



FH Salzburg
MultiMediaTechnology

Einführung in die Multiple Imputation im Medizinischen Bereich

Seminararbeit

AutorIn: Christoph Marko

BetreuerIn: DI Dr. Simon Ginzinger

Repository: <https://github.com/Chris392/template-thesis-latex>

Salzburg, Österreich, 05.07.2021

Einführung in die Multiple Imputation im Medizinischen Bereich

Christoph Marko

cmarko.mmt-b2019@fh-salzburg.ac.at

FH Salzburg

1 EINLEITUNG

In der Medizinischen Forschung ist es wichtig, dass Studienergebnisse mit fehlenden Daten sorgfältig analysiert und bearbeitet werden. Das soll verhindern, dass es zu einer Verzerrung des Ergebnisses kommt. Viele Medizinische Zeitschriften verlangen bereits einen korrekten Umgang mit Fehlenden Daten sowie eine Rechtfertigung für den gewählten statistischen Ansatz. (Lee und Simpson 2014, 163) Eine wichtige Methode zur Behandlung von fehlenden Daten ist die Multiple Imputation (MI).

Ziel dieser Arbeit ist es, einen Überblick über die MI im Medizinischen Bereich zu geben. Neben einer Erklärung der MI, wird auf wichtige Fragestellungen bei der Durchführung der MI eingegangen.

Ein wichtiger Teil bei der Durchführung der MI ist die Planung der MI und die Überprüfung des Datensatzes. Dabei wird im Forschungsblatt der Anteil von fehlenden Daten sowie die Art der fehlenden Daten erläutert. Des Weiteren muss die Methode der Imputation gewählt werden. Es werden einige Alternativen zur MI aufgezeigt und es wird auf den Vorteil von Hilfsvariablen bzw. Post-Baseline Untersuchungen eingegangen. Nach der Durchführung kann das Ergebnis der MI überprüft werden. Hierbei werden im Forschungsblatt unterschiedliche Mittel zur Qualitätsfeststellung beschrieben.

2 MULTIPLE IMPUTATION

2.1 Anwendungsbeispiel der MI

Angenommen es würde eine Studie über den Zusammenhang zwischen Blutdruck und dem Risiko eines Herzinfarkts geben. Dabei fehlen die Blutdruck Daten bei einigen Herzinfarkt-Patienten. Man kann nun aufgrund der anderen Daten die gesammelt wurden Rückschluss auf die fehlenden Daten ziehen. So können z.B. die fehlenden Blutdruck-Daten einiger Patienten ersetzt werden, indem andere Patienten mit gleichen Eigenschaften gesucht werden. Bei Patienten mit gleichen Eigenschaften (z.B. gleiches Alter, gleicher BMI) steigt auch die Wahrscheinlichkeit, dass der Blutdruck-Wert ähnlich ist.

Allerdings kann diese Vorgehensweise aufwändig werden. Vor allem, wenn es sehr viele Variablen gibt. Es wird z.B. schwierig werden, einen Patienten mit genau demselben BMI, Alter, Blutdruck etc. zu finden. Daher wird in der MI versucht, anhand der vorhandenen Daten und einer Verteilungsfunktion eine Vermutung über den fehlenden Wert zu bekommen. Man könnte z.B. davon ausgehen, dass die Blutdruckwerte der Bevölkerung Normalverteilt sind. Aufgrund dieser Vermutung kann der Mittelwert aller Blutdruckwerte gebildet und

für die fehlenden Daten eingesetzt werden. Für Komplexere Sachverhalte ist dieses Vorgehen allerdings nicht ausreichend. (Donders u. a. 2006, 1089)

Um den gesamten Sachverhalt besser darstellen zu können, werden z.B. mehrere voneinander abhängige Regressionsmodelle (Chained Equations) oder eine Mehrdimensionale Normalverteilung (Multivariate Normal Imputation) benötigt. (Spratt u. a. 2010, 489) Damit werden mehrere Eigenschaften berücksichtigt. Für die Studie könnten die Daten z.B. mithilfe von Chained Equations imputiert werden. Im Folgenden wird die Chained Equation Methode anhand eines einfachen Beispiels näher erläutert:

In der Chained Equation Methode werden mehrere Imputationen auf einzelnen Variablen durchgeführt. Angenommen es gibt die Variablen BMI, Alter und Blutdruck. In allen Variablen gibt es fehlende Daten. Zu Beginn kann für die fehlenden Daten z.B. einfach der Mittelwert eingesetzt werden. Also für alle fehlenden BMI-Daten wird der Mittelwert der BMI-Werte eingesetzt. Dasselbe passiert für Alter und Blutdruck. Danach wird auf jede Variable jeweils eine Imputation durchgeführt. Zuerst können z.B. die fehlenden BMI-Daten imputiert werden. Dazu werden zuerst die bereits vorhandenen imputierten Werte (im Falle des ersten Durchlaufs die Mittelwerte) der BMI-Variable gelöscht. Es wird daraufhin ein Regressionsmodell mit den BMI-Werten im Verhältnis zu Alter und Blutdruck erstellt. Angenommen es wird der BMI-Wert eines 60-Jährigen mit einem Blutdruckwert von 140-90 gesucht. Anhand des Regressionsmodells kann man nun z.B. ablesen, dass der BMI in Bezug auf Alter und Blutdruck 27 beträgt. Dieser Wert wird nun für den fehlenden Wert eingesetzt. Dasselbe passiert mit dem Alter im Bezug auf BMI und Blutdruck und dem Blutdruck in Bezug auf Alter und BMI. Dadurch entsteht ein neuer Datensatz mit den imputierten Werten. Auf diesen neuen Datensatz wird derselbe Vorgang nun mehrmals wiederholt. (Tilling u. a. 2016, 108) Das mehrfache Wiederholen des Vorgangs erhöht die Genauigkeit der imputierten Werte.

2.2 Funktionsweise der MI

Ausgangspunkt der MI ist das Analysis Modell. Das Analysis Modell ist ein Datensatz, in welchem Daten verschiedener Variablen fehlen können. In der ersten Phase der MI wird aus dem Analysis Modell ein Imputation Modell erstellt. Dazu werden dem Analysis Model noch zusätzliche Variablen, sogenannte auxiliary variables (Hilfsvariablen) hinzugefügt. Die fehlenden Daten werden daraufhin z.B. mit Hilfe der Chained Equations mehrfach imputiert. Durch die mehrfache Imputation entstehen mehrere Vervollständigte Datensätze. In der zweiten Phase werden die Vervollständigten Datensätze ana-

lysiert und gemittelt, wodurch der finale Datensatz ermittelt wird. Durch die Berechnung des Standard Errors kann die Unsicherheit der Daten ermittelt werden. (Lee und Simpson 2014, 163) Im Folgenden wird auf die beiden Phasen der MI, sowie auf die Begriffe Standarderror und Bias, näher eingegangen.

2.3 Phase 1: Mehrfache Imputation

Die Grundsätzliche Idee der MI ist es, ein Regressionsmodell zu erstellen, dass mögliche Werte für die fehlenden Werte beinhaltet. Das Regressionsmodell beinhaltet also eine Reihe an Werten, die für die fehlenden Werte plausibel sind. Aus diesem Regressionsmodell werden in Folge anhand einer Simulation randomisiert Werte gezogen. Angenommen, es würde sich um eine Normalverteilung handeln, dann würden die fehlenden Daten vereinfacht über folgende Formel aus Enders (2017, 8) gezogen werden:

$$Y_{(mis)} \sim N(\hat{Y}, \sigma_e^2) \quad (1)$$

„ $Y_{(mis)}$ ist ein Schätzwert für den fehlenden Wert. $\sim N$ bezeichnet die Normalverteilung, \hat{Y} ist der vorhergesagte Wert aus einer Regressionsgleichung und σ_e^2 ist die Residuenvarianz aus der Regression.“ (Enders 2017, 8) Die Normalverteilung stellt die Verteilung einer Variable, z.B. des Blutdrucks im Verhältnis zu den abhängigen Variablen dar. Die Varianz definiert den Mittelwert sowie die Streuung. Insgesamt ist die Imputation aus einer Gleichung die Summe eines Abweichungswertes (Varianz) und des vorhergesagten Wertes. (Enders 2017, 8). Der Schätzwert für einen fehlenden Wert wird also berechnet, indem man einen Wert entsprechend der abhängigen Variablen aus dem Bereich einer Gleichung zieht. Der Bereich wird dabei durch die Varianz bestimmt, die die Unsicherheit der Daten angibt. Das Zufällige ziehen von Schätzungen zu den fehlenden Werten wird mehrfach wiederholt, wodurch mehrere Vollständige Datensätze entstehen. Auf denen wird daraufhin die Analyse durchgeführt. (Lee und Simpson 2014, 163 ff.) Durch das Wiederholen der Imputationen mit unterschiedlichen, plausiblen Regressionsparametern, gleichen die Datensätze keiner Imputation der anderen. (Enders 2017, 8) Ein Grundprinzip der MI ist es also, mehrere Annahmen über die vorhandenen Daten zu machen,

2.4 Phase 2: Analyse der Daten

In der 2. Phase wird jeder der Imputierten Datensätze getrennt betrachtet. Wenn z.B. 20 Imputationen durchgeführt wurden, gibt es 20 Datensätze, die Schätzwerte für die fehlenden Werte beinhalten. Für jeden dieser Datensätze wird ein gemeinsamer Schätzwert (z.B. über das Arithmetische Mittel) und der Standard Error berechnet. Zuletzt werden die 20 Datensätze noch kombiniert, indem das Arithmetische Mittel der Schätzwerte sowie des Standard Errors gebildet wird. (Enders 2017, 10)

2.5 Standarderror

Der Standarderror ist ein Maß für die Unsicherheit der geschätzten Werte. (Donders u. a. 2006, 1098) Er ist somit auch ein Indikator für den Informationsverlust durch fehlende

Daten. (Hughes u. a. 2019, 1298) Angenommen es gibt in einer Studie fehlende Blutdruckwerte. Die Blutdruckwerte werden nun 3x mithilfe verschiedener Regressionsmodellen imputiert. Es entstehen somit 3 komplette Datensätze mit jeweils unterschiedlichen imputierten Blutdruckwerten. Ist der Mittelwert der imputierten Blutdruckwerte zwischen den 3 kompletten Datensätzen sehr unterschiedlich, dann wird der Standarderror hoch sein. Daraus kann man schließen, dass es eine hohe Unsicherheit in der Genauigkeit der Blutdruckdaten gibt und somit womöglich ein Bias vorhanden ist.

2.6 Bias

Unter Bias versteht man einen systematischen Fehler. Dieser Fehler entsteht, wenn der Einfluss von Eigenschaften über- oder unterbewertet wird. (Jakobsen u. a. 2017, 2) Angenommen, der BMI spielt keine Rolle bei der Entstehung von Herz-Kreislauferkrankungen. Sollte eine Studie den BMI als großen Faktor für die Entstehung von Herz-Kreislauferkrankungen miteinbeziehen, ist das Ergebnis vermutlich biased bzw. voreingenommen.

2.7 Praktische Anwendung der MI

Statistische Programme wie SAS und Stata unterstützen bereits MI unter der Verwendung des Conditional Regression Modells sowie der Joint Normal Distribution. (Lee und Simpson 2014, 164). Da dieses Paper einen Allgemeinen Überblick über die MI geben soll, wird auf die genaue Implementierung nicht mehr eingegangen.

Eine Implementierung der MI kann in Royston (2005) gefunden werden. Dort wird die Implementierung der MI in der MICE Methode näher erläutert. Die MICE-Methode basiert auf Chained Equations. Die Implementierung wird anhand von Ado-Dateien beschrieben. Ado-Dateien sind Dateien, welche vom Statistik Programm Stata verwendet werden.

2.8 Vorgehensweise MI

Bevor die MI durchgeführt werden kann. Sollten einige wichtige Punkte geklärt werden, die in den nächsten Kapiteln näher erläutert werden. So stellt sich unter anderem die Frage, wie groß der Anteil an fehlenden Daten ist und welche Auswirkungen dies hat. (Siehe Kapitel 3). Nach der Durchführung der MI sollte zu dem die Qualität der MI überprüft werden (siehe Kapitel 7).

3 ANTEIL AN FEHLENDEN DATEN

Es gibt in der Forschung unterschiedliche Empfehlungen, bis zu welchem Anteil an fehlenden Daten zuverlässige Ergebnisse erzielt werden können. Statistische Leitartikel gehen davon aus, dass es ab 10% an fehlenden Daten wahrscheinlich zu einem Bias kommen kann. Ab 40% sollte das Ergebnis der MI nur mehr als Hypothese gesehen werden. Die Beweislage, die diese Grenzen stützt, ist allerdings gering. (Madley-Dowd u. a. 2019, 63 f.)

In der Studie “The proportion of missing data should not be used to guide decisions on multiple imputation” wird argumentiert, dass der Anteil an fehlenden Daten nicht unbedingt ausschlaggebend für den Erfolg oder Misserfolg der MI steht. Viel mehr spielen die Art der fehlenden Daten und die Verwendung von Hilfsvariablen eine wichtige Rolle.

4 ART DER FEHLENDEN DATEN

Nach Definition von (Little & Rubin, 2002; Rubin, 1976) gibt es 3 verschiedene Arten von fehlenden Daten:

4.1 MCAR: Missing Completely at Random

Die Fehlenden Daten haben keine Verbindung zu den zu analysierenden Daten. Gründe für MCAR Daten können zum Beispiel ein Strom- oder Internetausfall sein. (Enders 2017, 8) Angenommen es kommt zu einem 24-Stündigem Stromausfall. Das Resultat daraus ist, dass 24-Stunden lang keine Ergebnisse von Untersuchungen gespeichert werden können. Bei dem Stromausfall handelt es sich um ein komplett zufälliges Event, dass vermutlich alle Typen von Patienten (alt, jung, groß, klein, ...) gleichermaßen betrifft. Man spricht daher bei den fehlenden Daten von MCAR Daten.

4.2 MAR: Missing at Random

Systematische Unterschiede zwischen den fehlenden Daten und den Untersuchten Daten können erklärt werden. So können zum Beispiel Blutdruckwerte v.a. bei jüngeren Personen fehlen, da bei diesen der Blutdruck weniger oft kontrolliert wird. (Sterne u. a. 2009, 157) An den vorhandenen Daten kann man nun erkennen, dass es bei Menschen mit Jüngerem Alter öfter zu fehlenden Blutdruckdaten kommt. Man könnte daher repräsentativ für die fehlenden Blutdruckdaten, die Blutdruckdaten anderer Junger Teilnehmer verwenden, bei welchen die Messung des Blutdrucks nicht vergessen wurde.

4.3 NMAR: Not Missing at Random

Es gibt systematische Unterschiede, welche nicht anhand der Untersuchten Daten erkannt werden können. NMAR Daten können zum Beispiel entstehen, wenn Patienten mit hohem Blutdruck häufiger aufgrund von Kopfschmerzen fehlen. (Sterne u. a. 2009, 157) Anhand der vorhandenen Daten kann man nun nicht erkennen, dass es eine Gruppe von Menschen mit Kopfschmerzen gibt, welche bei den Blutdruckmessungen fehlt.

NMAR-Daten können anhand von statistischen Tests nicht identifiziert werden. NMAR-Daten können also nur als solche durch vorhandenes Wissen interpretiert werden. Daher gibt es keine allgemein Gültige Methode für die Behandlung von NMAR-Daten. (Donders u. a. 2006, 1088)

4.4 Auswirkungen auf das Ergebnis der MI

MAR- und NMAR Daten sind schwer zu unterscheiden. NMAR können jedoch zu einem voreingenommenen Resultat führen. Um das Risiko eines voreingenommenen Ergebnisses durch NMAR Daten zu verringern, kann eine Sensitivitätsanalyse angewendet werden. (Sterne u. a. 2009, 157)

Wenn es sich um MCAR Daten handelt, können anstatt der MI einfachere Methoden wie die Complete Case Analysis in Betracht gezogen werden. (2006, 1088)

5 ALTERNATIVEN ZUR MI

5.1 Complete Case Analysis (CCA)

Bei der CCA werden Teilnehmer mit fehlenden Daten aus der Studie exkludiert. (Jakobsen u. a. 2017, 3)

In einer Simulation aus Heijden u. a. (2006) erzielte die CCA schlechtere Ergebnisse als die MI. Für die Simulation wurden Daten aus einer Studie zu Lungenembolie mit 398 Subjekten verwendet. Dabei gab es einen Unterschiedlichen Anteil an fehlende Daten bei einigen Variablen. Bei den fehlenden Daten handelte es sich nicht um MCAR Daten. Das Ergebnis der CCA schnitt dabei in allen Messgrößen im Vergleich zur single- und multiplen Imputation schlechter ab. Als Qualitätskriterien wurden u.a. der Standard Error sowie die Regressionskoeffizienten verwendet. Mögliche Qualitätskriterien werden in Kapitel XY nochmals näher erläutert.

Wenn es sich im Datensatz wie in der oben genannten Simulation nicht um MCAR-Daten handelt, dann ist das Ergebnis der CCA vermutlich voreingenommen. Grund dafür ist, dass die CCA systematische Unterschiede nicht berücksichtigt. (Sterne u. a. 2009, 157) Im Fall von MAR-Daten eignet sich daher die Single- oder Multiple Imputation besser. (Donders u. a. 2006, 1088)

5.2 Single Imputation

Bei der Single Imputation werden fehlende Daten nach einer Regel ersetzt. Von der Single Imputation gibt es mehrere Varianten. Bei der Last Observation Carried Forward (LOCF) werden fehlende Daten mit den Daten aus der letzten Untersuchung ersetzt. Bei der Worst Observation Carried Forward werden fehlende Daten mit dem schlechtesten untersuchten Ergebnis des Patienten ersetzt. Bei der Simple Mean Imputation wird der Mittelwert verwendet. Die Single Imputation hat die Annahme, dass fehlende Werte und in der Vergangenheit gemessene Werte identisch sind. Diese Annahme ist allerdings unrealistisch, weshalb diese Methode mit Bedacht eingesetzt werden soll. (Jakobsen u. a. 2017, 3)

Eine Simulation aus Young-Saver u. a. (2018) zeigt, dass mit der LOCF Methode ähnlich gute oder sogar bessere Ergebnisse erzielt werden können als mit der MI. In der Simulation wurden Daten aus einer Studie zu akutem Schlaganfall verwendet. Begründet wird das gute Abschneiden der LOCF u.a. mit der zeitlichen Nähe der letzten Untersuchung zu dem zu schätzen den Wert. Durch die zeitliche Nähe haben zeitrelevante Prozesse, wie Neuroplastizität oder Neuronale Regeneration (neuronale Prozesse) keinen so großen Einfluss. Zudem ist die Wahrscheinlichkeit von Vorfällen geringer, welche Einfluss auf die zu erhebenden Werte haben könnten. Dazu gehören z.B. Unfälle oder Herzinfarkte. (Young-Saver u. a. 2018, 3667). Wie gut die LOCF Methode im Vergleich zur MI abschneidet kann jedoch nicht allgemein gesagt werden und hängt maßgeblich von den verwendeten Daten ab.

5.3 Wann bietet sich der Einsatz von CCA oder Single Imputation an?

Die Anwendung der CCA bietet sich an, wenn es sich um MCAR-Daten handelt. Die Single Imputation sollte mit Bedacht eingesetzt werden. Sie kann jedoch unter Umständen ähnliche gute Ergebnisse wie die MI erzielen. Die LOCF Methode bietet sich vor allem an, wenn es relevante, zeitnahe Untersuchungen gibt und es sich bei den untersuchten Variablen um Zeitrelevante Prozesse handelt.

6 ZUSÄTZLICHE DATEN

Zusätzliche Daten können den Informationsgehalt erhöhen und damit zur Reduzierung des Bias beitragen.

6.1 Post-Baseline Untersuchungen

Die Simulation aus Young-Saver u. a. (2018) zum Thema akutem Schlaganfall zeigt die Bedeutung von Post-Baseline Untersuchungen. Post-Baseline Untersuchungen sind Untersuchungen, die nach einem Bestimmtem Ereignis erhoben wurden. Z.B. nach einem vorgefallenem Schlaganfall. Im Falle der Studie wurde der NIH Stroke Scale nach 24 Stunden und der mRS nach 7-10 Tagen erhoben. (der NIH Stroke Scale und der mRS sind beides Bewertungsskalen für die Symptome des Patienten). Die MI wurde einmal mit und einmal ohne Daten aus den Post-Baseline Untersuchungen durchgeführt. Im Ergebnis der Simulation schnitt die Multiple Imputation mit Post-Baseline besser ab als ohne. (siehe Abbildung 1)

6.2 Hilfsvariablen

Unter Hilfsvariablen (engl.: auxiliary variables) versteht man Variablen, welche dem Datensatz hinzugefügt werden, um bei der Imputation zu helfen. (Graham, Olchowski und Gilreath 2007, 211) Durch Hilfsvariablen wird die Menge an Information erhöht, wodurch im Optimalfall die Imputation verbessert werden soll. Hilfsvariablen können z.B. auch aus anderen Studien/Datensätzen herangezogen werden.

Eine Simulation in Madley-Dowd u. a. (2019) zeigt die Bedeutung der Hilfsvariablen. Die Simulation wurde anhand von Daten aus ALSPAC (Avon Longitudinal Study of Parents and Children) durchgeführt. In der ALSPAC Studie wurde die Auswirkung von Rauchen während der Schwangerschaft auf den Intelligenzquotienten des Kindes untersucht. Als Variablen wurde der IQ im Alter von 15 Jahren und ein Binärwert für das Rauchen während der Schwangerschaft verwendet. Als Hilfsvariablen wurde der IQ im Alter von 8 Jahren, die Verständlichkeit und der Redefluss im Alter von 9 Jahren, ein Binärwert für das Vorhandensein von Lernschwierigkeiten sowie die Mathematik- und Lesefähigkeiten des Kindes im 6. Schuljahr verwendet. Unter der Miteinbeziehung der Hilfsvariablen bei der MI verringerte sich der Standard Error im Vergleich zur MI ohne Hilfsvariablen von 0.504 auf 0.461. Der FMI (Fraction of Missing Information) verringerte sich von 0.724 auf 0.664. (Madley-Dowd u. a. 2019, 71) Der FMI ist ein Parameter, der den Verlust an Informationen darstellt. Der FMI kann zwischen 0 und 1 liegen, ein niedriger FMI ist anzustreben. (Madley-Dowd u. a. 2019, 65) Eine nähere Erläuterung des FMIs kann in Madley-Dowd u. a. (2019) gefunden werden.

Man sieht anhand der Reduktion von FMI und Standarderror, dass sich Hilfsvariablen positiv auf die MI auswirken können. Ausschlaggebend für den Effekt von Hilfsvariablen ist u.a. ob die Hilfsvariable genügend Information über die fehlenden Werte bereitstellt. (Hughes u. a. 2019, 1302) Ein Indikator für den Effekt einer Hilfsvariable auf die MI ist die Reduktion des Standarderrors.

In der Studie von Madley-Dowd u. a. (2019) reduzierte sich der Standarderror bei der Hinzugabe der Hilfsvariable "IQ im Alter von 8 Jahren" um 7.65%. Wurde die Hilfsvariable

"Learning Disabilities" hinzugefügt, erhöhte sich der Standarderror um 0,8%. (Madley-Dowd u. a. 2019, 71) Dies zeigt, dass nicht alle Hilfsvariablen zwingend den Informationsgehalt und damit die Genauigkeit des Ergebnisses verbessern. Eine sorgfältige Überprüfung/Auswahl von Hilfsvariablen ist daher von Vorteil.

Neben dem Standarderror, gibt es noch weitere Qualitätskriterien für die MI. Einige davon werden im nächsten Kapitel (Kapitel 7) näher beschrieben.

7 QUALITÄTSFESTSTELLUNG MI

7.1 Simulierte fehlende Daten und P-Value

Eine einfache Methode zur Qualitätsüberprüfung ist es, einen vollständigen Datensatz zu nehmen und die fehlenden Daten zu simulieren. Nach dieser Vorgehensweise wurde auch die Simulation aus (D. Young-Saver, 2018) durchgeführt. Hierbei wurde aus dem Vollständigen Datensatz auf Basis eines randomisierten Verfahrens Daten gelöscht. Daraus folgt, dass man nun die imputierten Daten mit den Daten des Vollständigen Datensatzes vergleichen kann. Ein möglicher Vergleich wäre z.B. die Bestimmung der exakten Übereinstimmungsrate. Hierbei wird der Anteil von exakten Übereinstimmungen zwischen den imputierten und originalen Werten bestimmt. Des Weiteren kann Spearmans Korrelationskoeffizient, oder die Mittlere (Absolute) Abweichung bestimmt werden. (Young-Saver u. a. 2018, 3664)

Ein weiterer Wert, mit dem der Unterschied zwischen vollständigem und imputiertem Datensatz verglichen werden kann, ist der P-Value. Der P-Value gibt an, wie groß der Unterschied zwischen zwei Datensätzen ist. Je nachdem, wie groß der gewählte Threshold ist, kann daraufhin eine Aussage darüber getroffen werden, ob die beiden Datensätze signifikant oder nicht signifikant unterschiedlich sind.

Signifikante Unterschiede zwischen vollständigem und imputiertem Datensatz sind ein Indikator für einen vorhandenen Bias.

7.2 Sensitivitätsanalyse

Vor der Durchführung der MI sollte eine Vermutung über die Art der fehlenden Daten gemacht werden. Sind die Daten MAR, MNAR oder MCAR? Angenommen es gibt eine Studie, die den Zusammenhang zwischen Blutdruck und Herz-Kreislauferkrankungen untersucht. Sollte ein Teil der Patienten diese Studie vorzeitig verlassen, könnte man vermuten, dass dies komplett zufällig geschieht. Die Daten wären somit MCAR. Es könnte aber auch sein, dass einige Patienten die Studie aufgrund von Kopfschmerzen wegen zu hohen Blutdrucks verlassen. Damit gäbe es einen systematischen Zusammenhang und die Daten wären MAR oder MNAR. In vielen Studien ist es schwer, eine sichere Vermutung über die Art der fehlenden Daten zu treffen. Mit Sensitivitätsanalysen können mehrere Vermutungen über die Art der fehlenden Daten überprüft werden. (Cro u. a. 2020, 2815)

Im Allgemeinen gibt es 2 Varianten der Sensitivitätsanalyse: δ -basiert und referenzbasiert. In der δ -basierten Variante wird dem zu erwarteten Wert noch ein Offset Variable δ (Delta) hinzugefügt, welche den Einfluss von fehlenden/nicht-beobachteten Teilnehmern bewertet. In der Referenzbasierten Sensitivitätsanalyse werden Werte aus

Missing Data Handling Method	Disability-Free (mRS 0-1) Actual p value with no missing data: 0.014					Reduced Disability (all 7 mRS ranks) Actual p value with no missing data: 0.035				
	5% Missing	10% Missing	15% Missing	20% Missing	25% Missing	5% Missing	10% Missing	15% Missing	20% Missing	25% Missing
No data used from cases with missing final outcome										
Complete Case Analysis	0.022	0.028	0.042	0.051	0.059	0.045	0.049	0.067	0.074	0.094
Worst Case Imputation	0.024	0.032	0.049	0.061	0.073	0.069	0.088	0.154	0.160	0.245
Baseline data only used from cases with missing final outcome										
Multiple Imputation – Baseline Variables Only	0.022	0.032	0.049	0.058	0.079	0.046	0.054	0.072	0.079	0.105
Post-baseline data used from cases with missing final outcome										
Last Observation Carried Forward	0.015	0.015	0.015	0.015	0.013	0.036	0.033	0.036	0.036	0.037
Multiple Imputation – Baseline + Post-Baseline Variables	0.017	0.018	0.020	0.021	0.022	0.036	0.035	0.037	0.039	0.043

Abbildung 1: Die MI mit Post-Baseline Variablen (orange umrahmt) schnitt bei jedem Prozentsatz an fehlenden Daten besser ab als die MI ohne Post Baseline Variablen (blau umrahmt). Die Simulation wurde mittels simulierter fehlender Daten durchgeführt. Berechnet wurde der P-Value. Eine nähere Erläuterung dazu kann in Kapitel 7 gefunden werden. Grafik aus Young-Saver u. a. (2018, 3667), die Beschriftung sowie die blaue/orange Markierung wurde nachträglich hinzugefügt.

anderen Gruppen des Versuchs imputiert, welche Referenz zu den Beobachteten Werten beinhaltet. (Cro u. a. 2020, 2815) Eine praktische Beschreibung der Durchführung der Sensitivitätsanalyse in Verbindung mit der MI kann in Cro u. a. (2020) gefunden werden.

8 FAZIT

Eine sorgfältige Planung und Durchführung der Imputation ist wichtig, damit es nicht zu verzerrten Ergebnissen kommt. Im ersten Schritt sollte der verwendete Datensatz überprüft werden. Bei den Verwendeten Daten sollte der Anteil an fehlenden Daten nicht zu hoch sein. In der Forschung gibt es jedoch Unterschiedliche Vorgaben, welche Menge an fehlenden Daten in Ordnung ist. Mehr Forschung wäre hier von Vorteil. Bei der Art der fehlenden Daten sollte es sich möglichst um MAR Daten handeln.

Im Zweiten Schritt sollte die zu verwendende Methode für die Behandlung fehlender Daten, sowie die Verwendung von zusätzlichen Variablen geplant werden. Handelt es sich um MCAR Daten, kann der Einsatz einer CCA überlegt werden. Gibt es relevante, zeitnahe Untersuchungen zu den fehlenden Daten, kann auch die Verwendung der Single Imputation in der LOCF-Variante gute Ergebnisse erzielen. Das Hinzufügen von Post-Baseline Untersuchungen oder Hilfsvariablen kann das Resultat der MI verbessern.

Nach der Durchführung der MI ist es zu empfehlen, eine Sensitivitätsanalyse durchzuführen. Vor allem, wenn es sich nicht sicher um MAR Daten handelt. Der Standard Error sollte nicht zu hoch sein, da hohe Werte auf einen Bias hinweisen können.

MI ist eines der wichtigsten Werkzeuge für die Behandlung von fehlenden Daten in vielen verschiedenen Bereichen. Wer sich z.B. mit Statistik oder AI beschäftigt, sollte einen Überblick über MI haben. MI ist ein komplexes Thema und auch dieses Forschungsblatt konnte nur einen kleinen Überblick zum Thema MI geben. Es ist daher von Vor-

teil, dass die meisten Statistischen Programme bereits MI unterstützen und so den Einstieg in die MI vereinfachen.

LITERATURVERZEICHNIS

- Cro, Suzie, Tim P. Morris, Michael G. Kenward und James R. Carpenter. 2020. »Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: A practical guide«. *Statistics in Medicine* 39 (21): 2815–2842. ISSN: 10970258. <https://doi.org/10.1002/sim.8569>.
- Donders, A. Rogier T., Geert J.M.G. van der Heijden, Theo Stijnen und Karel G.M. Moons. 2006. »Review: A gentle introduction to imputation of missing values«. *Journal of Clinical Epidemiology* 59 (10): 1087–1091. ISSN: 08954356. <https://doi.org/10.1016/j.jclinepi.2006.01.014>.
- Enders, Craig K. 2017. »Multiple imputation as a flexible tool for missing data handling in clinical research«. *Behaviour Research and Therapy* 98:4–18. ISSN: 1873622X. <https://doi.org/10.1016/j.brat.2016.11.008>.
- Graham, John W., Allison E. Olchowski und Tamika D. Gilreath. 2007. »How many imputations are really needed? Some practical clarifications of multiple imputation theory«. *Prevention Science* 8 (3): 206–213. ISSN: 13894986. <https://doi.org/10.1007/s11121-007-0070-9>.
- Heijden, Geert J.M.G. van der, A. Rogier T. Donders, Theo Stijnen und Karel G.M. Moons. 2006. »Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: A clinical example«. *Journal of Clinical Epidemiology* 59 (10): 1102–1109. ISSN: 08954356. <https://doi.org/10.1016/j.jclinepi.2006.01.015>.

Hughes, Rachael A., Jon Heron, Jonathan A.C. Sterne und Kate Tilling. 2019. »Accounting for missing data in statistical analyses: Multiple imputation is not always the answer«. *International Journal of Epidemiology* 48 (4): 1294–1304. ISSN: 14643685. <https://doi.org/10.1093/ije/dyz032>.

Diese Arbeit hat folgende Länge (gezählt mit texcount):
detailtexcountbody

Jakobsen, Janus Christian, Christian Gluud, Jørn Wetterslev und Per Winkel. 2017. »When and how should multiple imputation be used for handling missing data in randomised clinical trials - A practical guide with flowcharts«. *BMC Medical Research Methodology* 17 (1): 1–10. ISSN: 14712288. <https://doi.org/10.1186/s12874-017-0442-1>.

Lee, Katherine J., und Julie A. Simpson. 2014. »Introduction to multiple imputation for dealing with missing data«. *Respirology* 19 (2): 162–167. ISSN: 13237799. <https://doi.org/10.1111/resp.12226>.

Madley-Dowd, Paul, Rachael Hughes, Kate Tilling und Jon Heron. 2019. »The proportion of missing data should not be used to guide decisions on multiple imputation«. *Journal of Clinical Epidemiology* 110:63–73. ISSN: 18785921. <https://doi.org/10.1016/j.jclinepi.2019.02.016>.

Royston, Patrick. 2005. »Multiple imputation of missing values: Update of ice«. *Stata Journal* 5 (4): 527–536. ISSN: 1536867X. <https://doi.org/10.1177/1536867X0500500404>.

Spratt, Michael, James Carpenter, Jonathan A.C. Sterne, John B. Carlin, Jon Heron, John Henderson und Kate Tilling. 2010. »Strategies for multiple imputation in longitudinal studies«. *American Journal of Epidemiology* 172 (4): 478–487. ISSN: 00029262. <https://doi.org/10.1093/aje/kwq137>.

Sterne, Jonathan A.C., Ian R. White, John B. Carlin, Michael Spratt, Patrick Royston, Michael G. Kenward, Angela M. Wood und James R. Carpenter. 2009. »Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls«. *BMJ (Online)* 339 (7713): 157–160. ISSN: 17561833. <https://doi.org/10.1136/bmj.b2393>.

Tilling, Kate, Elizabeth J. Williamson, Michael Spratt, Jonathan A.C. Sterne und James R. Carpenter. 2016. »Appropriate inclusion of interactions was needed to avoid bias in multiple imputation«. *Journal of Clinical Epidemiology* 80:107–115. ISSN: 18785921. <https://doi.org/10.1016/j.jclinepi.2016.07.004>.

Young-Saver, Dashiell F., Jeffrey Gornbein, Sidney Starkman und Jeffrey L. Saver. 2018. »Handling of Missing Outcome Data in Acute Stroke Trials: Advantages of Multiple Imputation Using Baseline and Postbaseline Variables«. *Journal of Stroke and Cerebrovascular Diseases* 27 (12): 3662–3669. ISSN: 15328511. <https://doi.org/10.1016/j.jstrokecerebrovasdis.2018.08.040>.