουμε στατιστική μεθοδολογία που επιτρέπει την ανάλυση κυρίως συνεχών μέτα-αναλυτικών δεδομένων με πιο εξελιγμένες μεθόδους, χρησιμοποιώντας ταυτόχρονα ευρέως διαδεμένο στατιστικό λογισμικό.

Προτείνουμε τη μέθοδο των ψεύδο ατομικών δεδομένων για συνεχή δεδομένα που μετρώνται μία φορά (μετά τη θεραπεία) και την επέκτασή της ώστε να ενσωματώνει τιμές πριν από τη θεραπεία (τιμές βάσης), επιτρέποντας στους ερευνητές να αναλύουν τα συγκεντρωτικά δεδομένα χωρίς να υποθέτουν ότι οι διακυμάνσεις των καταλοίπων έχουν γνωστές και σταθερές τιμές, το οποίο συμβαίνει στις κλασικές μέτα-αναλυτικές μεθόδους που στηρίζονται σε συγκεντρωτικά δεδομένα. Η μέθοδος των ψεύδο ατομικών δεδομένων, προσφέρει λοιπόν, τη δυνατότητα εκτίμησης των εντός των μελετών διακυμάνσεων των καταλοίπων, υποθέτοντας ρεαλιστικά σενάρια. Απλούστερες περιπτώσεις, όπου οι εντός των μελετών διακυμάνσεις των καταλοίπων μπορούν να θεωρηθούν ότι διαφέρουν ανάλογα με την θεραπεία, αλλά όχι με τη μελέτη μπορούν να ερευνηθούν με τον έλεγχο λόγου πιθανοφανειών. Αυτός ο έλεγχος μεταφράζεται στη διερεύνηση ετερογενών αποκρίσεων στη θεραπεία, εάν για παράδειγμα η διακύμανση στην ομάδα που υποβλήθηκε σε θεραπεία είναι μεγαλύτερη από ό,τι στην ομάδα ελέγχου.

Η μέθοδος των ψεύδο ατομικών δεδομένων προσφέρει εύκολη ρύθμιση στους βαθμούς ελευθερίας του παρονομαστή (της στατιστικής συνάρτησης του συγκεντρωτικού μεγέθους επίδρασης της θεραπείας), όπου αντικαθιστά την κανονική κατανομή με την t-κατανομή με συγκεκριμένους βαθμούς ελευθερίας χρησιμοποιώντας λογισμικό για γραμμικά μικτά μοντέλα. Έτσι τα διαστήματα εμπιστοσύνης για τις παραμέτρους ενδιαφέροντος, κυρίως το μέγεθος επίδρασης της θεραπείας, θα είναι πιο μεγάλα και επομένως πιο συντηρητικά.

Η επέκταση της μεθόδου των ψεύδο ατομικών δεδομένων ώστε να ενσωματώσει τις τιμές βάσης (αρχικές τιμές) παρουσιάζει ιδιαίτερο ενδιαφέρον καθώς η βιβλιογραφία στη σύνθεση συνεχών δεδομένων δεν είναι ομόφωνη σχετικά με μια προτιμώμενη ιεραρχία ανάλυσης, ιδιαίτερα μεταξύ των προσεγγίσεων που χρησιμοποιούν τελικές τιμές ή μεταβολές τιμών. Ελλείψει ατομικών δεδομένων, η σύνθεση εκτιμητών ANCOVA από όλες τις μελέτες είναι ομόφωνα η καταλληλότερη προσέγγιση, ωστόσο τέτοιες εκτιμήσεις σπάνια δημοσιεύονται στην πράξη. Η μέθοδος που προτείνουμε και η προσέγγιση των ανακτημένων εκτιμήσεων ANCOVA, παρέχουν άμεση λύση σε αυτό το πρόβλημα και ελπίζουμε ότι ενδεχόμενως να ξεκινήσουν ένα κίνημα αλλαγής ώστε να πραγματοποιούνται πιο πολλές ANCOVA μέτα-αναλύσεις στο μέλλον. Ο ANCOVA εκτιμητής είναι πιο αποτελεσματικός από τους εκτιμητές της διαφοράς των τελικών τιμών και των μεταβολών τιμών, αλλά ευελπιστούμε ότι η εισαγωγή της μεθόδου ANCOVA ψεύδο ατομικών δεδομένων θα καταστήσει το δίλημμα μεταξύ τελικών και μεταβολών τιμών, περιττό. Η μέθοδος των ψεύδο ατομικών δεδομένων επιτρέπει επιπλέον τη διερεύνηση των επιδράσεων αλληλεπίδρασης θεραπείας και τιμής βάσης, η οποία είναι στατιστικά ο καταλληλότερος τρόπος για να διερευνηθεί το αντίκτυπο των χαρακτηριστικών στο επίπεδο του ασθενούς στη θεραπεία και δεν υποφέρει από οικολογική μεροληψία και/ή χαμηλή ισχύ όπως η τεχνική της μέτα-παλινδρόμησης στο επίπεδο της μελέτης.

Μια βασική συμβολή της προσέγγισης των ψεύδο ατομικών δεδομένων είναι ότι αντιμετωπίζει το θέμα της πρόσβασης στα πραγματικά ατομικά δεδομένα, η οποία έγινε ακόμη πιο δύσκολη από τότε που τέθηκε σε ισχύ ο νέος Γενικός Κανονισμός για την Προστασία των Δεδομένων το 2018. Ενώ ο κανονισμός αποσκοπεί στην προστασία των ιδιωτικών δεδομένων, περιπλέκει αρκετά την πρόσβαση ακόμη και σε ανώνυμα ατομικά δεδομένα. Η μέθοδος των ψεύδο ατομικών δεδομένων απλοποιεί αυτό το ζήτημα καθώς βασίζεται εξ ολοκλήρου σε πληροφορίες, οι οποίες μπορούν να δημοσιοποιηθούν χωρίς ανησυχίες περί απορρήτου, τα επαρκή στατιστικά στοιχεία. Επιπλέον, η προτεινόμενη μέθοδος θα μπορούσε να λειτουργήσει κάτω από ένα πλαίσιο όπου κάποιες μελέτες δημοσιεύουν συγκεντρωτικά δεδομένα και άλλες ατοιμικά δεδομένα, δημιουργώντας ψεύδο ατομικά για τις πρώτες υποομάδες μελετών και στοιβάζοντάς τα με τα πραγματικά ατομικά δεδομένα των υπόλοιπων μελετών.

Στην απλούστερη περίπτωση, που εξετάζουμε στο Κεφάλαιο 2, δεν υπάρχουν διαθέσιμες τιμές βάσης (ή δεν λαμβάνονται υπόψη) και έτσι η «φυσική» μέτα-αναλυτική εκτιμήτρια είναι η διαφορά των τελικών τιμών. Η προτεινόμενη σε ένα στάδιο μέθοδος ψεύδο ατομικών δεδομένων εκτιμά αυτή τη μέση διαφορά ως τον συντελεστή παλινδρόμησης της μεταβλητής της θεραπείας στο γραμμικό μικτό μοντέλο. Αυτό το μέτρο μεγέθους επίδρασης είναι χρήσιμο όταν τα αποτελέσματα μετρούνται στην ίδια κλίμακα σε όλες τις μελέτες, αλλά υπάρχουν περιπτώσεις όπου χρησιμοποιούνται διαφορετικές κλίμακες για τη μέτρηση του ίδιου αποτελέσματος, για παράδειγμα, τη σοβαρότητα της κατάθλιψης χρησιμοποιώντας κλίμακες HAM_D ή MADRS. Το κατάλληλο μέτρο μεγέθους επίδρασης σε αυτές τις περιπτώσεις είναι η τυποποιημένη μέση διαφορά (SMD). Η επέκταση της μεθόδου των ψεύδο ατομικών δεδομένων στις τυχαιοποιημένες μέσες διαφορές είναι δυνατή μετά από απαραίτητους μετασχηματισμούς των δεδομένων πριν την ανάλυση. Εάν δημιουργηθεί το σύνολο των ψεύδο ατομικών δεδομένων, τότε, ανά μελέτη, η μέση τιμή του αποτελέσματος στην ομάδα ελέγχου πρέπει να αφαιρεθεί από τη μέση τιμή του αποτελέσματος στην ομάδα που υποβλήθηκε στη θεραπεία. Αυτή η διαφορά στη συνέχεια διαιρείται με μια κοινή τυπική απόκλιση, συνήθως τη συγκεντρωτική τυπική απόκλιση μεταξύ των δύο ομάδων. Ωστόσο, δεν είναι απλό, καθώς υπάρχουν διάφοροι εκτιμητές (SMD), όπως Glass' Δ, Cohen's d, Hedges' g και η βιβλιογραφία είναι ασυνεπής όσον αφορά την εκτίμηση της διακύμανσης του SMD. Υποθέτοντας ότι κάποιος έχει λάβει το μετασχηματισμένο αποτέλεσμα, επιλέγοντας μεταξύ των εκτιμητών (SMD) στο επίπεδο συμμετεχόντων, τότε ένα μοντέλο μικτών επιδράσεων μπορεί να εφαρμοστεί παρόμοια με τα μοντέλα που παρουσιάζονται στο στο Κεφάλαιο 2 υποθέτοντας σταθερές και/ή τυχαίες επιδράσεις για την τυποποιημένη μέση τιμή στην ομάδα ελέγχου και στην ομάδα θεραπείας. Περισσότερες λεπτομέρειες μπορείτε να βρείτε στα άρθρα των Goldstein et al., 2000 και Walwyn and Roberts, 2017, που αναφέρονται σε μοντέλα επιδράσεων στρωματοποίησης.

Η μέθοδος των ψεύδο ατομικών δεδομένων του Κεφαλαίου 2 μπορεί να επεκταθεί στη μέτα-ανάλυση δικτύου σχετικά εύκολα, προσθέτοντας σταθερούς και τυχαίους όρους στο γραμμικό μικτό μοντέλο για κάθε επιπρόσθετη θεραπεία. Επιπλέον, ενώ η μέθοδος προτείνεται για τον παράλληλο σχεδιασμό δύο ομάδων σε συγκριτικές κλινικές μελέτες, μπορεί να προσαρμοστεί σε διαφορετικό πειραματικό σχεδιασμό — τον σχεδιασμό διασταύρωσης δύο ομάδων, αν είναι διαθέσιμη η πληροφορία της συσχέτισης των τιμών των δύο θεραπειών για κάθε συμμετέχοντα.

Επεκτάσεις της ANCOVA μεθόδου ψεύδο ατομικών δεδομένων που περιγράφονται στο Κεφάλαιο 3 στη μέτα-ανάλυση δικτύων και στον σχεδιασμό διασταύρωσης δύο θε-

ραπειών είναι επίσης πιθανές. Ένα μελλοντικό ερευνητικό θέμα που πρέπει να εξεταστεί είναι η ενσωμάτωση άλλων συν-μεταβλητών εκτός της τιμής βάσης, υποθέτοντας ότι θα είναι διαθέσιμα τα κατάλληλα συγκεντρωτικά δεδομένα. Σε αυτή την περίπτωση, τα επαρκή στατιστικά στοιχεία θα είναι οι μέσες τιμές και ο πίνακας διακύμανσης-συνδιακύμανσης για κάθε ομάδα θεραπείας.

Τέλος, η διατριβή πάσχει από τους ίδιους εγγενείς περιορισμούς του πλαισίου σύνθεσης δεδομένων, όπως η μεροληψία δημοσίευσης. Αυτή η μεροληψία μπορεί να προκύψει όταν η μετα-ανάλυση (δικτύου) δεν περιέχει όλα τα δεδομένα σχετικά με το ερευνητικό πρόβλημα αλλά απλώς ένα υποσύνολο τους, που συχνά αποτελείται από μελέτες που παρουσιάζουν στατιστικά σημαντικά αποτελέσματα. Σε αυτή την περίπτωση, μια μετα-ανάλυση τέτοιων μελετών μπορεί να παρέχει ένα στρεβλό μέγεθος επίδρασης της θεραπείας, το οποίο δεν μπορεί να διορθωθεί με τις προτεινόμενες μεθόδους μας. Επιπλέον, τα συνδυασμένα αποτελέσματα που εκτιμώνται σε μέτα-αναλύσεις (δικτύου) μπορεί να είναι πιο ακριβή από αυτά που εκτιμήθηκαν στις πρωτογενείς μελέτες, υπό την προϋπόθεση ότι οι μελέτες που συμβάλλουν σε αυτά είναι υψηλής ποιότητας.

Αυτή η διατριβή προσφέρει πρόοδο στον τομέα της σύνθεσης συνεχών δεδομένων προσφέροντας λύση σε βασικά προβλήματα, όπως τη σύνθεση δεδομένων όταν οι τιμές βάσης διαφέρουν μεταξύ των συγκρινόμενων ομάδων και την απώλεια δεδομένων των συμμετεχόντων. Φιλοδοξούμε ότι η ANCOVA ψεύδο ατομικών δεδομένων θα γίνει η παραδοσιακή μέθοδος μέτα-ανάλυσης τέτοιων δεδομένων στο μέλλον, και ότι η R-Shiny εφαρμογή που αναπτύξαμε ενθαρρύνει τους χρήστες προς αυτή την κατεύθυνση.