universität wien

Diplomarbeit

Titel der Arbeit

Software für Meta-Analysen: Systematischer Review und Usability-Studie

Verfasserin

Sandra Geislinger

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, im März 2012

Studienkennzahl:	298
Studienrichtung:	Psychologie
Betreuer:	Assistenzprof. Privatdoz. MMag. DDDr. Martin Voracek

Danksagung

Diese Diplomarbeit ist Dominik gewidmet, der mich nicht nur bei der Erstellung dieser Arbeit, sondern auch während den letzten Jahre meines Studiums immer unterstützt hat.

Außerdem möchte ich meinem Betreuer Assistenzprof. Privatdoz. MMag. DDDr. Martin Voracek und auch Mag. Jakob Pietschnig für die professionelle Hilfestellung und die anregenden und hilfreichen Gespräche danken.

Zudem gilt mein Dank meinen Freundinnen Petra, Martina und Mimoza, ohne deren Hilfe und Geduld ich diese Arbeit nicht so schnell fertigstellen hätte können, und auch meinen Eltern, die mich während meines Studiums unterstützt haben.

Herzlichen Dank dafür!

INHALTSVERZEICHNIS

VORWORT 7

I. REVIEW 9

1. META-ANALYSEN: THEORETISCHER HINTERGRUND 10

1.1 Geschichte 10
1.2 Methoden 10
1.3 Vorteile 14
1.4 Probleme 15
1.5 Software 16

2. PROGRAMMÜBERSICHT 17

2.1 Softwaresuche 17
2.2 Beurteilung der numerischen und grafischen Funktionen 18
2.3 Vorstellung der Programme 18
2.3.1 Spezielle Meta-Analysen-Software für DOS 18
2.3.2. Spezielle Meta-Analysen-Software für Windows 25
2.3.3 Allgemeine Statistikprogramme mit Meta-Analysen-Funktion 31
2.3.4 Programme zur Effektstärkenberechnung 36
2.4 Konklusio 38

II. USABILITY-STUDIE 39

1. USABILITY: THEORETISCHER HINTERGRUND 40

1.1 Definition 40
1.2 Entwicklung 41
1.3 Usability Engineering 41
1.4 DIN EN ISO 9241-110 Grundsätze der Dialoggestaltung 42
1.5 Usability-Evaluation 46
1.5.1 Ziele 46
1.5.2 Methoden 47
1.5.2.1 Analytische Evaluation 48
1.5.2.2 Empirische Evaluation 50
1.5.2.3 Formative und summative Evaluation 52

2. STUDIENDESIGN 53

2.1 Methode 54
2.1.1 Fragestellung 54
2.1.2 Verwendeter Fragebogen 55
2.1.3 Beschreibung der Stichprobe 56
2.1.4 Erhebungsmethode 58
2.2 Ergebnisse 59
2.2.1 Usability Assesment: Auswertung nach dem IsoMetrics-Manual 59
2.2.2 Weitere Berechnungen 63
2.3. Diskussion 66

2.3.1 Interpretation der Ergebnisse 66
2.3.2 Probleme früherer Arbeiten 68
2.3.3 Einschränkungen 69
2.3.4 Zusammenfassung und Ausblick 69
3. SCHLUSSWORT 71
4. ABBILDUNGSVERZEICHNIS 72
5. TABELLENVERZEICHNIS 72
6. ANHÄNGE 72
LITERATURVERZEICHNIS 82

Vorwort

Diese Arbeit soll einen umfassenden Überblick über die Geschichte der Software für Meta-Analysen geben. Meta-Analysen sind sowohl in den Sozialwissenschaften, als auch in der Medizin und Wirtschaft, mittlerweile ein unabdingbares Mittel zur Auswertung, Datenanalyse und grafischen Darstellung der Daten.

Im Weiteren wird das Konzept der Usability erläutert und sechs Programme, die häufig in wissenschaftlichen Arbeiten zum Einsatz kommen, werden hinsichtlich ihrer Usability evaluiert.

Anmerkung: In der vorliegenden Arbeit werden personenbezogene Formulierungen immer in ihrer maskulinen Form wiedergegeben, um das inhaltliche Verständnis zu erleichtern. Diese Formulierungen sind jedoch geschlechtsneutral zu verstehen.

I. Review

Dieses Kapitel soll einerseits einen allgemeinen Überblick zur Meta-Analyse und ihrer computergestützten Durchführung schaffen, andererseits eine umfassende Beschreibung der zur Verfügung stehenden Programme geben.

1. Meta-Analysen: Theoretischer Hintergrund

1.1 Geschichte

Eines der frühesten Beispiele quantitativer Synthese lieferte Pearson (1904), indem er die Ergebnisse von fünf Studien über die Wirksamkeit von Typhusimpfungen unter Berücksichtigung der Gewichtung der Studien zusammenführte.

Der Begriff „meta-analysis“ wurde erstmals Mitte der 1970er Jahre von Glass (1976) verwendet. Er betont die Wichtigkeit dieser Art von Datenanalyse, da die Menge an publizierter Forschung zu vielen Themen rapide zunehme und so ein Überblick über die Ergebnisse verschafft werden könne. Glass und Smith (1977) waren es auch, die die erste derart benannte Analyse zur Wirksamkeit von Psychotherapie durchführten.

Etwa zur gleichen Zeit wurden ähnliche Ansätze zur Zusammenführung einzelner Studien von Schmidt und Hunter (1977), Cooper und Rosenthal (1980) und Rosenthal und Rubin (1982; 1986) entwickelt und vorangetrieben. Der Begriff Meta-Analyse umfasst all diese Ansätze von den genannten und auch anderen Forschern.

Es gab jedoch auch kritische Sichtweisen in Bezug auf dieses neue Verfahren. Eysenck (1978) etwa bezeichnete Meta-Analysen als „an excercise in mega-silliness“. Trotzdem hat sich dieses Verfahren durchgesetzt und zählt heute zu den höchsten Standards in der Forschung.

1.2 Methoden

Meta-Analysen repräsentieren die Ergebnisse jeder Studie in Form von Effektstärken. In den letzten 35 Jahren wurden tausende Meta-Analysen durchgeführt und vielfältige Berechnungsverfahren für Meta-Analysen entwickelt. Der folgende Abschnitt wird wichtige etablierte Methoden kurz vorstellen.

Effektstärke

Als Effektstärke wird ein statistisches Maß bezeichnet, welches die Größe eines Effekts wiedergibt. Für dichotome Endpunkte sind das relative Risiko (RR) oder die Odds Ratio (OR) üblich, für kontinuierliche Endpunkte sind in Einzelstudien die standardisierte mittlere Differenz (SMD) und in Meta-Analysen die gewichtete mittlere Differenz (weighted mean difference, WMD) gebräuchliche Effektmaße (Deutsches Cochrane Zentrum, 2011).

Fixed-Effect-Modell

Das Fixed-Effect-Modell ist ein statistisches Modell zur Zusammenfassung von Ergebnissen einzelner Studien. Es beruht auf der Annahme, dass alle Studien den gleichen Effekt schätzen und Unterschiede nur durch zufällige Abweichung bzw. verschieden große Stichproben der Primärstudien bedingt sind. Die Ungenauigkeit des Gesamteffektes ist damit nur durch die Variation innerhalb der einzelnen Studien beeinflusst. Beispiele sind die Inverse-Varianz-Methode, das Peto-Modell und die Mantel-Haenszel Odds Ratio.

Random-Effects-Modell

Das Random-Effects-Modell ist ein statistisches Modell zur Berechnung zusammengefasster Ergebnisse; im Gegensatz zum Fixed-Effect-Modell werden hierbei die Effektunterschiede zwischen verschiedenen Primärstudien berücksichtigt. Da sowohl die Variation, die innerhalb der Studien beobachtet wird, als auch die Variation zwischen den Studien in die Genauigkeit der Schätzung des gemeinsamen Effekts eingehen, wird dieses Modell in der Praxis häufig angewandt, weil selten von einem identischen Effekt ausgegangen werden kann. Als Beispiel kann die Methode nach DerSimonian & Laird genannt werden.

Forest Plot

Der Forest Plot stellt die Einzelergebnisse aller in die Analyse eingeschlossenen Studien zusammen mit dem Ergebnis der kombinierten Analyse dar. Typischerweise wird die Punktschätzung der einzelnen Studien durch ein schwarzes Quadrat dargestellt, durch das eine horizontale Linie läuft, welche das Konfidenzintervall beschreibt. Der Schätzwert der gesamtem Meta-Analyse wird durch einen Diamanten ganz unten in der Grafik repräsentiert (Lewis & Clarke, 2001). Die Grafik hilft, sich in kurzer Zeit einen Überblick über gemeinsame Trends der Einzelstudien zu verschaffen.

Nach Lewis und Clarke (2001) wurde der Forest Plot spätestens ab den 1970er Jahren verwendet. So stellten Freiman, Chalmers, Smith und Kuebler (1978) die Ergebnisse einiger Studien in Form von horizontalen Linien für die Konfidenzintervalle mit einer Kennzeichnung der Punktschätzung dar. Dies war allerdings keine Meta-Analyse und die einzelnen Studienergebnisse wurden deshalb auch nicht zu einem insgesamten Ergebnis kombiniert.

Kumulative Meta-Analyse

Als kumulative Meta-Analyse bezeichnet man das wiederholte Durchführen einer Meta-Analyse mit inkrementeller Erhöhung der Anzahl der Studien. Die Studien können hierbei nach verschiedenen Kriterien sortiert in die Analyse eingehen, üblicherweise werden sie chronologisch nach dem Jahr ihrer Erscheinung kumuliert (Lau, Schmid & Chalmers, 1995). Die grafische Darstellung erfolgt als kumulativer Forest Plot, der ein schnelles Erkennen von Trends oder auch plötzlicher Sprünge in den Studienergebnissen ermöglicht.

Beurteilung des Publikationsbias

Nach Borenstein (2005) sollte eine Reihe von Analysen vollzogen werden, um das Vorhandensein eines Publikationsbias zu erkennen. Im Folgenden werde ich auf einige gängige Verfahren bzw. Darstellungsweisen kurz eingehen.

Funnel Plot

Der Funnel Plot ist ein Streudiagramm, im welchem die Effektgrößen verschiedener Studien gegen ein Maß der Präzision wie Studiengröße oder Kehrwert der Varianz, oder eine andere Studieninformation aufgetragen wird. Durch diese grafische Methode kann das Vorhandensein eines Publikationsbias geprüft werden (Deutsches Cochrane Zentrum, 2011).

Kleine Studien erscheinen weiter unten in der Grafik und streuen breiter um die Mittelachse, große Studien hingegen befinden sich weiter oben und näher an der Mittellinie. Die erwartete Verteilung ähnelt der Form eines (sich nach oben entleerenden) Trichters (Light & Pillemer, 1984). Sind die Werte symmetrisch verteilt, so spricht das gegen das Vorhandensein eines Publikationsbias.

Trim-and-Fill

Mit Hilfe der Trim-and-Fill-Methode (Duval & Tweedie, 2000) wird der asymmetrische Teil des Funnel Plots entfernt, nachdem geschätzt wurde, wie viele Studien sich darin befinden. Aufgrund der verbleibenden Studien wird ein neuer Effektschätzer festgelegt. Anschließend werden die gelöschten Studien und ihre zum neuen Schätzer symmetrischen Gegenstücke in den Plot eingefügt, um daraus einen korrigierten Effektschätzer zu berechnen. Die neu erzeugten Punkte, die die fehlenden Studien symbolisieren, werden optisch von den ursprünglich vorhandenen abgehoben

Begg und Mazumdars Rangkorrelationstest

Der Rangkorrelationstest nach Begg und Mazumdar (1994) berichtet die Rangkorrelation (Kendalls Tau) zwischen der standardisierten Effektstärke und den Varianzen (oder Standardfehlern) dieser Effekte. Tau ist ähnlich zu anderen Korrelationen zu interpretieren, also ein Wert von null zeigt an, dass kein Zusammenhang zwischen den Effektstärken und der Genauigkeit besteht, Abweichungen von Null hingegen sind ein Indikator für einen Zusammenhang (Begg & Berlin, 1988; Begg & Mazumdar, 1994).

Eggers Regression

Die Methode der linearen Regression nach Egger, Smith, Schneider und Minder (1997) dient ebenfalls der Beurteilung des Publikationsbias. Nach Borenstein (2005) unterscheidet sie sich von der vorigen Methode insofern, dass nicht die Ränge, sondern die tatsächlichen Effektstärken und ihre Genauigkeit zur Berechnung verwendet werden.

Fail-Safe-N

Rosenthal (1979) schlägt vor, die Anzahl der Studien mit Nulleffekt zu berechnen, die man zum tatsächlichen Effekt der Meta-Analyse addieren müsste, damit der kombinierte Effekt nicht länger statistisch signifikant wäre. Nach Borenstein (2005) wurde dieses Verfahren zuerst “file drawer”-Analyse und auf Anregung von Cooper (1979) “failsafe *N*” benannt.

1.3 Vorteile

Lipsey und Wilson (2009) sehen die Vorteile einer Meta-Analyse wie folgt:
Durch die Durchführung einer Meta-Analysen werden Forschungsergebnisse strukturiert zusammengefasst und dokumentiert und können so jederzeit nachvollzogen werden. Außerdem repräsentieren sie Studienergebnisse komplexer und differenzierter als die bis dahin üblichen Arten von Reviews. Vor allem die Verwendung von Effektstärken verhilft dazu, Ergebnisse unterschiedlicher Gewichtung aus verschiedenen Studien zu vergleichen. So können Effekte oder Zusammenhänge aufgedeckt werden, die bei Verwendung anderer Techniken eines Reviews, beispielsweise dem qualitativen bzw. narrativen Review, verborgen bleiben würden. Meta-Analysen liefern einen organisierten Weg, um mit Informationen aus einer großen Anzahl von Studien im Sinne eines Reviews umzugehen. So besteht die Möglichkeit, fast unzählige Studien und Details daraus festzuhalten

1.4 Probleme

Sharpe (1997) sieht drei grundsätzliche Gefahren der Validität von Meta-Analysen. Er betitelt diese als „apples and oranges“, „file drawer“ und „“garbage in, garbage out“.

„garbage in, garbage out“

Die methodische Qualität der Primärstudien unterliegt oft einer starken Variation. Das gefährdet wiederum die Validität der Meta-Analyse, wenn bei der Integration zwischen methodisch guten und schlechten Arbeiten nicht unterschieden wird.

„apples and oranges“

Bei der Literatursuche und -auswahl ist sehr wichtig, die Fragestellung zu spezifizieren sowie genaue Ein- und Ausschlusskriterien festzulegen, da einzelne Studien häufig zu heterogen sind, um hinsichtlich eines bestimmten Effekts zusammen analysiert zu werden. Dies ist auch als Uniformitätsproblem bekannt.

„file drawer“

Meist werden Studien, die einen signifikanten Effekt nachweisen, eher veröffentlicht als solche, denen dies nicht gelingt. Diese Verzerrung spiegelt sich folglich in der zu einem bestimmten Thema verfügbaren Literatur wider. So meint Sharpe (1997), dass eine Meta-Analyse, die nur auf Basis von veröffentlichter Literatur durchgeführt werde, die Gefahr der Überschätzung des Effekts mit sich bringe, oder sogar ein Effekt gefunden würde, der in Wirklichkeit nicht existiere.

Eine weitere Problematik stellen fehlende Studiendaten dar, die, sollte man ihrer nicht habhaft werden können, einen Ausschluss der Studie aus der Analyse bedingen. Entscheidet man sich dazu, die Studien trotzdem in die Berechnungen einfließen zu lassen, geht man die Gefahr einer Verzerrung des gesamten Ergebnisses der Analyse ein (Yuan & Little, 2009).

1.5 Software

Auch wenn eine Meta-Analyse mit sehr wenigen Studien manuell durchgeführt werden könnte, empfiehlt sich der Einsatz eines Computers. Zur Vorbereitung des Datensatzes können neben der eigentlichen Statistik-Software auch spezielle Programme verwendet werden, wie z.B. FileMaker Pro, Paradox, dBASE, FoxPro oder MS Access, die eine Verwaltung hierarchischer Daten ermöglichen. Einige statistische Softwarelösungen verfügen auch über ein eigenes Modul zur Dateneingabe. Dies hat den Vorteil, dass die Daten für die nachfolgende Berechnung nicht mehr konvertiert werden müssen (Lipsey & Wilson, 2009).

Die Meta-Analyse nutzt spezielle statistische Methoden und Techniken, welche nicht in allen Statistikprogrammen enthalten sind; diese spezifischen Anforderungen erfüllen eigens auf diesen Zweck spezialisierte Programme. Mittlerweile stehen zur Durchführung von Meta-Analysen einige Programme zur Auswahl; neben ziemlich kostspieligen gibt es auch kostenfrei verfügbare, die sich in ihrer Qualität nicht zwingend unterscheiden.

Grundsätzlich kann Software in zwei Gruppen eingeteilt werden. Zum einen sind das kommandogesteuerte Programme, das heisst, man gibt einen bestimmten Befehl zur Durchführung einer Aktion ein. Diese Befehle können entweder selbst geschrieben oder bereits vorgefertigt für bestimmte Zwecke aus Büchern oder von Internetseiten entnommen sein. Diese Art von Software erfordert ein gewisses Maß an Expertise und eine relativ lange Einarbeitungszeit, ist aber dafür meist flexibler in der Gestaltung und weist einen höheren Grad an Individualisierbarkeit auf. Andererseits gibt es menügesteuerte Programme, in denen Funktionen durch Klicken mit der Maus oder Auswahl von Funktionen in Menüleisten abgerufen werden können. Diese Programme können grundsätzlich schneller erlernt werden, da die Funktionen durch Ausprobieren erprobt werden können. Beide Programmarten sind auch im Bereich der Software für Meta-Analysen vertreten.

In der Vergangenheit gab es schon einige Reviews (Franke, 1992; Normand, 1995; Borenstein, 2005; Bax, Yu, Ikeda & Moons, 2007) über Software für Meta-Analysen, die sich jedoch zumeist nur einiger weniger Programme annahmen oder bestimmte Arten von Software, beispielsweise DOS Programme oder allgemeine

Statistikprogramme, außen vor ließen. Diese Arbeit soll einen Überblick über jegliche verfügbare Art von Software bieten, die den Nutzer erlaubt, meta-analytische Berechnungen durchzuführen. In den Review werden auch Programme mit einbezogen, die derzeit nicht gewartet werden, um die von den übrigen Arbeiten hinterlassenen Lücken zu schließen.

Diese Arbeit soll Wissenschaftlern und Forschern eine Orientierung bei der Fragestellung geben, welche Software ihrem Zweck gerecht wird. Die Wahl der Software wird sich zum Teil nach der Expertise, der Verfügbarkeit und den Anforderungen richten.

2. Programmübersicht

2.1 Softwaresuche

Gestartet habe ich meine Suche mit den im Artikel von Bax et al. (2007) erwähnten Programmen. Aus dem Literaturverzeichnis kann man weitere Reviews (Franke, 1992; Normand, 1995) entnehmen, die sich mit Software für Meta-Analysen beschäftigen. Dann habe ich eine Google-Web-Recherche mit den Suchbegriffen Meta-Analysen Software, meta analysis software, meta analysis, meta analyses, program, macro und software, auch verschieden kombiniert, durchgeführt. Die so gefundenen Websites habe ich auf Relevanz geprüft und Informationen über Programme extrahiert. Da sich laufend Änderungen ergeben, sind in Anhang 1 alle Websites aufgelistet, die Links zu Programmen darstellen oder sonstige relevante Information bieten. So kann man sich über eventuelle Updates selbst informieren. Auch in folgenden Datenbanken habe ich eine Stichwortsuche mit den selben Begriffen durchgeführt: ISI Web of Knowledge, PubMed, Scopus, PsychInfo und PsychArticles. Der Review erhebt den Anspruch, die Programmpalette relativ erschöpfend anzugeben und es sind sowohl Windows- als auch DOS-basierte Programme darin enthalten. Falls Programme nicht gefunden bzw. vergessen wurden, bitte gerne Info an den Autor.

2.2 Beurteilung der numerischen und grafischen Funktionen

Alle verfügbaren Programme wurden hinsichtlich ihrer grafischen und numerischen Funktionen untersucht. Es wurden nicht alle Programme erschöpfend getestet, so kann es auch passieren, dass bestimmte Funktionen trotz intensiver Beschäftigung und Suche nicht genannt werden.

2.3 Vorstellung der Programme

Nachfolgend werden alle gefundenen Programme, sortiert nach Software, die speziell für meta-analytische Berechnungen entwickelt wurde und allgemeinen Statistikprogrammen, die entweder Funktionen für Meta-Analysen beinhalten oder für welche Macros vorhanden sind, aufgelistet. Außerdem wird zwischen Programmen für DOS und Windows unterschieden.

2.3.1 Spezielle Meta-Analysen-Software für DOS

Advanced BASIC Meta-Analysis (ABMA) 1.10

Das DOS-Programm wurde ab 1989 zusammen mit einem Buch auf Diskette vertrieben und ist der Nachfolger von Mullen und Rosenthal´s BASIC Meta-Analysis: Procedures and Programs (1985). Text und Software beziehen sich größtenteils auf Techniken, die von Rosenthal und Rubin (1982; 1986) entwickelt und durch zahlreiche Bücher und Artikel verbreitet wurden. Das Buch kann nach Franke (1992) als eine Einführung in die Thematik der Meta-Analyse betrachtet werden. Vom Wert einer Meta-Analyse über Literatursuche und Kodierung der Literatur bis hin zum „file drawer problem“ und der Interpretation der Ergebnisse wird darin das gesamte Vorgehen beschrieben.

Die Daten werden direkt in das Programm eingegeben und können nicht aus anderen Quellen importiert werden. Die Datensätze können gespeichert und später nachbearbeitet werden. Dem Nutzer stehen eine Reihe von Heterogenitäts- und Homogenitätstests zur Wahl, auch eine limitierte Anzahl von Grafiken kann erstellt werden. Das Programm sieht jedoch keine Ausgabe an einen Drucker vor.

BASIC Meta-Analysis

BASIC Meta-analysis wurde auf einer Diskette zusammen mit einem Buch (Mullen & Rosenthal, 1985) vertrieben und besteht aus insgesamt 16 einzelnen Programmpaketen. Die Daten müssen direkt im Programm in Form von Effektstärken, der Richtung des Effekts und einer Gewichtung für jede Studie eingegeben werden. Es sind nur sehr grundlegende meta-analytische Berechnungen vorgesehen und Grafiken lassen sich nicht erzeugen.

DSTAT

DSTAT 1 konnte, als es noch gewartet wurde, zusammen mit einem Buch erworben werden, und ist mittlerweile per Email an den Autor Blair T. Johnson erhältlich. Das Buch ist eine Art User Guide und orientiert sich an Methoden von Hedges und Olkin (1985), enthält aber auch eine kurze Einführung zum Thema Meta-Analysen.

Das Programm wurde ursprünglich für den Einsatz im psychologischen Kontext entwickelt. Die Daten werden in Form von Vierfeldertafeln, Korrelationskoeffizienten, Teststatistiken, *p*-Werten oder gemischt eingegeben. Diese Werte werden anschließend in standardisierte Effektmessungen umgewandelt. Klinisch relevante Maßzahlen wie Risk Difference, relatives Risiko oder Odds Ratio können mit DSTAT nicht berechnet werden. Außerdem gibt es keine Möglichkeit zur grafischen Ergebnisdarstellung, weshalb der Nutzen für die Verwendung im medizinischen Bereich relativ eingeschränkt ist (Sterne, Egger & Sutton, 2001).

Alle Berechnungsmöglichkeiten unter Version 1.x arbeiten unter der Annahme des Fixed-Effect-Modells. Der Autor selbst empfiehlt, einige Analysemethoden dieser Version nicht mehr zu verwenden.

DSTAT 2.x war zum Zeitpunkt der Softwaresuche noch in einem sehr frühen Entwicklungsstadium und wurde vom Autor als Testversion zur Verfügung gestellt. Das Programm benötigt eine Java-Laufzeit-Umgebung.

Diese Version bietet die Durchführung von Berechnungen unter der Annahme des Fixed- und Random-Effects-Modells, die Erstellung von Grafiken wie Funnel und

Forest Plot und verschiedene Effektstärkenberechnungen, die einen höheren Grad an Genauigkeit aufweisen als jene in Version 1.x (Johnson, 2010).

EasyMA

EasyMA ist eine DOS-Software für Meta-Analysen von klinischen Studien mit dichotomen Outcomes und kann online kostenfrei unter http://www.spc.univ-lyon1.fr/easyma.dos/ heruntergeladen werden. Das Programm wird allerdings nicht mehr gewartet und ist der Vorgänger der späteren Windows-Version WEasyMA. Eine kurze Beschreibung der Funktionen findet sich in einem Artikel von Cucherat, Boissel, Leizorovicz und Haugh (1997), und auch ein Manual zum Programm ist erhältlich.

Die Dateneingabe erfolgt direkt im Programm in einer Tabelle. Es besteht die Möglichkeit, den Publikationsbias zu bestimmen, indem die Anzahl der Studien ohne oder mit negativem Effekt berechnet wird, die zur Analyse hinzugefügt werden müssten, damit der insgesamt errechnete Effekt nicht mehr signifikant ist. Sensitivitätsanalysen sind in Form von Subgruppenanalysen und kumulativer Meta-Analyse möglich (Cucherat et al., 1997).

Die Durchführung einer Meta-Analyse ist sowohl unter Annahme des Fixed-, als auch Random-Effects-Modells möglich. Die Ergebnisse können grafisch in Form eines Forest Plots dargestellt werden.

EpiMeta

Das Programm ist kostenfrei im Internet unter http://ftp.cdc.gov/pub/Software/epimeta/ als Download verfügbar und wurde für Meta-Analysen von epidemiologischen Daten entwickelt. Als Hilfestellung steht ein Manual zur Verfügung. Trotz DOS-Interface ist die Benutzung einer Maus möglich.

Die Studiendaten müssen entweder in Form von dichotomen Daten oder Effektstärken manuell im Programm eingegeben werden. Analysen sind sowohl unter Annahme des Fixed-, als auch Random-Effects-Modells möglich, auch Subgruppenanalysen können durchgeführt werden. Die Ergebnisse können grafisch in Form eines Forest Plots dargestellt und als Report gedruckt werden.

*Fast*Pro 1.7*

Das Programm wurde auf Diskette zusammen mit einem Manual vertrieben, welches sowohl die Installation als auch Berechnungsmöglichkeiten erklärt. Die Methoden werden von Eddy, Hasselblad, & Shachter (1992) beschrieben. Als Format für die Dateneingabe in das Programm sind beispielsweise Mittelwerte mit Standardabweichungen, Odds Ratios oder Regressionskoeffizienten möglich. Die Berechnungen erfolgen unter Annahme des Fixed- oder Random-Effects Modells. Grafische Darstellungen sind nicht möglich.

Im Gegensatz zu einigen anderen Programmen, die zu dieser Zeit entwickelt wurden, wie DSTAT oder True Epistat, werden mit Hilfe dieses Programmes keine Teststatistiken, sondern Parameterschätzungen verglichen. Es können nicht nur Meta-Analysen aufgrund von Daten aus vergleichenden, sondern auch aus Studien mit nur einer Versuchsgruppe berechnet werden (Normand, 1995).

HepiMA

HepiMA ist eine Software vorwiegend für epidemiologische Meta-Analysen, also dichtome Daten. Das Programm läuft ursprünglich unter DOS, ist aber auch in einer Windows-Version verfügbar und kann auf Anfrage beim Autor Bahi Takkouche bezogen werden.

Das Programm ist für Epidemiologen und andere Forscher gedacht, die eine Meta-Analyse mit Beobachtungsstudien als Grundlage durchführen wollen und bei denen das relative Risiko den Interessensfokus bildet. Ein wesentlicher Vorteil des Programmes ist, dass es eine Option für die Eingabe von angepassten Schätzungen des relativen Risikos inklusive Konfidenzintervalle der Individualstudien gibt, was die gängigste Art und Weise in der epidemiologischen Literatur ist, Ergebnisse zu präsentieren.

HepiMA wendet zusätzlich zu asymptotischen Methoden auch Bootstrapping-Methodologien an. Um asymptotische Berechnungen durchzuführen, ist eine große Anzahlt an Primärstudien nötig, um die Validität zu sichern. Dieser Grundsatz findet allerdings in der Realität oft keine Beachtung. Deswegen berechnet HepiMA für jeden

der asymptotischen Heterogenitätstest eine parametrische Bootstrap-Version mit 1000 Replikationen. Bootstrapping ist ein Verfahren, bei dem aus den vorhandenen Datensätzen neue Datensets zufällig gezogen werden und die empirische Verteilung des gewünschten Verfahrens aus einer großen Anzahl dieser Datensets berechnet wird (Costa-Bouzas, Takkouche, Cadarso-Suarez & Spiegelman, 2001). Außerdem kommen neue Methoden zur Schätzung der Größe der Heterogenität, die von den Autoren selbst entwickelt wurden, zum Einsatz.

Alle Berechnungen können unter Annahme des Fixed- und Random-Effects-Modells durchgeführt werden. Grafische Darstellungen in Form von Forest und Funnel Plot sind nur mit der Windows-Version der Software möglich.

Hunter Schmidt MA Program Package

Das Programmpaket wurde von Frank Schmidt und Huy Le entwickelt und kann auf Anfrage bezogen werden. Daten aus Primärstudien müssen in Form von Korrelationen (*r*) oder Effektstärken (*d*) manuell eingegeben werden. Obwohl das Programm relativ neu ist, lässt es sich nicht mit einer Maus bedienen.

Eine Besonderheit des Programms ist, dass es in der Voreinstellung 80-prozentige „Credibility-Intervalle“ anstelle von Konfidenzintervallen verwendet. Zum momentanen Zeitpunkt sind nur Berechnungen unter der Annahme des Random-Effects-Modells möglich. Studien können, beispielsweise zum Zweck von Sensitivitätsanalysen, einfach entfernt werden (Roth, 2008).

Meta-analysis 5.3 Statistics Software for Meta-Analysis

Das kostenlose DOS-Programm wurde von Ralf Schwarzer an der FU Berlin speziell für Meta-Analysen von kontinuierlichen Daten entwickelt und kann zusammen mit einem Manual von der Homepage des Autors unter http://userpage.fu-berlin.de/~health/meta_e.htm heruntergeladen werden.

Die Software kann sowohl Analysen aufgrund von *p*-Werten als auch von Effektstärken (*d* oder *r*) rechnen und die Dateneingabe erfolgt im integrierten Texteditor. Die Visualisierungsmöglichkeiten des Programms beschränken sich auf ein Stamm-

Blatt-Diagramm von Korrelationskoeffizienten zur Darstellung der Häufigkeitsverteilung. Des weiteren stehen zur Vorbereitung und Transformation der Rohdaten einige Routinen bereit (Schwarzer, 1997).

MetaDos

MetaDos ist ein menügesteuertes DOS-Programm, das sich mit einer Maus bedienen lässt. Es kann auf Anfrage kostenlos beim Autor bezogen werden. Die verwendeten statistischen Verfahren orientieren sich an den von Hunter und Schmidt (1990) entwickelten Methoden. Daten aus Primärstudien müssen in Form von Effektstärken (*d)* und Korrelationen (*r*) manuell in den Editor eingegeben werden. Dieser erlaubt die gleichzeitige Bearbeitung mehrerer Dateien, wie auch den Austausch ihrer Inhalte durch kopieren und einfügen. Eine Gruppierung für Sensitivitätsanalysen wird mittels Etikettierung ermöglicht (Stauffer, 1996).

Der Output öffnet sich in einem separaten Fenster und kann gedruckt werden. Eine grafische Darstellung der Ergebnisse ist in diesem Programm nicht vorgesehen.

Meta-Stat 1.5

Meta-Stat ist ein kostenfreies Programm speziell für die Berechnung von Meta-Analysen und einer der Autoren des Programms, Gene Glass, war es, der 1976 den Begriff „Meta-Analyse“ erstmals verwendete. Unter http://ericae.net/meta/metastat.htm ist es zum freien Download verfügbar.

Analysen in der Tradition nach Hedges und Olkin (1985), Schmidt und Hunter (1977) und Hunter und Schmidt (1990) werden unterstützt. Es beinhaltet auch ein „plain statistics“-Tool für allgemeine statistische Berechnungen wie etwa Effektstärken. Außerdem sind unter anderem Heterogenitätsstatistiken und eine Fail-Safe-*N* Berechnung möglich (Rudner, 2003*)*.

Meta-Test

Meta-Test wurde ebenfalls von Joseph Lau entwickelt und ist kostenfrei auf Anfrage beim Autor verfügbar. Das Programm wurde speziell für Meta-Analysen von diagnostischen Testdaten entwickelt und Daten werden nur in Form von Anzahl der Patienten und Anzahl von Events in jeder Gruppe bzw. true +, false -, false + und true - inklusive der Kovariaten akzeptiert. Diese müssen im Programm selbst eingegeben werden (Sterne et al., 2001).

Sensitivitäts- und Spezifitätsanalysen und auch ROC-Kurven-Analysen können durchgeführt und grafisch dargestellt werden, die Plots können per Druckfunktion exportiert werden. Die Analysen können sowohl unter der Annahme des Fixed- als auch Random-Effects-Modells durchgeführt werden.

TRUE EPISTAT 5.3

TRUE EPISTAT wurde als statistisches und grafisches Softwarepaket für epidemiologische und klinische Studien auf Disketten vertrieben. Das Programm beinhaltet ein Manual, in welchem beschrieben ist, welche Formeln für die jeweiligen Rechenoperationen verwendet werden.

Neben meta-analytischen Berechnungen können mit dem Programm beispielsweise noch parametrische und parameterfreie Tests, ANOVA und ANCOVA, Analysen über Messwiederholungen, verschiedene Regressionsanalysen und ROC-Analysen durchgeführt werden. Alle Ergebnisse können auch grafisch dargestellt werden (Oster, 1998).

Die Daten müssen direkt in das Programm eingegeben werden. Hierzu werden Studienname, Stichprobengröße, die Richtung des Effekts pro Studie, die Art der Teststatistik (*d, t* oder *z*) zusammen mit dem Wert für jede Studie.

Das Meta-Analysen-Tool arbeitet unter der Voraussetzung, dass in jeder Studie zwei Gruppen verglichen werden und die Anzahl von Einheiten innerhalb jeder Gruppe ausgeglichen ist. Außerdem muss jede Studie mindestens acht Beobachtungen aufweisen, um in die Meta-Analyse eingeschlossen werden zu können (Normand, 1995).

2.3.2. Spezielle Meta-Analysen-Software für Windows

Comprehensive Meta-Analysis (CMA)

CMA wird von Biostat durch ein Team von Experten – u.a. Michael Bornstein und Hanna Rothstein – entwickelt und kostenpflichtig unter www.meta-analysis.com vertrieben, die aktuelle Version ist 2.2.057. Beim Start des Programmes steht dem Nutzer ein Tutorial bei den ersten Schritten zur Seite, für weitergehende Informationen kann auf das umfassende Handbuch zurückgegriffen werden. Zur Vertiefung werden von den Entwicklern regelmäßig Kurse angeboten.

Die Dateneingabe erfolgt entweder direkt im Programm oder per Import einer Tabellen-Datei. Das Alleinstellungsmerkmal von CMA ist, dass es Studiendaten in nahezu allen Formaten, auch verschiedene innerhalb einer Meta-Analyse akzeptiert. Es bietet die Möglichkeit, die Studien zu gruppieren, z.B. nach Art der Behandlung oder stichprobenspezifischen Faktoren, sowie eine Datenbank aller Studien inklusive Abstracts und Literaturverzeichnis zu erstellen.

Die Software beherrscht sämtliche gängigen Berechnungsmethoden wie Fixed-Effect-Modell, Random-Effects-Modell, Sensitivitäts- und Subgruppenanalysen, Berechnungen zur Beurteilung des Publikationsbias inklusive einer Trim-und-Fill-Darstellung nach Duval und Tweedie (2000), Heterogenitäts- und Homogenitätsanalysen und hochauflösende Plots; zur Beurteilung des Einflusses von Moderatorvariablen stehen Optionen wie Mixed-Effects-ANOVA und Metaregression bereit (Sterne et al., 2001).

Die Grafiken können als Datei oder direkt nach PowerPoint oder Word exportiert werden. Ebenso lässt sich auswählen, die Berechnungen mit den von Revman benutzten Optionen durchzuführen. Das Look and Feel orientiert sich an Windows und mag zur wahrgenommen Benutzerfreundlichkeit beitragen.

MetaAnalyst

MetaAnalyst wurde als DOS-Programm von Joseph Lau entwickelt, auf Basis dessen später eine Windows-Version entstand; diese kann über http://tuftscaes.org/meta_analyst/ kostenfrei heruntergeladen werden.

Die DOS-Version ist nur für einfache und kumulative Meta-Analysen für klinische Stichproben mit binäre Daten verwendbar, während die Windows-Version auch den Umgang mit kontinuierlichen und diagnostischen Daten beherrscht.

Beim Öffnen des Windows-Programms erscheint im unteren Bereich des Bildschirms eine kontextbezogene Hilfefunktion, die sich auch ausblenden lässt. Daten können direkt im Programm eingegeben oder aus einer Excel-Datei importiert werden, die Maske dazu wird mit einer Drag-and-Drop-Funktion angepasst.

Alle Berechnungen sind sowohl unter Annahme von Fixed- als auch Random-Effects Modell möglich. Zusätzlich zur normalen Meta-Analyse können kumulative Meta-Analysen, Sensitivitätsanalysen, Metaregression und Subgruppenanalysen durchgeführt und übliche Plots dargestellt werden. Nach Auswahl der Parameter werden alle Ergebnisse und Plots der Analyse sofort angezeigt.

MetaDiSc

MetaDiSc ist eine Software speziell für Meta-Analysen von diagnostischen Daten. Die aktuelle Version 1.4 ist kostenfrei für den akademischen Gebrauch und kann zusammen mit einem Manual, in welchem die statistischen Methoden erklärt werden, von ftp://ftp.hrc.es/pub/programas/metadisc/Metadisc_update.htm heruntergeladen werden.

Die Dateneingabe ist entweder direkt im Programm möglich, genauso können die Daten aber auch aus anderen Spreadsheets, z.B. Excel, kopiert, oder aus einem Textfile importiert werden. Die Ergebnisse der Primärstudien müssen in Form einer Vier-Felder-Tafel vorliegen (positive und negative Likelihood Ratios oder Spezifität und Sensitivität).

Man kann verschiedene grafische Ergebnisdarstellungen erzeugen, die auch exportiert werden können. Zur Beurteilung der Heterogenität wird der Grad der

Variabilität zwischen den Studienergebnissen evaluiert, indem Sensitivität und Spezifität in einem Forest Plot abgebildet werden. Die Metaregression verwendet die Moses-Shapiro-Littenberg-Methode.

Statistisches Pooling für Sensitivität und Spezifität, Likelihood Ratios und diagnostische Odds Ratios, allgemein und in Subgruppen, sind unter Annahme des Fixed- und Random-Effects-Modells möglich (Zamora; Abraira, Muriel, Khan & Coomarasamy, 2006).

MetAnalysis

Diese Software war nur in Kombination mit einem Buch (Leandro & Gallus, 2005) zu erwerben und wurde vorwiegend zur Analyse von Daten aus klinischen Studien entwickelt. Mittlerweile wird es allerdings nicht mehr vertrieben und gewartet.

Das Buch beschäftigt sich zum Großteil mit der Planung und Durchführung von Meta-Analysen, verschiedenen Modellen und Möglichkeiten zur Berechnung. Das letzte Kapitel kann als Softwaremanual inklusive Hilfestellungen zur Installation, Dateneingabe und Berechnung verstanden werden. Im Programm selbst gibt es keine Hilfeoptionen, dafür kann man im Buch aber mehr Informationen finden. Die binären Daten müssen direkt im Programm eingegeben werden und können nicht aus anderen Quellen importiert werden.

Man hat die Wahl zwischen einer Intention-to-Treat oder Per-Protocol-Analyse, zu diesem Zweck können auch Drop-Outs eingegeben werden. Die Ergebnisse werden jeweils für Fixed- und Random-Effects-Modell angezeigt und es besteht die Möglichkeit einer Subgruppenanalyse. Auch alle gängigen Plots können dargestellt werden (Leandro & Gallus, 2005)

MetaWin

MetaWin ist ein kostenpflichtiges Programm speziell für die Durchführung von Meta-Analysen und das letzte Versionsupdate 2.1.5 ist Anfang 2007 erschienen. Die Software kann über die Internetseite http://www.metawinsoft.com/ bezogen werden.

Die Daten können in drei verschiedenen Formaten (Vierfeldertafeln, Mittelwerte mit Standardabweichungen und Korrelationskoeffizienten) entweder direkt im Programm eingegeben oder aus einer Tabelle importiert werden. Vor der Durchführung der eigentlichen Analyse, die sowohl unter der Fixed- als auch Random-Effects-Modellannahme möglich ist, müssen zuerst die Effektstärken berechnet werden, welche für jede weitere Berechnung oder Darstellung eines Plots angewählt werden müssen.

Neben der Darstellung verschiedener Plots wie Histogramme, Forest, Funnel, Radial und Quantilplots können außerdem Fail-Safe-*N*-Berechnungen, nonparametrische Resampling-Verfahren, kumulative Meta-Analysen, Sensitivitäts- und Heterogenitätsanalysen durchgeführt werden. Gemäß Bax et al. (2007) basieren alle Berechnungen in MetaWin auf *t*-Verteilungen. Bei Bedarf können Hilfeoptionen im Programm abgerufen oder das ausführliche Manual konsultiert werden.

Metaxis

Gemäß Sterne et al. (2001) war Metaxis als kostenpflichtiges Programm bei Update Software Ltd in Entwicklung und dafür designed, alle Aspekte eines Systematischen Reviews durchzuführen, inklusive einer Meta-Analyse. Die Durchführung sollte die Definition der Fragestellung des Reviews, Auswahlkriterien für die Studien, Datenextraktion und -analyse sowie die Verwaltung der Quellen beinhalten.

Dieses Programm konnte bei der Suche für diese Arbeit allerdings nicht gefunden werden. Es ist anzunehmen, dass es entweder nie oder nur für kurze Zeit vertrieben wurde.

MIX (Meta-analysis with Interactive eXplanations)

MIX ist ein statistisches Add-In für Excel und sowohl als kostenfreie Lite Version, in der allerdings keine neuen Datensätze erstellt werden können, als auch als kostenpflichtige Professional Version verfügbar. Die aktuelle Version ist 2.0.1.4 und benötigt Excel 2007 oder 2010. Die ältere Versionsreihe 1.x für Excel 2000 wird nicht mehr vom Hersteller gewartet, kann aber noch von http://www.meta-analysis-made-easy.com/ heruntergeladen werden.

Um die Software zu starten, muss Microsoft Excel mit Solver-Plug-In installiert sein und das Programm öffnet sich als weiteres Ribbon zusätzlich zu den Excel-Funktionen. Die Daten können sowohl aus Tabellen kopiert und importiert als auch direkt im Programm eingegeben werden. Es werden relativ viele Datenformate akzeptiert, innerhalb einer Meta-Analyse müssen die eingegebene Daten allerdings das gleiche Format haben.

Das Programm kann alle grundlegenden und fortgeschrittenen Berechnungen durchführen und sehr viele Grafiken hochauflösend generieren. Auch zur Beurteilung des Publikationsbias gibt es mehrere Tests und grafische Darstellungen sowie eine Trim-und-Fill-Option.

Durch das Windows-Look-and-Feel mag es gerade ungeübten Benutzern angenehm erscheinen. Außerdem positiv zu erwähnen ist, dass das Programm von der Autorengruppe validiert wurde, was in einem eigenen Paper (Bax et al., 2006) nachgelesen werden kann.

ReviewManager

RevMan ist das offizielle Tool der Cochrane Collaboration zur Durchführung systematischer Reviews und die aktuelle Version 5 kann von der Internetseite http://ims.cochrane.org/revman heruntergeladen und für den akademischen Gebrauch bzw. die Erstellung eines Cochrane Reviews kostenfrei verwendet werden. Das Programm bringt ein Tutorial sowie einen User Guide mit. Seit der Version 5 können zusätzlich zu Interventions-Reviews auch diagnostische und methodologische Reviews durchgeführt werden (The Cochrane Collaboration, 2011).

Das Programm dient dazu, komplette Reviews mit strukturiertem Text und Tabellen mit in-und exkludierten Studien durchzuführen. Das Meta-Analysenmodul von RevMan erlaubt die Berechnung von Fixed-Effect-, Random-Effects-, Subgruppen- und Sensitivitätsanalysen, indem manuell bestimmte Studien ausgeschlossen werden; auch verschiedene Plots können generiert werden. Die Berechnung einer Metaregression ist bislang nicht möglich. Die Notwendigkeit zur Eingabe diverser Studiendaten wirkt sich nachteilig auf die Verwendung des Programms zu rein meta-analytischen Zwecken aus.

Synthesizer 1.0

Der Synthesizer 1.0 ist eine von Zlatan Krizan erstellte Excel-Datei mit hinterlegten Formeln, die meta-analytische Berechnungen ermöglicht, und steht zum kostenlosen Download unter www.psychology.iastate.edu/~zkrizan/Synthesizer.htm bereit. Die Dateneingabe und die vorgesehenen Berechnungen werden im dazugehörigen Manual beschrieben.

Es können Meta-Analysen über Korrelationen und standardisierte Mittelwertsunterschiede gerechnet und Kontraste von Korrelationen oder Reliabilitäten beziehungsweise standardisierten Mittelwertsunterschiede erzeugt werden.

WeasyMA

WeasyMA wurde aufbauend auf EasyMA entwickelt und kostenpflichtig vertrieben. Allerdings wird es mittlerweile nicht mehr gewartet und kann auch nicht mehr erworben werden.

Das Programm dient der Berechnung von Meta-Analysen über klinische Studien, diese können sowohl unter der Annahme des Fixed- als auch Random-Effects-Modells durchgeführt werden. Auch kumulative Meta-Analysen sowie Subgruppenanalysen können berechnet werden. Alle erforderlichen Schritte werden ausführlich in dem dazugehörigen Manual beschrieben.

Die Dateneingabe muss direkt im Programm erfolgen, Einfügen aus der Zwischenablage ist nicht möglich. Ein großer Vorteil des Programms ist, dass die Ergebnisse innerhalb kürzester Zeit zur Verfügung stehen. Wie schon in einigen anderen Programmen können auch hier nur Daten im Format von Vierfelder-Tafeln eingegeben werden. An grafischen Darstellungen bietet es einen Radial, L'Abbé, Forest und Funnel Plot (Bax et al., 2007) Es gibt die Möglichkeit, druckbare HTML-Reports der Ergebnisse zu erstellen.

2.3.3 Allgemeine Statistikprogramme mit Meta-Analysen-Funktion

Die sogenannten „General Purpose Statistical Packages“, also allgemeinen Statistikprogramme, enthalten keine bereits vorhandenen Optionen zur Berechnung von Meta-Analysen, es gibt aber von Nutzern geschriebene frei verfügbare Programme bzw. Macros, die meist im Internet zum Download zur Verfügung stehen.

NCSS 07.1.21

In NCSS können die Daten in den Formaten Hazard Ratio, korrelierte und unkorrelierte Proportionen oder Mittelwerte mit Standardabweichungen entweder aus einer Tabelle importiert oder direkt im Programm eingegeben werden.

Nach der Dateneingabe können alle Parameter im Menü festgelegt werden, wobei eine Hilfe zu jedem Eingabefeld angezeigt wird, sobald man den Cursor dahin bewegt. Der gesamte Output aller gewählten Optionen wird in einem eigenen Fenster angezeigt. Die Analysen können sowohl unter Annahme des Fixed- als auch des Random-Effects-Modells durchgeführt werden.

Es besteht die Möglichkeit zur Darstellung von Forest, Radial und L’Abbé Plot, ein Funnel Plot kann allerdings nicht generiert werden. Alle grafischen Darstellungen können individuell angepasst werden. Optionen zur Beurteilung des Publication Bias fehlen in diesem Programm gänzlich.

R 2.13.0

Dieses Programm basiert auf der Programmiersprache S und man kann die Befehle der Software S-Plus teilweise mit nur kleinen Modifikationen auch in dieser Software nutzen. Aber auch speziell für R sind eigene Meta-Analysen-Befehle unter http://cran.r-project.org/web/packages/ verfügbar. Die grafischen Outputs sind von sehr guter Qualität (Sterne et al., 2001). Metafor ist das umfassendste dieser Pakete und wird deshalb im folgenden näher beschrieben. Weitere Pakete sind MAc, MAd, MADAM, meta, metacor, mvmeta, rmeta und diverse auf spezielle Einsatzzwecke für Meta-Analysen zugeschnittene Pakete. Eine kurze Beschreibung der bereitgestellten

Funktionen dieser Pakete und Downloadlinks befinden sich auf oben genannter Internetseite.

Daneben existiert noch das metaSEM-Paket (http://metasem.r-forge.r-project.org/). Es benutzt die Methode der Strukturgleichungsmodelle zur Berechnung uni- und multivariater Meta-Analysen.

R - Metafor Package 1.6 Das Metafor Package ist ein frei verfügbares und Open-Source Add-On zur Durchführung von Meta-Analysen mit der Statistiksoftware R. Es besteht aus einer Reihe von Funktionen, die dem Nutzer ermöglichen, verschiedene meta-analytische Berechnungen, unter anderem Metaregressionen oder kumulative Meta-Analysen, durchzuführen. Eine Vielzahl von Graphen können mit dem Paket generiert werden, sowie beispielsweise Forest-, Funnel- und Radialplots. Außerdem ist es möglich, den Einfluss des Publikationsbias mit Hilfe des Rangkorrelationstests und Eggers Regressionstests und der Trim-und-Fill Methode zu bestimmen. Daten aus Primärstudien werden in mehreren Formaten akzeptiert.

Auf der Website findet man Neuigkeiten und Updates, eine detaillierte Beschreibung der Funktionen, Anleitung zu Download und Installation des Paketes sowie einige Publikationen von Meta-Analysen, die Metafor oder den Vorgänger Mima für ihre Berechnungen verwendet haben. Das Metafor Package wird laufend weiterentwickelt und mit neuen Funktionen und Optionen ausgestattet (Viechtbauer, 2010).

Statistical Analysis System (SAS)

SAS 9.3 enthält keine Optionen zur Berechnung von Meta-Analysen; es gibt allerdings ein Buch von Wang und Bushman (1999), welches sich mit der Durchführung von Meta-Analysen in SAS beschäftigt, und alle darin genannten Routinen stehen sind im Internet unter http://ftp.sas.com/samples/A55810 zu finden.

Berechnungen nach dem Fixed- und Random-Effects-Modell können durchgeführt werden, allerdings ist die Qualität der grafischen Darstellungen nicht vergleichbar mit anderen kostenpflichtig vertriebenen Programmen (Sterne et al., 2001).

Andere Meta-Analysen-Macros für SAS stellt David Wilson als „Meta-analysis macros for SAS, SPSS, and Stata“ im Internet unter

http://mason.gmu.edu/~dwilsonb/ma.html zur Verfügung. Bei diesen Macros werden die Effektstärken nicht automatisch berechnet, sondern müssen anderweitig vorbereitet werden.

Zusätzlich gibt es die SAS Macro Suite von Kuss and Koch (1996) zur Durchführung von Meta-Analysen mit binären Daten als Grundlage. Diese Macros bieten auch die Darstellung gängiger Diagramme wie Funnel, Forest oder Galbraith Plot.

S-Plus

Dieses Programm basiert auf der statistischen Programmiersprache S, aus der R entwickelt wurde, und ist kostenpflichtig. Es enthält keine bereits vorhandenen Optionen zur Berechnung von Meta-Analysen, es gibt aber von Nutzern geschriebene frei verfügbare Programme bzw. Macros. Die grafischen Ausgaben sind von sehr guter Qualität (Sterne et al., 2001).

Prinzipiell lassen sich alle meta-analytischen Berechnungen in S-Plus programmieren. Manche R-Macros sind kompatibel, es existieren aber auch eigens für S-Plus vorgefertigte Befehle.

IBM SPSS Statistics

SPSS 20.0 ist ein allgemeines Statistikprogramm und enthält keine bereits vorhandenen Optionen zur Berechnung von Meta-Analysen. Auch für dieses Programm gibt es von Nutzern geschriebene frei verfügbare Programme bzw. Macros, die im Internet zum Download zur Verfügung stehen.

Einige davon sind von David Wilson, der unter http://mason.gmu.edu/~dwilsonb/ma.html auch Macros für Stata und SAS zur Verfügung stellt. Diese entstammen dem Buch „Practical meta-analysis“ (Lipsey & Wilson, 2001) und werden unter http://mason.gmu.edu/~dwilsonb/ma.html zum freien Download angeboten. Die Effektstärken für die weitere Analyse müssen in einem anderen Programm, z.B. SPSS, vorbereitet werden.

Auch Rustenbach (2003) liefert SPSS-Routinen zusammen mit seinem Buch „Meta-Analyse: Eine anwenderorientierte Einführung“. Das Buch kann sowohl als eine Art Anleitung zur Nutzung der Macros als auch als eine allgemeine Information zu Meta-Analysen gesehen werden.

Stata 12

Stata ist ein Programm, das für den ungeübten Nutzer eine relativ lange Einarbeitung erfordert, da es durch Befehlseingabe gesteuert wird. Die Befehle können auch mit Hilfe einer Dialogbox erstellt werden, was die Eingabe etwas erleichtert. Für geübte Nutzer von Stata gibt es allerdings eine große Auswahl von Macros für die Berechnung von Meta-Analysen.

Die Daten können direkt im Programm in eine Tabelle eingegeben oder aus einer Windows-basierten Tabelle kopiert werden, es werden allerdings nur sehr wenig Formate für die Dateneingabe akzeptiert (Borenstein, Hedges, Higgins, & Rothstein (2009).

Sharp und Sterne haben 1997 erstmals ein Standard-Meta-Analysen-Verfahren für Fixed- und Random-Effects-Analysen entwickelt; es erzeugt numerischen und grafischen Output in Form eines Forest Plots.

Die gesammelten Artikel des Stata-Journals wurden 2009 als Buch “Meta-Analysis in Stata: An Updated Collection from the Stata Journal“ veröffentlicht. Mit den Stata-Macros können verschiedene Berechnungen durchgeführt werden, wie beispielsweise Meta-Analysen basierend auf Fixed- und Random-Effects-Modell, Sensitivitätsanalysen, Subgruppenanalysen und Metaregression, Beurteilung des Publikationsbias sowie verschiedene grafische Darstellungen der Ergebnisse. Diese können sehr leicht in Word, Power Point oder auch andere Programme exportiert werden.

Im Folgenden werden anhand von Stata exemplarisch einige grundlegende Funktionen bzw. Befehle vorgestellt, die im Großteil der mit Befehlen arbeitenden Programme verwendbar sind. Die Namensgebung dieser Funktionen ist abhängig vom jeweiligen Programm (Sterne, Bradburn & Egger, 2001).

metan: Dieser Befehl bietet Methoden zur Berechnung von Meta-Analysen von Studien mit zwei Gruppen. Mit binären Daten kann die Effektstärkenmessung die Differenz zwischen Proportionen (Risk Difference), das Verhältnis von zwei Proportionen (Risk Ratio oder Relatives Risiko) oder Odds Ratio. Bei kontinuierlichen Daten kann sowohl die beobachtete als auch die standardisierte Mittelwertsdifferenz verwendet werden. Für beide Arten von Daten kann eine Analyse sowohl aufgrund des Fixed- als auch des Random-Effects-Modells berechnet werden.

meta: Diese Funktion verwendet die „Inverse-Variance"-Gewichtung, um Fixed- und Random-Effects-Modelle zu berechnen, und optional einen Forest Plot zu generieren. Man benötigt im Gegensatz zum meta-command für jede Studie Variablen, die die Effektstärken inklusive Standardfehler beinhalten.

metacum: Dies ist ein Befehl zur Durchführung von kumulativen Meta-Analysen. Die Ergebnisse können numerisch als auch grafisch dargestellt werden.

metainf: Dieser Befehl dient dazu, den Einfluss einzelner Studien auf das Gesamtergebnis der Meta-Analyse zu zeigen. Deshalb wird zur Berechnung jeweils eine Studie ausgeschlossen. In der Standardeinstellung wird eine Fixed-Effect-Analyse angezeigt.

metabias: Dieser Befehl führt den von Begg und Mazumdar (1994) und von Egger et al. (1997) vorgeschlagenen Test zur Asymmetrie des Funnel Plots aus.

metareg: Mit diesem Befehl wird eine Metaregression berechnet (Sterne, Bradburn & Egger, 2001).

2.3.4 Programme zur Effektstärkenberechnung

Clintools Software Version 4.1C

Diese kostenpflichtige Software besteht aus mehreren Einzelprogrammen, von denen einige meta-analytische Berechnungen unterstützen. Zur Durchführung von Meta-Analysen ist nur der Effect Size Generator geeignet. Damit können verschiedene Effektstärkenberechnungen, beispielsweise Cohen‘s *d* und Hedges‘ *g*, durchgeführt werden, außerdem berechnet das Programm alle dazugehörigen Konfidenzintervalle. Auf Grundlage der Effektstärken, die in einer Datenbank gespeichert werden, kann mit Hilfe des Programmes eine Meta-Analyse berechnet werden. Auch eine Reihe von Fail-Safe-*N* Berechnungen sind inkludiert (Clintools, 2008).

Weiters beinhaltet die Software Heterogenitätstests und einige andere statistische Verfahren, die allerdings nicht speziell für meta-analytische Berechnungen geeignet sind. Für anfallende Probleme steht eine Hilfe-Datei zur Verfügung. Insgesamt bietet das Programm wenig Raum für Individualisierbarkeit.

Weitere Programmteile sind der Odds Ratio Generator, der Reliable Change Generator und der Random Number Generator. Diese Funktionen sind nicht spezifisch für Meta-Analysen, sondern können für diverse mathematische und statistische Verfahren eingesetzt werden.

Effect Size (ES) 1.0

ES ist ein Programm von William Shadish zur Berechnung bzw. Umrechnung von Effektstärken und kann über http://www.assess.com/ erworben werden Es können sowohl Mittelwert, Standardabweichung und Stichprobengröße als auch dichotome Daten als Ausgangspunkt eingegeben werden. Außerdem können mit der Software andere statistische Verfahren wie *t*-Test, *F*-Test und Varianzanalyse durchgeführt werden.

Meta-Analysis Calculator

Mit dem Meta-Analysis Calculator kann man online auf der Site von Larry C. Lyons (http://www.lyonsmorris.com/ma1/index.cfm) kostenlos Effektstärken-berechnungen bzw. Transformationen von kontinuierlichen Daten in allen gängigen Formaten berechnen. Nach Auswahl des Datenformates wird man nacheinander nach relevanten Daten der jeweiligen Studien gefragt. Diese werden temporär in einer Tabelle gespeichert, die sich jedoch als Excel-Datei herunterladen lässt.

Meta-Analysis Easy to Answer (META)

META ist ein frei verfügbares Programm von David Kenny und kann zusammen mit der Programmdokumentation kostenfrei im Internet unter http://davidakenny.net/meta.htm heruntergeladen werden. Zusätzlich zur DOS Version existiert auch eine Windows Version, die der Autor zusammen mit Al Ziming entwickelt hat. Das Programm kann verschiedene Berechnungen, die im Zusammenhang mit einer Meta-Analyse stehen, durchführen und ist wohl vor allem hilfreich für die Umrechnung von Effektstärken. Diese können beispielsweise nach Stichprobengröße oder Varianz der Studien sortiert werden. Auch Homogenitätstests über die Studien können durchgeführt werden.

Meta-Analysis Tool

Das Meta-Analysis Tool besteht aus dem EffectSize Calculator Converter und dem ForestPlot Tool. Beides sind Excel-Vorlagen, die die namensgebenden Funktionen zur Verfügung stellen und können im Internet von http://psych.cf.ac.uk/home2/mat/ heruntergeladen werden.

METASTATS

METASTATS ist eine Programmsammlung für den programmierbaren wissenschaftlichen Taschenrechner HP49G, von der auch ein Emulator für Windows-

PCs verfügbar ist. Es lassen sich diverse statistische Verfahren wie Effektstärkenberechnungen, *t*-Tests und Korrelationsanalysen durchführen.

Practical Meta-Analysis Effect Size Calculator

Der Practical Meta-Analysis Effect Size Calculator ist eine webbasierte Anwendung zur Berechnung bzw. Transformation von Effektstärken. Weitere meta-analytische Berechnungen können damit nicht durchgeführt werden.

Software for Permutation Methods: A Distance Function Approach
Statistical Software by Paul W. Mielke Jr.

Diese große Kollektion von ausführbaren DOS- und Windows Programmen ist frei verfügbar. Neben vielen anderen statistischen Verfahren, die für meta-analytische Berechnungen nicht von Bedeutung sind, beinhaltet es die Berechnung von Fisher's kombinierten *p*-Werten.

2.4 Konklusio

Bei der Recherche zu diesem Review stellte sich heraus, dass viele ältere Programme nicht mehr gewartet werden bzw. nicht mehr zur Verfügung stehen. Manche der DOS-Programme wurden durch Windows-Versionen ersetzt und ein bedeutender Anteil der Software ist kostenfrei zum Download verfügbar.

Die befehlsgesteuerten Programme weisen oft eine viel größere Funktionsbreite auf als solche mit grafischer Benutzeroberfläche; allerdings erfordern diese ein gewisses Maß an Expertise.

II. Usability-Studie

1. Usability: Theoretischer Hintergrund

1.1 Definition

Im deutschen Sprachraum wird Usability häufig sowohl im Sinne von Gebrauchstauglichkeit als auch Benutzerfreundlichkeit verwendet. Dabei bezieht sich der Begriff nicht nur auf Software, sondern auf alle Produkte, mit denen der Nutzer in irgendeiner Weise interagiert. Es existieren viele Definitionen, die im Kern alle das gleiche beschreiben, allerdings verschiedene Aspekte in den Vordergrund stellen.

Nach Nielsen (2000), einem Mitbegründer des Begriffs und in der Folge Experte auf diesem Gebiet, ist Usability durch schnelle und einfache Erlernbarkeit (learnability), Möglichkeit zur effizienten Nutzung (efficiency), Einprägsamkeit (memorability), Fehlertoleranz (errors) und zufriedenstellende Nutzung (satisfaction) gegeben.

In DIN EN ISO 9241-110 ist Usability definiert als "das Ausmaß, in dem ein Produkt durch bestimmte Nutzer in einem bestimmten Nutzungskontext genutzt werden kann, um bestimmte Ziele effektiv, effizient und zufriedenstellend zu erreichen".

Eine relativ triviale Definition von Usability liefert auch Krug (2002, S. 5): „Schließlich meint Usability einfach nur, dass man darauf achten soll, dass etwas richtig funktioniert: dass eine Person mit durchschnittlichen (oder auch unterdurchschnittlichen) Fähigkeiten und Erfahrungen das Ding – sei es eine Website, oder ein Kampfjet oder eine Drehtür – in beabsichtigter Weise benutzen kann, ohne hoffnungslos frustriert zu werden."

Jordan (1998) hat ein Drei-Komponenten-Modell der Usability entwickelt, welches Guessability, Learnability und Experienced User Performance (EUP) beinhaltet. Diese Begriffe beschreiben die Effektivität, Effizienz und Zufriedenheit, mit der der Nutzer jeweils bei erstmaliger, wiederholter und routinierter Anwendung mit einem bestimmten Produkt bestimmte Aufgaben ausführen kann.

Die beschriebenen Definitionen beziehen sich generell auf die rationalen Aspekte der Nutzung und beinhalten die einfache Zielerreichung, Zweckgebundenheit, Intuitivität und Einprägsamkeit eines Produkts, wobei Jordan (1998) noch zwischen dem Erfahrungsgrad der Benutzer unterscheidet.

Der Terminus Usability ist somit klar von der oft damit verwechselten User Experience abzugrenzen. Die User Experience meint die Gesamtbeurteilung eines interaktionsfähigen Produktes, in die neben der durch die praktischen Aspekten beschriebenen Usability auch die emotional geprägte Benutzerfreundlichkeit eingeht. Hierzu gehört beispielsweise das „Look & Feel“, was die Darstellung der Benutzeroberfläche sowie individuelle Interaktion mit dem Produkt meint. Somit ist Usability nur als ein Teilbereich der User Experience zu verstehen.

1.2 Entwicklung

Grundsätzlich stellt die Usability keine eigenständige Disziplin wie die Ergonomie dar, sondern eine Qualität eines technischen Systems. Sie ist ein Ziel der Gestaltung nach den Erkenntnissen der Ergonomie und kommt generell aus dem Bereich der Human-Computer-Interaction (HCI) (Sarodnick & Brau, 2011). Die Beschäftigung mit diesem Thema hat in den letzten 20 Jahren erheblich zugenommen, nicht zuletzt wegen der nachgewiesenen Steigerung der Produktivität durch Produkte bzw. Systeme mit hoher Usability. Mittlerweile gibt es ein weites Feld an Literatur, das sich mit der Usability von Software beschäftigt, Ratschläge an Programmierer und Designer richtet und gleichzeitig Evaluationsverfahren vorschlägt.

1.3 Usability Engineering

Bei der Entwicklung von Software sind meist nur Experten beteiligt, die andere Anforderungen und Erwartungen an diese haben als die späteren User. Dieser Mangel an Vorstellung davon, welche Schwierigkeiten Benutzer später mit der Software haben können, kann große Probleme mit sich bringen, die sich beispielsweise in Form von längeren Bearbeitungs- und Einarbeitungszeiten sowie Verringerung von Produktivität äußern.

Mittlerweile wird schon während des Entwicklungsprozesses versucht, eine möglichst hohe Usability und Benutzerorientierung zu erzeugen. Diesen Prozess

bezeichnet man als Usability Engineering. Die Phasen des Usability-Engineerings lassen sich nach Sarodnick und Brau (2011, S. 88) in folgende Schritte einteilen:

a. Analyse der Arbeit und des Arbeitsumfeldes
b. Analyse der Benutzergruppen
c. Bestimmung von Anforderungen
d. Entscheidung über Funktionalität und Ableitung eines Handlungs- und Bedienungskonzeptes
e. Entwicklungsbegleitende Evaluation und Verbesserung des Systems
f. Einführung und Schulung
g. Weiterentwicklung.

Der von Nielsen (2000) vorgeschlagene Usability Engineering Lifecycle besteht gar aus elf Schritten, wobei für eine benutzerfreundliche Produktgestaltung nicht zwangsläufig alle Schritte berücksichtigt werden müssen. Beide Ansätze weisen Ähnlichkeiten auf, wobei Nielsen (2000) noch eine etwas genauere Beschreibung liefert. In jedem Fall ist es schon in diesem Stadium sehr wichtig, sich nach den Interessen der späteren Nutzer zu richten und möglichst genau deren Bedürfnisse zu evaluieren.

1.4 DIN EN ISO 9241-110 Grundsätze der Dialoggestaltung

In der ISO-Norm 9241, die den deutschen Titel „Ergonomie der Mensch-System-Interaktion“ trägt, werden die Qualitätsrichtlinien zur Sicherstellung der Ergonomie interaktiver System beschrieben. Die Norm ersetzt die ISO 9241-10 von 1995 (deutsche Fassung DIN EN ISO 9241-10 von 1996) und seit 2006 liegt nun DIN EN ISO 9241-110 als Nachfolger der DIN EN ISO 9241-10 öffentlich vor. Teil 110 beschäftigt sich im speziellen mit den Grundsätzen zur Dialoggestaltung. Die sieben Grundsätze werden mit den Begriffen Aufgabenangemessenheit (suitability for the task), Selbstbeschreibungsfähigkeit (self-descriptiveness), Erwartungskonformität (conformity with user expectations), Lernförderlichkeit (suitability for learning), Steuerbarkeit (controllability), Fehlertoleranz (error tolerance) und Individualisierbarkeit (suitability for individualization) bezeichnet.

Aufgabenangemessenheit (suitability for the task)

Aufgabenangemessenheit ist laut der ISO Norm 9241-110 (2006) dann gegeben, wenn Funktionalität und Dialog auf die charakteristischen Eigenschaften der Arbeitsaufgabe zugeschnitten sind. Die Software soll den User ohne vermeidbare Belastungen dabei unterstützen, seine Aufgabe möglichst schnell und effizient erledigen zu können. Dazu gehört die Vermeidung überflüssiger Schritte und Informationen, die verwirren könnten, als auch die Bearbeitung inhaltlich zusammenhängender Arbeitsschritte in einer Folge anzubieten.

Es muss gewährleistet sein, dass der Nutzer mit geringem Einsatz von Zeit, Geduld, Gedächtnis- und Transferleistung an das Ziel gelangt. (Rampl, 2007)

Selbstbeschreibungsfähigkeit (self-descriptiveness)

Ein Programm ist selbstbeschreibungsfähig, wenn es dem User ermöglicht, zu jeder Zeit zu erkennen, wo in der Anwendung er sich gerade befindet bzw. welche Aktionen möglich sind und wie sie durchgeführt werden können. Auch der Systemzustand soll klar ersichtlich sein (DIN EN ISO 9241-110, 2006).

Beispiele dafür sind unter anderem ein Fortschrittsbalken, der die verbleibende Dauer einer Bearbeitung, die gerade durchgeführt wird, anzeigt oder ein Dateneingabefeld, welches das notwendige Format der Daten erkennen lässt.

Erwartungskonformität (conformity with user expectations)

Erwartungskonformität bedeutet gemäß ISO Norm 9241-110 (2006), dass ein Dialog den aus dem Nutzungskontext heraus vorhersehbaren Belangen des Nutzers sowie allgemein anerkannten Konventionen entspricht.

Dieser Dialoggrundsatz beruht darauf, dass menschliches Handeln sowohl durch erlernte Verhaltensmuster als auch durch lernpsychologische Faktoren geprägt ist. So können Zusammenhänge umso schneller erfasst werden, je schneller bekannte Muster erkannt werden. Variieren nun wichtige Merkmale zur Erkennung von beispielsweise Dateneingabe oder Erstellen einer Grafik, kann der Nutzer kein Muster speichern, das ihm die Nutzung eines anderen Programms erleichtern würde. Im Gegensatz zu

unerfahrenen Benutzern haben Menschen mit größerer Expertise jedoch bereits bestimmte Erwartungen an Programme hinsichtlich der Konformität.

Lernförderlichkeit (suitability for learning)

Ein Dialog sollte den Benutzer beim Erlernen des Dialogsystems unterstützen und anleiten. Das Ausmaß dieser Systemeigenschaft wird durch die Lernförderlichkeit beschrieben (DIN EN ISO 9241-110, 2006).

Das bedeutet, dass ein System sowohl in einem gewissen Ausmaß intuitiv erlernbar sein sollte („learning by doing"), als auch Hilfe-Funktionen beinhaltet, die leicht zu finden und einfach zu verstehen sind. Viele Programme bieten auch ein Tutorial, welches sich beim Öffnen der Software automatisch startet. Alle grundlegenden Schritte werden kurz erklärt und können gleich geübt werden. Vor allem unerfahrenen Nutzern erleichtert das den Einstieg von selbst.

Steuerbarkeit (controllability)

Steuerbarkeit beschreibt die Dialogeigenschaft, die den Benutzer in die Lage versetzt, während der Aufgabenerledigung Richtung und Geschwindigkeit des Dialogablaufs steuern zu können. (DIN EN ISO 9241-110, 2006).

Dieses Dialogprinzip beeinflusst in einem hohen Ausmaß die User Experience, da es sich auf die individuell erlebte Kontrollmöglichkeit auswirkt und somit auch affektive Aspekte mit einschließt. Dem Nutzer sollte es beispielsweise möglich sein, seine Ergebnisse zu jedem Zeitpunkt zu speichern bzw. alte Daten bei Bedarf schnell wieder abrufen zu können. Eine einfache Navigation innerhalb des Programms zu jeder Zeit sowie der Wechsel zwischen verschiedenen Datensätzen, die bearbeitet werden sollen, stellen wichtige Kriterien dafür da. Auch muss es möglich sein, die Aufgabe, an der gearbeitet wird, jederzeit ohne Konsequenzen abzubrechen bzw. Teile davon rückgängig zu machen.

Fehlertoleranz (suitability for individualization)

Fehlertoleranz ermöglicht dem Benutzer, trotz erkennbar fehlerhafter Eingaben das beabsichtigte Arbeitsergebnis entweder mit keinem oder mit minimalem Korrekturaufwand zu erreichen. Dies wird mit folgenden Mitteln erreicht: Fehlererkennung und -vermeidung (Schadensbegrenzung); Fehlerkorrektur oder Fehlermanagement, um mit Fehlern umzugehen, die sich ereignen (DIN EN ISO 9241-110, 2006).

Die potentiellen Fehler, die Nutzer verursachen oder die ihnen begegnen können, lassen sich nach Rampl (2007) wie folgt klassifizieren:

Vermeidbare Fehler

Diese Art von Fehlern treten aufgrund mangelnder Beschäftigung mit dem Nutzerverhalten auf und wären bei sorgfältiger Auseinandersetzung mit der Zielgruppe vermeidbar.

Bekannte, nicht vermeidbare Fehler

Nicht alle bekannten Fehler lassen sich vermeiden. Ein Vertippen mit der Tastatur ist nur ein Beispiel für Fehler, mit denen man rechnen muss, weil sie sich nicht ausschließen lassen. Für alle vorhersehbaren Fehler muss es deshalb einfache, klar erkennbare Korrekturmöglichkeiten geben.

Nicht antizipierbare Fehler

In diese Klasse von Fehlern fallen all diejenigen, die entweder aufgrund unerwarteten Nutzerverhaltens passieren oder bei Sonderfällen, die im Vorfeld schwer identifizierbare Programmierfehler auslösen.

Beispiele für ein fehlertolerantes Programmverhalten sind der „Zurück“-Button in Microsoft Office Programmen, die Rückmeldung eines Programms über fehlerhafte oder unvollständige Dateneingabe oder die Autokorrektur-Option von Microsoft Office Word.

Individualisierbarkeit

Individualisierbarkeit wird in der ISO Norm 9241-110 (2006) durch die Anpassungsfähigkeit des Systems an individuelle Fähigkeiten und Bedürfnisse definiert. Demzufolge soll der Benutzer die Mensch-System-Interaktion und die Darstellung von Informationen dahingehend ändern können.

Dies ist besonders wichtig, wenn ein User ein System oft nutzt bzw. für geübte Benutzer, um das System für häufig verwendete Aufgaben anzupassen. In vielen Programmen ist es möglich, die Funktionstasten mit individuell gewählten Befehlen zu belegen oder in der Bedienoberfläche selbst sogenannte Shortcuts zu erstellen, die die Arbeit erleichtern und die benötigte Zeit unter Umständen wesentlich verkürzen können. In den meisten Programmen ist auch möglich, eigene Vorlagen zu erstellen und auf diese immer wieder zurück zu greifen.

1.5 Usability-Evaluation

1.5.1 Ziele

Nach Gediga und Hamborg (2002, S. 41) lassen sich die Evaluationsziele im Bereich der Software-Ergonomie üblicherweise durch drei einfache Fragen charakterisieren:

1. „Which is better?“ Die Evaluation zielt auf einen Vergleich zwischen alternativen Softwaresystemen, z. B., um für einen bestimmten Anwendungszweck das software-ergonomisch beste Produkt auszuwählen, um im Rahmen des Prototypings zwischen Designalternativen zu entscheiden oder die Optimierung von Versionen zu kontrollieren. [...]

2. „How good?“ Bei dieser Zielsetzung geht es um die Bestimmung der Ausprägung bestimmter gewünschter oder geforderter Systemeigenschaften. Hierzu zählt die Bewertung von Software zum Ende des Entwicklungsprozesses [...] oder auch die Überprüfung von Software auf Normkonformität etc.

3. „Why bad?“ Die Evaluation zielt darauf, Schwachstellen als Ausgangspunkt für die Gestaltung oder direkt Gestaltungsvorschläge zu liefern. [....]

Die ersten beiden Ziele kann man dem Konzept der summativen Evaluation zuordnen. Diese wird durchgeführt, nachdem der Entwicklungsprozess des Produkts bereits beendet ist und findet beispielsweise Anwendung, wenn mehrere Programme verglichen werden sollen. Das dritte Ziel ist der Klasse der formativen Evaluation zuzuordnen, welche bereits während der Produktentwicklung stattfindet und zur Verbesserung dienen soll.

1.5.2 Methoden

Die Evaluationsmethoden lassen sich nach mehreren Prinzipien unterteilen. Eine sehr gängige Unterscheidung basiert darauf, welche Art von Gutachtern für die Evaluation herangezogen wird. Man unterscheidet die expertenorientierten oder analytischen Methoden, sowie die benutzerorientierten oder empirischen Methoden (Schweibenz & Thissen, 2003). Eine analytische Evaluation wird anhand von Expertenmeinungen oder strukturiertem Testen einiger Funktionen vollzogen. Beispiele hierfür sind Aufgabenanalysen, Checklisten, Expertenbeurteilungen und Cognitive Walkthroughs.

Bei den benutzerorientierten bzw. empirischen Methoden der Usability-Evaluation werden künftige, reale Benutzer mit dem Produkt konfrontiert und während der Interaktion mit diesem Produkt beobachtet (Schweibenz & Thissen 2003). So lässt man „echte” Nutzer das Produkt beurteilen bzw. testen, beispielsweise zum Vergleich mehrerer Programme oder auch zur Beurteilung einer in einem Betrieb verwendeten Software. Dies geschieht unter anderem durch Usability Tests, Befragungen der Nutzer (mündlich oder mit Hilfe von Fragebögen), Verhaltensbeobachtung, Verwendung von Log Files oder Eye- bzw. Mousetracking. Beide Methoden inklusive einiger Beispiele werde ich im Folgenden erläutern.

1.5.2.1 Analytische Evaluation

Der Ausdruck Analytische Evaluation ist synonym mit dem Terminus “Usability Inspection”. Dieser Begriff wurde nach Nielsen (1994) zum ersten Mal öffentlich bei der ACM CHI’90, einer Konferenz zur Mensch-Computer-Interaktion in Seattle, am 04. April 1990 verwendet. Dort wurden auch die beiden ersten Methoden, der Cognitive Walkthrough und die Heuristische Evaluation, vorgestellt. Zwei Jahre später, auf der CHI’92, wurde beschlossen, den Terminus „Usability Inspection“ für alle ähnlichen Methoden zu verwenden (Nielsen, 1994). In den darauffolgenden Jahren wurden viele weitere Verfahren zur Usability-Evaluation von Benutzerschnittstellen entwickelt, deren gemeinsame Basis die Beurteilung und Prüfung durch Experten bildet. Diese Experten können laut Nielsen (1994) sowohl Usability Professionals, aber auch Softwareentwickler oder Nutzer mit entsprechendem Fachwissen sein.

Diese Art der Analyse wird durch sogenannte Usability Professionals durchgeführt, die das Produkt basierend auf Heuristiken beurteilen. Expertenorientierte Verfahren lassen sich bereits mit einer geringen Anzahl von Gutachtern durchführen. Diese Methoden haben daher den Vorteil, dass sie relativ schnell und flexibel durchführbar und auswertbar sind. Da diese Methoden nicht so strengen formellen Anforderungen unterliegen wie die benutzerorientierten Methoden, werden sie häufig auch als informelle Methoden der Evaluation bezeichnet (Schweibenz & Thissen 2003).

Heuristische Evaluation

Die heuristische Evaluation steht nach Nielsen (1994) als eine eher formlose Methode zur Verfügung, um von Usability Experten beurteilen zu lassen, ob ein User Interface den bestehenden Prinzipien der Benutzerfreundlichkeit (Heuristiken) entspricht. Jeder Gutachter untersucht das Produkt alleine. Die Prüfer sollten ihre Ergebnisse einander erst mitteilen, nachdem alle Untersuchungen abgeschlossen worden sind. So wird eine unabhängige und unbeeinflusste Prüfung gewährleistet.

Diese Art der Evaluation wird vorwiegend als Teil des Usability Engineerings verwendet und ist so optimalerweise ein Teil des ganzen Entwicklungsprozesses.

Ein von Nielsen (2000) durchgeführtes Experiment zeigt, dass Usability Experten mit Hilfe heuristischer Evaluation mehr Probleme identifizieren als normale Anwender.

Nielsen und Molich (1990) liefern folgendes Bild zum Einfluss der Gutachterzahl:

- Ein Gutachter entdeckt 38% der Usability Probleme.
- Fünf Gutachter entdecken 70% der Usability Probleme.
- Zehn Gutachter entdecken 89% der Usability Probleme.

Die Anzahl der Probleme wurde von den Autoren bestimmt.

Aus Sicht der Kosten-Nutzen-Rechnung hat sich in der Praxis gezeigt, dass es genügt, ca. drei bis fünf Beurteiler einzusetzen.

Guideline Review

Der Guideline Review prüft Software anhand von umfangreichen Richtlinien. Die Einhaltung dieser Guidelines soll gewährleisten, dass das Produkt bestimmte Merkmale aufweist, die schließlich Usability erzeugen.

Die Methode erfordert ein hohes Maß an Expertise und außerdem einen immensen Zeitaufwand, so dass in der Praxis meist darauf verzichtet wird. Oft werden hingegen nur kurze Checklisten vorgegeben, um mehrere Programme zu vergleichen und sich einen Überblick zu verschaffen.

Die Guideline Inspection wird in zwei Teilen durchgeführt. Im ersten Teil werden Usability-Kriterien überprüft, was einer Standard Inspection im weiteren Sinn entspricht. Beim zweiten Teil handelt es sich um eine Feature Inspection, die das Vorhandensein von notwendigen Informationen untersucht (Binder, 2006).

Cognitive Walkthrough

Der Cognitive Walkthrough ist eine Usability Inspection-Technik. Meist kommt sie aufgrund ihrer kostengünstigen Durchführung schon früh in der Entwicklung eines Produktes zum Einsatz.

Es wird in einer relativ detaillierten Art und Weise simuliert, wie ein User mit dem Interface arbeitet und jeder Schritt in der Computer-Human-Interaction wird genau durchleuchtet. Somit wird geprüft, ob der Nutzer das Programm korrekt bzw. in der vorgesehenen Weise bedient. Der Teilnehmer hält fest, wie er bestimmte Operationen durchführt, und so können Produktentwickler Hilfestellung zur Problembehebung erhalten (Nielsen, 1994).

Die Technik basiert auf den theoretischen Grundlagen der Kognitionsforschung und besonderer Fokus wird auf die Erlernbarkeit bzw. Lernförderlichkeit des Programmes gelegt. Die Struktur des Programmes soll von Beginn an nachvollziehbar sein und so dem Nutzer die intuitive Bedienung erleichtern.

Pluralistic Walkthrough

Bei dieser Methode wird, wie auch beim Cognitive Walkthrough, die Usability einer Software mittels Durchdenken der Arbeitsabläufe ermittelt. In Treffen, zu denen sich User, Entwickler und nach Bedarf auch andere Personengruppen zusammenfinden, wird gemeinsam ein virtuelles Szenario durchgegangen und es werden Usability-Aspekte der verschiedenen Schritte des Szenarios diskutiert. Die Teilnehmer zeichnen ihren Lösungsweg detailliert auf und der Moderator oder ein erfahrener Nutzer erklärt im Anschluss die richtige Lösung. Die Ergebnisse werden diskutiert und die Ursache eventueller Abweichungen erarbeitet. In einer anderen Art des Verfahrens werden User von Experten durch das Programm geführt und die optimale Nutzung der Programmfeatures sowie die schnellste Zielerreichung werden aufgezeigt.

1.5.2.2 Empirische Evaluation

Usability-Test

Ein Usability-Test zählt zu jenen Evaluierungsverfahren, die den größten personellen und auch räumlichen Aufwand erfordern. Man benötigt Testpersonen, die bestimmte Kriterien erfüllen müssen sowie einen gut eingeschulten Testleiter. Bei jeder

Erhebung müssen Anforderungen und Ziele definiert und dann Methoden dafür ausgewählt werden. Nur so ist ein gewinnbringendes Gelingen gesichert.

Nach Nielsen (2000) ist ein Nutzertest mit echten Benutzern die beste Usability Methode und durch nichts zu ersetzen, weil er direkte und grundlegende Informationen darüber liefert, wie Personen die Software benutzen und was ihre exakten Probleme mit einer bestimmten zu testenden Anwenderschnittstelle sind.

Eyetracking, Mousetracking

Bei Eyetracking und Mousetracking findet die Usability-Evaluation mithilfe von Blickbewegungs- bzw. Cursorpositionsmessungen statt. Somit kann eine Analyse der Aufmerksamkeit der Probanden durchgeführt werden (Völkel, 2007).

Beim Eyetracking werden sowohl die Bewegungen sowie die Verweildauer auf bestimmten Punkten aufgezeichnet. Man kann auch ohne die Kommentare der Testpersonen feststellen, wie lang sie sich mit den einzelnen Programmfeatures auseinandersetzen bzw. welche Punkte am Bildschirm bevorzugt betrachtet werden.
Beide Methoden kommen oft kombiniert zum Einsatz, so dass beispielsweise nachvollzogen werden kann, ob nach längerer Begutachtung eines Menüpunkts dieser auch tatsächlich angeklickt worden ist.

Thinking Aloud

Die Methode des lauten Denkens fordert den User dazu auf, seine Gedanken auszusprechen und seine Tätigkeiten zu kommentieren, während er sich mit dem zu bewertenden Objekt auseinandersetzt. Durch die Verbalisierung der Gedanken kann die Auseinandersetzung des Nutzers mit den Dialogen deutlich besser als mit anderen Methoden nachvollzogen und Probleme, die sich aus Fehlannahmen oder Missverständnissen ergeben, erkannt werden (Nielsen, 2000).

Das Verfahren ist vor allem in den USA weit verbreitet und nach Nielsen (2000) die einzige Methode, die auch allein durchgeführt werden kann. Die Testperson führt dabei vorgegebene Aufgaben durch und die Kommentare und Bewegungen sowie die digitalen Bewegungen werden durch Videoaufnahmen bzw. Log Files festgehalten.

Durch manche direkten Aussagen können bereits notwendige Änderungen erkannt und durchgeführt werden.

Fokusgruppen

Beim Einsatz der Fokusgruppen-Technik diskutiert eine Gruppe von meist sechs bis neun Personen über das Produkt. Dieses Verfahren kommt häufig schon vor der Markteinführung zum ersten Mal zur Anwendung. Der größte Vorteil liegt darin, dass speziell im Bereich Software sowohl Entwickler als auch Nutzer an der Sitzung teilnehmen können und so ein direkter Austausch stattfindet.

Geleitet werden diese Sitzungen von einem gut geschulten Moderator, der dafür verantwortlich ist, dass die Diskussion am Laufen gehalten wird, dass sich alle Teilnehmer gleichermaßen an dem Gespräch beteiligen können und dass die gewünschten Themen angesprochen werden (Krueger & Casey, 2000).

Logfile-Auswertung

Gerade im Bereich der Softwareentwicklung kommt oft die Logfile-Auswertung zum Einsatz. Bei der „Logfile Analyse“ (Nielsen, 1994) wird automatisch aufgezeichnet, welche Funktionen des Programms wie oft verwendet werden. Dadurch kann unter anderem festgestellt werden, wie oft einzelne Funktionen benutzt werden. So kann beispielsweise der Zugriff darauf in einer zukünftigen Programmversion erleichtert werden. Der Vorteil des Verfahrens ist, dass die Beurteilung darüber, wie gut der Nutzer sich bei vorgegebenen Aufgaben im Programm zurechtfindet, nach objektiven Kriterien vorgenommen werden kann und nicht auf der Einschätzung des Nutzers selbst beruht. Aus Datenschutzgründen müssen die Teilnehmer vor der Durchführung eines derartigen Tests unbedingt über die Aufzeichnungen in Kenntnis gesetzt werden.

1.5.2.3 Formative und summative Evaluation

Eine andere mögliche Art der Unterscheidung der Evaluationsmethoden kann basierend auf dem Zeitpunkt, zu dem die Erhebung durchgeführt wird, vorgenommen

werden. Dabei unterscheidet man zwischen formativen und summativen Verfahren der Evaluation.

Die formative Evaluation findet bereits während der Produktentwicklung statt und bietet somit die Möglichkeit zur Verbesserung. Kaspar, Hamborg, Sackmann und Hesselmann (2010) konnten mit einer Studie zur Effektivität formativer Evaluation bei der Entwicklung gebrauchstauglicher Software zeigen, dass sich die Anzahl der Usability-Probleme nach einem Evaluations-Gestaltungszyklus signifikant reduziert.

Die summative Evaluation dagegen meint die abschließende Beurteilung eines Produktes und findet erst nach Fertigstellung statt. Sie kann auch dazu dienen, mehrere Produkte bzw. Programme zu vergleichen. Es werden größtenteils Gesamteinschätzungen erhoben und bewertet.

2. Studiendesign

Ein Usability-Test hat eine viel größere Aussagekraft als eine einfache Befragung der Benutzer über ihre Zufriedenheit, Akzeptanz oder Handhabung der Software. Der Anwender ist dazu angehalten, bestimmte Aufgaben zu bearbeiten und wird so direkt mit dem System konfrontiert. Diese Eigenschaften gaben den Ausschlag für die Wahl dieser Methode. Auch Nielsen (2000, S. 165) sieht eindeutige Vorteile in diesem Verfahren:

„User testing with real users is the most fundamental usability method and is in some sense irreplaceable, since it provides direct information about how people use computers and what their exact problems are with the concrete interface being tested."

Aufgrund der großen Anzahl verfügbarer Programme zur Berechnung von Meta-Analysen ist es schwer, auf den ersten Blick die richtige Entscheidung zu treffen. Neben speziellen Anforderungen an die Software und eigener Erfahrungen fließt auch die Benutzerfreundlichkeit in die Programmwahl mit ein. Da es kaum umfassende Studien über diese speziellen Statistikprogramme gibt, widmet sich vorliegende Untersuchung einem Vergleich von sechs verschiedene Programmen hinsichtlich ihrer Usability. Zu diesem Zweck wurde jedes Programm unmittelbar nach der Bearbeitung vorgegebener

Aufgaben von allen Teilnehmern mittels des Isometrics S Verfahrens beurteilt. Dies ist nach Gediga, Hamborg und Willumeit (2000) das Hauptanwendungsgebiet des IsoMetrics.

In Anlehnung an die Studie von Bax et al. (2007) wurden für den Test sechs Programme ausgewählt, auch um die Belastung für die Teilnehmer in vertretbarem Ausmaß zu halten. Ein weiteres Einschlusskriterium neben einer grafischen Benutzeroberfläche war, dass die Programme aktuell noch vermarktet bzw. gewartet werden, weswegen nur vier der fünf von Bax et al. (2007) getesteten Programmen berücksichtigt werden konnten. Um ein breiteres Feld abzudecken, wurden zusätzlich noch ein allgemeines Statistikprogramm mit meta-analytischen Funktionen und ein kostenfreies Programm, das erst seit 2008 verfügbar ist, in die Studie aufgenommen. Nähere Informationen zu diesen Programmen können in Anhang 2 nachgelesen werden.

Die Beurteiler wiesen unterschiedliche Erfahrung in der Durchführung von Meta-Analysen auf und setzten sich aus Studenten, Absolventen oder Lehrbeauftragten der Fakultät für Psychologie der Uni Wien zusammen. Die Erfahrung wurde als Kovariate in die Berechnung mit aufgenommen. Als Testinstrument wurde der Fragebogen Isometrics verwendet, auf den im Vorfeld schon genauer eingegangen wurde. Im Folgenden wird die Stichprobe sowie die Durchführung der Studie im Detail beschrieben.

2.1 Methode

2.1.1 Fragestellung

Mittels der vorliegenden Usability-Studie soll geklärt werden, welchem Programm von den Teilnehmern die höchste Benutzerfreundlichkeit zugesprochen wird. Dabei sind signifikante Mittelwertsunterschiede sowie der Einfluss von Geschlecht, Kompetenzlevel und Vorerfahrung von Interesse.

2.1.2 Verwendeter Fragebogen

Einleitung und demografische Daten

Im ersten Teil des Fragebogens wurden die Teilnehmer nochmals über den Hintergrund der Studie aufgeklärt. Anschließend wurden sie gebeten, Angaben zu Schulbildung, Alter, Anzahl bereits berechneter Meta-Analysen und Vorerfahrung mit meta-analytischen Programmen im Allgemeinen zu machen.

Angaben zur Durchführung

In diesem Teil des Fragebogens wird erklärt, welche Aufgaben in den Programmen bearbeitet werden sollen. Es wurde die Eingabe von Daten, Berechnung einer Analyse nach Fixed- und Random-Effects-Modell, Durchführung einer kumulativen Meta-Analyse und Darstellung von Funnel und Forest Plot abgefragt. Die Aufgaben sollten die grundlegenden Funktionen zur Durchführung einer Meta-Analyse abdecken und deren Bearbeitung sich gleichzeitig in einem annehmbaren zeitlichen Rahmen bewegen. Zusätzlich werden die Teilnehmer darauf hingewiesen, dass es sein kann, dass nicht jedes Programm über alle der genannten Funktionen verfügt.

Anschließend wird die Bearbeitungsreihenfolge der Programme vorgegeben. Um Reihenfolgeneffekte zu vermeiden, wurde den Teilnehmern chronologisch nach dem Zeitpunkt ihrer Testung ein anderes Startprogramm sowie eine andere Programmreihenfolge zugewiesen. Sechs Sequenzen wurden so gewählt, dass bei jeder davon ein anderes Programm an erster Stelle steht und sich keine direkte Programmabfolge wiederholt, das heisst, jedes Programm ist in jeder Sequenz von einem unterschiedlichen Programm gefolgt. Dies wurde als ausreichend randomisiert angesehen. Diese Abfolgen wurden mit annähernd gleicher Häufigkeit vorgegeben:

CMA – MIX – MetaAnalyst – NCSS – MetaWin – RevMan
MetaAnalyst – RevMan – MIX – MetaWin – CMA – NCSS
MetaWin – MetaAnalyst – CMA – RevMan – NCSS – MIX
MIX – NCSS – RevMan – CMA – MetaAnalyst – MetaWin

NCSS – CMA – MetaWin – MIX – RevMan – MetaAnalyst
RevMan – MetaWin – NCSS – MetaAnalyst – MIX – CMA

IsoMetrics S

Zuletzt wurde die Bearbeitung des IsoMetrics S Fragebogens erklärt und ein Itembeispiel gegeben. Dieses validierte Inventar zur summativen Usability-Evaluation von Gediga et al. (2000) wird anschließend zur Beurteilung der einzelnen Programme vorgegeben. Das Verfahren bezieht sich auf die ISO Norm 9241 und beinhaltet, wie auch die Norm, die Teilbereiche Aufgabenangemessenheit (15 Items), Selbstbeschreibungsfähigkeit (12 Items), Steuerbarkeit (11 Items), Erwartungskonformität (8 Items), Fehlerrobustheit (15 Items), Individualisierbarkeit (6 Items) und Erlernbarkeit (8 Items). Die Beurteilung erfolgt mit Hilfe einer fünfstufigen Likertskala, wobei eine Bewertung von eins „stimmt nicht“, zwei „stimmt wenig“, drei „stimmt mittelmäßig“, vier „stimmt ziemlich“ und von fünf „stimmt sehr“ bedeutet.

2.1.3 Beschreibung der Stichprobe

Die Studie richtete sich an Studenten, Absolventen und Mitarbeiter des Arbeitsbereichs Forschungsmethoden der Psychologischen Fakultät der Universität Wien, die sich bereits in irgendeiner Art und Weise mit der Thematik Meta-Analysen beschäftigt haben. Meta-Analysen stellen ein großes Interessensgebiet dieses Arbeitsbereichs dar, so dass den Studierenden des zweiten Abschnitts eine Vielzahl an Veranstaltungen zu diesem Thema angeboten werden. Einschlusskriterium zur Teilnahme an der Usability-Studie war, mindestens eine dieser Veranstaltungen besucht zu haben, um die entsprechende Expertise mitzubringen. Die potentiellen Teilnehmer erhielten eine Einladung per E-Mail, in der die Studie kurz beschrieben und der Terminvergabemodus erklärt wurde.

Insgesamt haben sich 46 Personen zur Teilnahme bereiterklärt, wovon 28 (61%) weiblich und 18 (39%) männlich waren. Das Alter der Teilnehmer bewegte sich

zwischen 23 und 46 Jahren, wobei das Mittel 28.5 Jahre beträgt. 33 der Personen waren Studenten, drei Absolventen und zehn Mitarbeiter der Fakultät für Psychologie.

Da diese sich in ihrem Erfahrungsgrad bezüglich Meta-Analysen unterschieden, wurde eine objektive Klassifizierung der Teilnehmer, basierend auf Kriterien wie zum Beispiel Teilnahme an fachspezifischen Seminaren oder selbst durchgeführten Meta-Analysen, in zehn Level vorgenommen, wobei Level eins die geringste und Level zehn die höchste Kompetenz widerspiegelt. Die Verteilung wird in Abbildung 1 dargestellt, die Zuordnung zu den Rängen erfolgte nach folgenden Kriterien:

Level 1: Mitarbeiter des Arbeitsbereichs Forschungsmethoden mit abgeschlossenem Doktorat und mehreren veröffentlichten Meta-Analysen

Level 2: Mitarbeiter des Arbeitsbereichs Forschungsmethoden und mindestens eine veröffentlichte Meta-Analyse

Level 3: Absolventen des Arbeitsbereichs Forschungsmethoden im Doktoratsstudium oder mit bevorstehender Veröffentlichung einer Meta-Analyse

Level 4: Absolventen des Arbeitsbereichs Forschungsmethoden mit abgeschlossener Diplomarbeit zum Thema

Level 5: Studienassistenten des Arbeitsbereichs Forschungsmethoden mit mindestens einem absolvierten Seminar zum Thema

Level 6: Diplomanden mit mindestens zwei abgeschlossenen fachspezifischen Seminaren

Level 7: Diplomanden mit einem abgeschlossenen fachspezifischen Seminar

Level 8: Personen, die gleichzeitig die Kriterien von Level 9 und 10 erfüllen

Level 9: Personen mit Meta-Analyse als Diplomarbeitsthema

Level 10: Belegung des Fachliteraturseminars zum Thema Meta-Analysen, welches zum Zeitpunkt der Testung noch nicht abgeschlossen war.

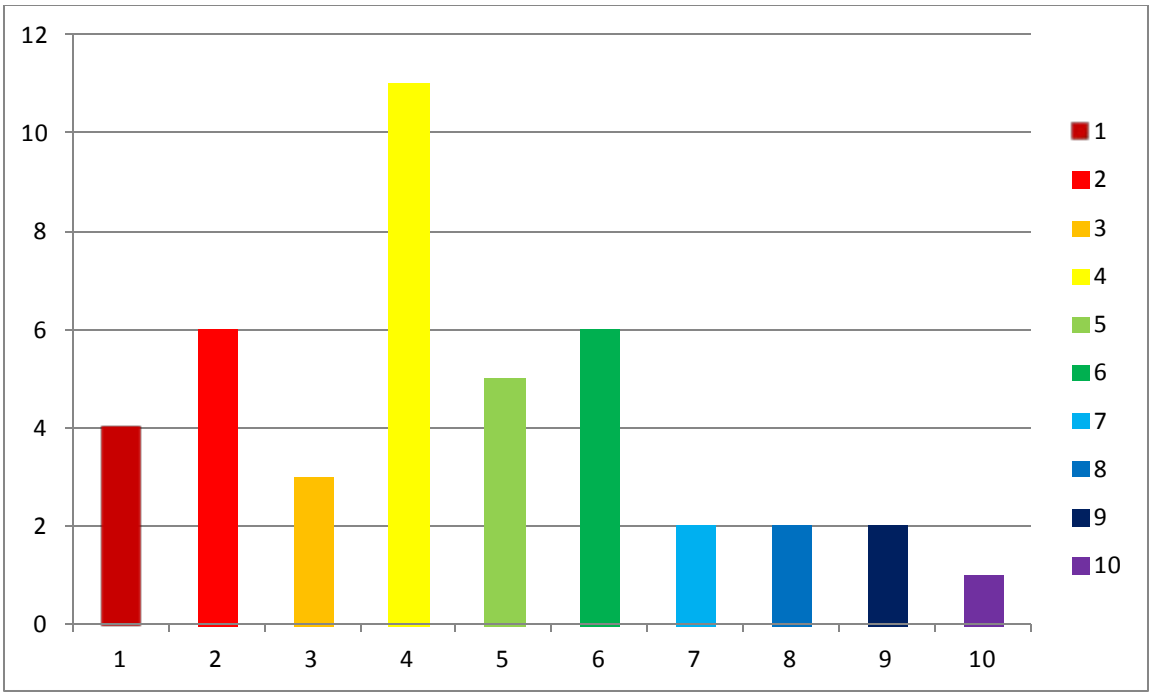

Abbildung 1: Verteilung der Teilnehmerlevel

2.1.4 Erhebungsmethode

Um ein angemessenes Verhältnis von Versuchspersonen und Versuchsleiter herzustellen, wurden maximal drei Teilnehmer gleichzeitig getestet. Nach Gediga et al. (2000) kann ein Versuchsleiter bis zu vier Teilnehmer betreuen.

Nach Ausfüllen des allgemeinen Fragebogenteils wurde den Teilnehmern ein PC, an dem die zu bearbeitenden Programme bereits geöffnet waren, zur Verfügung gestellt. Die Bearbeitungsreihenfolge der Aufgaben wurde den Teilnehmern freigestellt. Um bei Scheitern an der Dateneingabe fortfahren zu können, bestand die Möglichkeit, durch die Versuchsleiterin einen vorgefertigten Datensatz laden zu lassen.

2.2 Ergebnisse

2.2.1 Usability Assesment: Auswertung nach dem IsoMetrics-Manual

Die quantitative Auswertung der Testbögen erfolgte gemäß dem IsoMetrics-Manual. Dabei wurden vier Teilnehmer eliminiert, da sie mehr als 20% der Items nicht oder mit „Keine Angabe“ beantwortet haben.

Die Skalen Fehlerrobustheit und Individualisierbarkeit wurden ebenfalls von den weiteren Auswertungen ausgeschlossen, da ein Großteil der Items nicht oder mit „Keine Angabe“ beantwortet wurden. Laut Manual (Gediga et al., 2000) ist das Fortlassen ganzer Subskalen methodisch unproblematisch. So konnte außerdem verhindert werden, noch mehrere Teilnehmer auszuschließen. Außerdem wurden die Items A.14 und T.13 aus den selben Gründen eliminiert und die im Manual vorgesehene Korrektur angewandt.

Der theoretisch mögliche Wertebereich erstreckt sich von eins („stimmt nicht“) bis fünf („stimmt sehr“). Die Kenngrößen für die Gesamtbeurteilung der sechs Softwaresysteme werden in Tabelle 1 dargestellt:

Tabelle 1: Arithmetisches Mittel aller Programme über alle Skalen hinweg

Programm	$\bar{x}$	95% KI	*s*
CMA	3.86	[3.73; 4.01]	0.47
MetaAnalyst	3.95	[3.79; 4.11]	0.52
MetaWin	2.87	[2.72; 3.01]	0.45
MIX	3.64	[3.44; 3.85]	0.65
NCSS	3.51	[3.34; 3.68]	0.54
RevMan	3.18	[3.02; 3.34]	0.52

Die grafische Darstellung der Mittelwerte mit aufgesetzten Konfidenzintervallen ist in Abbildung 2 zu sehen.

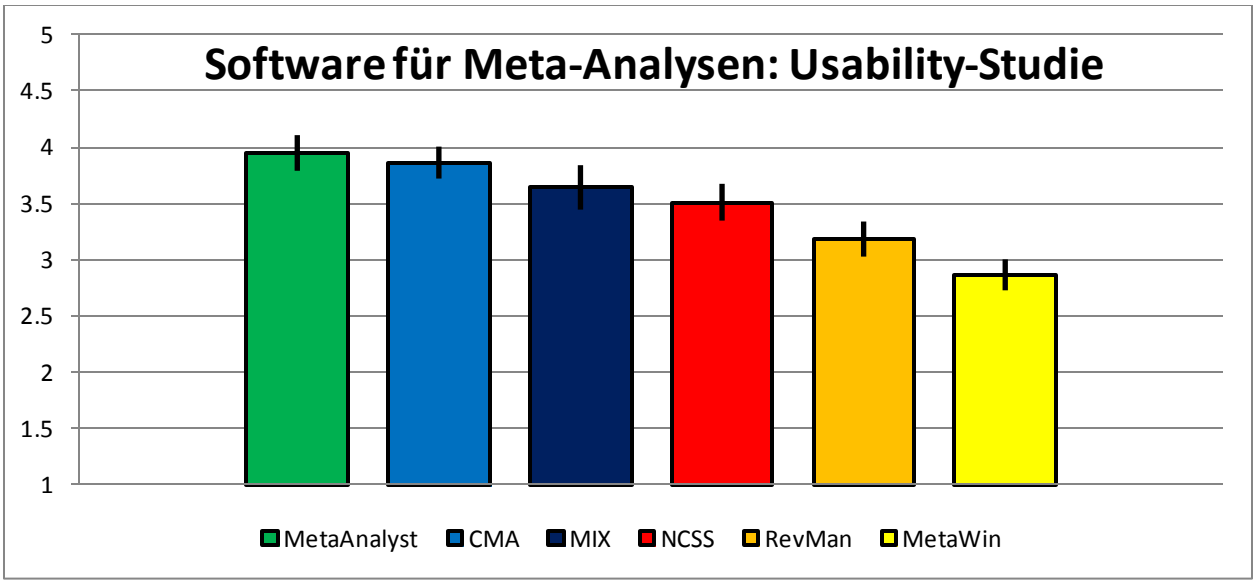

Abbildung 2: Mittelwerte mit Konfidenzintervall aller Programme

In Tabelle 2 bis 6 werden die Mittelwerte der einzelnen Skalen innerhalb der sechs Programme zusammen mit ihrem 95-prozentigen Konfidenzintervall und Standardabweichung dargestellt.

Tabelle 2: Aufgabenangemessenheit

Programm	$\bar{x}$	95% KI	*s*
CMA	3.99	[3.80; 4.18]	0.61
MetaAnalyst	4.12	[3.91; 4.33]	0.67
MetaWin	2.66	[2.48; 2.84]	0.58
MIX	3.62	[3.41; 3.84]	0.69
NCSS	3.24	[3.04; 3.43]	0.64
RevMan	3.01	[2.80; 3.22]	0.67

Tabelle 3: Selbstbeschreibungsfähigkeit

Programm	$\bar{x}$	95% KI	*s*
CMA	3.94	[3.77; 4.10]	0.52
MetaAnalyst	3.54	[3.31; 3.77]	0.74
MetaWin	2.84	[2.67; 3.02]	0.56
MIX	3.39	[3.13; 3.64]	0.81
NCSS	3.42	[3.23; 3.61]	0.61
RevMan	3.19	[3.04; 3.35]	0.49

Tabelle 4: Steuerbarkeit

Programm	$\bar{x}$	95% KI	*s*
CMA	3.66	[3.46; 3.85]	0.63
MetaAnalyst	3.85	[3.68; 4.02]	0.55
MetaWin	2.90	[2.73; 3.07]	0.54
MIX	3.57	[3.36; 3.78]	0.68
NCSS	3.58	[3.39; 3.77]	0.62
RevMan	3.20	[2.99; 3.42]	0.70

Tabelle 5: Erwartungskonformität

Programm	$\bar{x}$	95% KI	*s*
CMA	3.95	[3.77; 4.14]	0.59
MetaAnalyst	4.03	[3.87; 4.18]	0.49
MetaWin	3.12	[2.94; 3.29]	0.57
MIX	3.87	[3.76; 4.08]	0.66
NCSS	3.69	[3.49; 3.89]	0.64
RevMan	3.45	[3.24; 3.66]	0.68

Tabelle 6: Erlernbarkeit

Programm	$\bar{x}$	95% KI	*s*
CMA	3.87	[3.57; 3.98]	0.66
MetaAnalyst	4.20	[4.01; 4.39]	0.62
MetaWin	2.82	[2.58; 3.05]	0.74
MIX	3.76	[3.50; 4.02]	0.84
NCSS	3.63	[3.41; 3.85]	0.70
RevMan	3.05	[2.80; 3.32]	0.83

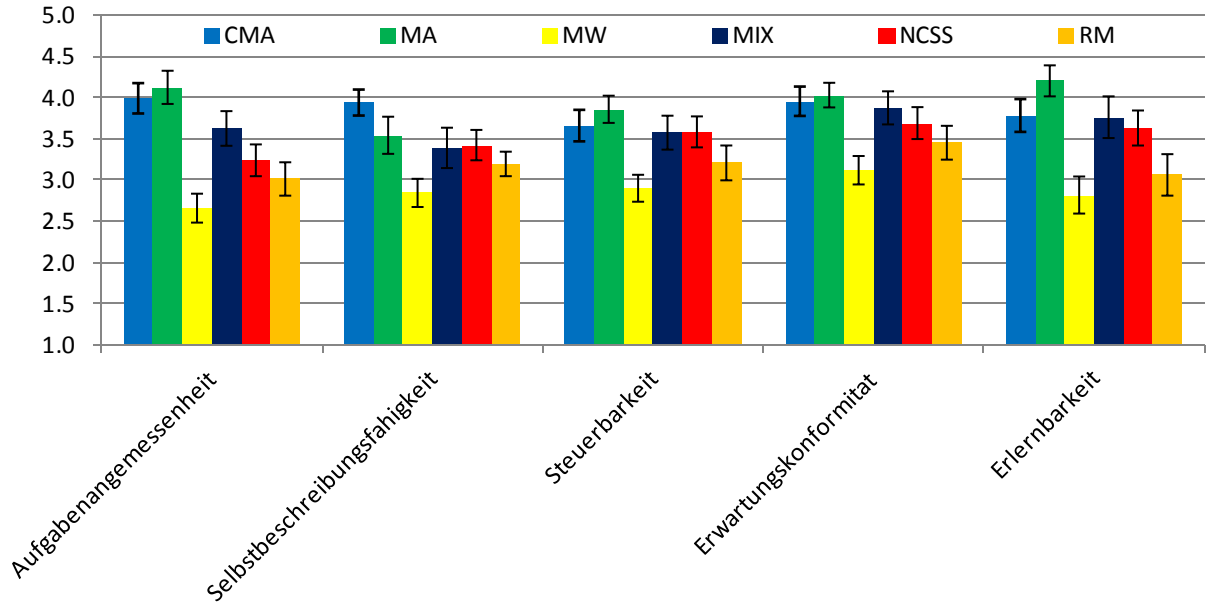

Abbildung 3: Mittelwerte mit Konfidenzintervall aller Skalen

Alle weiteren statistischen Analysen wurden mittels im Folgenden beschriebenen Verfahren in SPSS durchgeführt.

2.2.2 Weitere Berechnungen

Vergleich von Mittelwertsunterschieden

Zur Beantwortung der Fragestellung, ob es einen signifikanten Mittelwertsunterschied der Usability-Scores der getesteten Programme gibt, wurde nach Überprüfung der Voraussetzungen eine Varianzanalyse mit Messwiederholung gerechnet. Die Ergebnisse der Tests der Innersubjekteffekte zeigen, dass signifikante Unterschiede zwischen der Beurteilung der Usability der einzelnen Programme bestehen ($p < .01$).

Es wurden Bonferroni-korrigierte Post-Hoc-Tests verwendet, um Mittelwertsunterschiede herauszufinden. Im Folgenden werden einige ausgewählte Ergebnisse der Tests dargestellt. Alle weiteren Werte können Anhang 3 bis 8 entnommen werden.

Gesamtscore

Bezüglich des Gesamtscores unterscheidet sich MetaWin mit einem arithmetischen Mittel von 2.87 als einziges Programm von allen anderen und schneidet am schlechtesten ab. Den zweitniedrigsten Wert erzielte RevMan mit einem Mittelwert von 3.18. Dieses Programm unterscheidet sich signifikant von allen Programmen außer NCSS, das mit einem Mittelwert von 3.51 darüber liegt. Zwischen den drei Programmen, die am besten plaziert wurden nämlich MetaAnalyst ($\bar{x}$ = 3.95), CMA ($\bar{x}$ = 3.86) und MIX ($\bar{x}$ = 3.64) konnten keine signifikanten Unterschiede festgestellt werden.

Aufgabenangemessenheit

In der Skala Aufgabenangemessenheit unterscheidet sich Metaanalyst mit dem höchsten Mittelwert von 4.12 signifikant von allen Programmen außer CMA ($\bar{x}$ = 3.98). Auch hier schnitt MetaWin ($\bar{x}$ = 2.66) wieder am schlechtesten ab und erzielt nur im Vergleich mit RevMan ($\bar{x}$ = 3.01) kein signifikantes Ergebnis.

Selbstbeschreibungsfähigkeit

In der Skala Selbstbeschreibungsfähigkeit hebt sich das Programm CMA ($\bar{x}$ = 3.94) signifikant von den anderen Programmen ab. Wie schon beim Gesamtscores erzielte MetaWin ($\bar{x}$ = 2.84) signifikant das schlechteste Ergebnis.

Steuerbarkeit

Bezüglich der wahrgenommenen Steuerbarkeit der Programme ist es, wie schon in der Skala Aufgabenangemessenheit, erneut MetaWin, das sich mit einem arithmetischen Mittel von 2.90 signifikant von allen Programmen außer RevMan ($\bar{x}$ = 3.20) unterscheidet. MetaAnalyst ($\bar{x}$ = 3.85), CMA ($\bar{x}$ = 3.66), MIX ($\bar{x}$ = 3.57) und NCSS ($\bar{x}$ = 3.58) weisen untereinander keine signifikanten Unterschiede auf.

Erwartungskonformität

In der Skala Erwartungskonformität unterscheidet sich wiederum MetaWin ($\bar{x}$ = 3.12) signifikant von allen Programmen außer RevMan ($\bar{x}$ = 3.45), zwischen den anderen Programmen gibt es keine signifikanten Unterschiede.

Erlernbarkeit

In der Skala Erlernbarkeit erzielte MetaAnalyst mit einem arithmetischen Mittel von 4.20 die höchste Bewertung und unterscheidet sich von allem Programmen außer MIX ($\bar{x}$ = 3.76). Wie schon in den meisten anderen Skalen beobachtet werden konnte, wurde MetaWin ($\bar{x}$ = 2.82) am schlechtesten bewertet und nur RevMan ($\bar{x}$ = 3.06) unterscheidet sich nicht signifikant davon.

Einfluss von Prädiktorvariablen

Aufgrund der Tatsache, dass annähernd die Hälfte der Personen über Vorerfahrung mit dem Programm CMA verfügte, schien es sinnvoll, durch eine Robustheitsanalyse zu überprüfen, ob die Usability-Beurteilung dadurch verändert wird. Die Güte von Kompetenzlevel, Geschlecht und Vorerfahrung der Teilnehmer mit CMA als Prädiktor für die Beurteilung der Usability wurde mit einer multiplen Regressionsanalyse untersucht. Es konnte jedoch kein signifikanter Effekt festgestellt werden, was bedeutet, dass keiner der drei Faktoren Einfluss auf die Beurteilung genommen hat (korrigiertes *R*-Quadrat = 0.009). Die genauen Werte sind in Tabelle 8 zu sehen.

Tabelle 7: Multiple Regression

	Standardfehler	β	*p*
Level	0.036	-.178	.320
CMA-Erfahrung	0.162	.273	.127
Geschlecht	0.152	.128	.429

Reliabilitäten

Die Cronbach-α-Werte der einzelnen Skalen liegen im Bereich von .721 bis .922. Dies bedeutet, dass die Items innerhalb einer Skala sehr hoch miteinander korrelieren und damit die interne Konsistenz der Skalen sehr hoch ist. Die einzelnen Skalenwerte können Anhang 9 entnommen werden.

2.3. Diskussion

2.3.1 Interpretation der Ergebnisse

Ziel dieser Untersuchung war, sechs Programme, die sich sowohl bezüglich der Bedienung als auch der Kosten unterscheiden, hinsichtlich ihrer Usability zu bewerten und eine Vergleichbarkeit zwischen den Programmen und auch zur Studie von Bax et al. (2007) herzustellen. Zu diesem Zweck wurde der von Gediga und Hamborg (2000) erstellte und validierte Fragebogen IsoMetrics S verwendet. Dieser beruht auf den Grundsätzen zur Dialoggestaltung der Norm DIN EN ISO 9241-110, welche einen europäischen Konsensus zur Usability von Software verwirklicht.

Die Ergebnisse der Usability-Studie zeigen, dass sowohl bezüglich des Gesamtscores als auch hinsichtlich fünf von sechs Skalen das Programm MetaAnalyst am besten abschneidet. Es hebt sich allerdings nicht signifikant von allen anderen Programmen ab. MetaAnalyst kann von der Website des Autors heruntergeladen werden und liegt zum momentanen Zeitpunkt nur in einer Beta-Version vor. Bemerkenswert ist, dass es neben RevMan das einzige kostenlose Programm ist. Ein Grund für die hohe Bewertung mag sein, dass nach Durchführung der Analyse sofort alle verfügbaren Ergebnisse vorliegen, ohne diese einzeln anwählen zu müssen. So besteht der Output immer aus den eingegebenen Studiendaten, deskriptiver Statistik, dem Ergebnis der Meta-Analyse und einem Forest und Funnel Plot. Unten im Bildschirm wird automatisch ein Fenster mit Tipps und Hilfestellungen eingeblendet, was gerade bei erstmaliger Anwendung sehr hilfreich sein kann.

CMA liegt im Gesamtscore und auch in den meisten Dimensionen knapp unter MetaAnalyst. Es ist wohl das Programm mit den umfassendsten Funktionen innerhalb

der Studie, gleichzeitig aber auch das teuerste. Gerade in der Dimension Selbstbeschreibungsfähigkeit erzielte es den höchsten Wert, was daran liegen könnte, dass der Benutzer mit einem automatisch eingeblendeten Tutorial durch die ersten Schritte geleitet wird. So werden Dateneingabe, Erzeugung notwendiger Spalten und die Berechnung der Effektstärken genau erklärt. Trotzdem wurde deutlich, dass auch bei diesem Programm noch Raum zur Verbesserung der Usability besteht, da in keiner der fünf Dimensionen herausragende Werte erzielt werden konnten.

Die frühere Arbeit von Bax et al. (2007) kam zu dem Ergebnis, dass MIX, welches von Mitgliedern derselben Autorengruppe entwickelt wurde, die höchste Usability aufweist. Dieses Ergebnis konnte in vorliegender Studie nicht repliziert werden. In der Gesamtbeurteilung erzielte das Programm den dritten Rang und hob sich in keiner Skala besonders hervor. Auch das den meisten Teilnehmern der Studie wohl bekannte Excel-Interface konnte zu keinem besseren Ergebnis führen. Allerdings verfügt die Software, wie auch CMA, über ein sehr breites Spektrum an Funktionen.

NCSS ist ein allgemeines Statistikprogramm und deshalb nicht spezialisiert auf die Durchführung von Meta-Analysen. So verfügt es nicht über alle abgefragten Funktionen, was die Beurteilung durchaus beeinflussen mag. Insgesamt erreichte das Programm aber trotzdem in allen Skalen einen mittleren Wert, was mitunter an der ansprechenden Gestaltung liegen könnte.

RevMan erleichtert zwar die Durchführung eines systematischen Reviews, weil der Benutzer quasi durch den ganzen Prozess geleitet wird, wenn allerdings nur meta-analytische Berechnungen durchgeführt werden sollen, ist die geforderte Eingabe bestimmter Studiendaten überflüssig und vereinnahmt relativ viel Zeit. Außerdem verfügt auch dieses Programm nicht über alle gängigen Funktionen, was wohl zusätzlich zu der relativ schlechten Beurteilung beiträgt.

MetaWin erzielte die schlechtesten Bewertungen im Test. Auch dieses Programm erfordert viele unnötige Schritte vom Anwender. Schon bei der Eingabe von Studiendaten müssen die Spalten einzeln in ein numerisches Format umgewandelt werden, da sonst keine Analysen durchgeführt werden können. Anschließend wird die Berechnung der Effektstärken vorausgesetzt, worauf der Benutzer allerdings nicht explizit hingewiesen wird. Bei jeder weiteren Berechnung müssen die Daten, die als

Grundlage dienen, erneut angewählt werden, was außerdem anfangs nicht sofort ersichtlich sein mag. Um wieder zur Datenansicht zurückzukommen, nachdem der Output präsentiert wurde, muss das Output-Fenster minimiert werden. Insgesamt scheint es, als hätten die Entwickler dieses Programms die Grundregeln der Usability nicht berücksichtigt.

Überraschend ist, dass weder Kompetenzlevel noch Vorerfahrung der Teilnehmer einen signifikanten Einfluss auf die Beurteilung der Usability nehmen. Es wäre zu erwarten, dass vor allem die Vorerfahrung mit dem Programm CMA die Beurteilung des selben beeinflusst. Die Ergebnisse zeigen aber, dass die wiederholte Nutzung eines Programmes nicht unbedingt eine veränderte Beurteilung impliziert.

2.3.2 Probleme früherer Arbeiten

Frühere Reviews oder Usability-Studien (Franke, 1992; Normand, 1995; Borenstein, 2005; Bax, Yu, Ikeda & Moons, 2007) haben sich zwar bemüht, einen Vergleich zwischen mehreren Programmen herzustellen und Komponenten der Usability abzufragen, allerdings ohne den Einsatz eines validierten Messverfahrens. Viele der vorhandenen Reviews oder Kapitel über Software für Meta-Analysen sind von Entwicklern verschiedener Programme mitverfasst und implizieren daher meist schon einen gewissen Bias, sowohl in der Beurteilung als auch in der Vorstellung der Software. Wie schon erwähnt, schnitt MIX in der Usability-Studie, die von Bax et al. (2007) publiziert wurde, am besten ab. Auch in einem Buchkapitel von Borenstein (2009) werden lediglich CMA, RevMan und Stata angeführt. Hierzu ist zu erwähnen, dass Borenstein und einige der Herausgeber an der Entwicklung von CMA mitwirkten, andere auch zur Entwicklung von Algorithmen für die anderen beiden Programme beigetragen haben.

Im aktuellsten publizierten Meta-Analysen-Software-Review (Bax et al., 2007) werden nur Windows-Programme erwähnt. Auch Software mit kommandogesteuertem Interface wurden gänzlich vernachlässigt. Diese Arbeit bietet einen Überblick über jegliche Art von Software zur Berechnung von Meta-Analysen.

2.3.3 Einschränkungen

Ein Kritikpunkt dieser Studie ist die Bearbeitungszeit, die den Teilnehmern zur Beurteilung der Programme zur Verfügung stand; diese betrug im Durchschnitt insgesamt 90 bis 120 Minuten. Um sich ausgiebig mit einem Programm zu befassen und zu evaluieren, ob die individuellen Bedürfnisse damit zur Genüge abgedeckt werden, bedarf es sicher mehrerer Stunden intensiver Beschäftigung. Deshalb wurden auch nur wenige Aufgaben vorgegeben, von denen allerdings angenommen wird, dass sie für die meisten Anwender grundlegende Funktionen darstellen. Allerdings war dies auch nicht das Hauptziel der Erhebung, sondern durch die Evaluation sollte vor allem die Usability der Programme erhoben werden.

DOS-Programme bzw. Programme, die nicht menügesteuert sind, wurden von der Usability-Studie ausgeschlossen, da sie sich aufgrund zumeist grundverschiedener Anwendergruppen schlecht mit grafisch gesteuerten vergleichen lassen. Für Programmierer oder Experten im Umgang mit befehlsgesteuerter Software können diese Programme sehr einfach zu bedienen und gegebenenfalls auch zu modifizieren sein, für einen ungeübten Nutzer ist die Handhabung teilweise unmöglich ohne zuvor aufwändig eine Programmiersprache erlernt zu haben. Viele der DOS-Programme werden heutzutage nicht mehr unterstützt bzw. gewartet, bei mit der Software auftretenden Problemen kann dann kein Support kontaktiert werden.

Menügesteuerte Programme mit einer grafischen Oberfläche dagegen sind, da sie ihre Funktionen zum Anschauen und Ausprobieren zur Verfügung stellen, meist einfacher zu bedienen und deshalb wahrscheinlich beim Großteil der Nutzer beliebter.

2.3.4 Zusammenfassung und Ausblick

Der Review und die Usability-Studie hat gezeigt, dass ein Großteil der Anforderungen von vielen Programmen gleich gut erfüllt werden kann. Es hat sich herausgestellt, dass kostenpflichtige Programme nicht grundsätzlich besser abschneiden als frei verfügbare, wie auch die guten Testergebnisse von MetaAnalyst zeigen.

Die Individualisierbarkeit fast aller Programme, vor allem der mit grafischer Benutzeroberfläche, bietet durchaus noch Verbesserungspotential, beispielsweise durch die Möglichkeit, eigene Module in vorhandene Programme zu integrieren. Dies ist bislang nur bei befehlsgesteuerten Programmen möglich. Auch eine bessere Austauschbarkeit der Daten zwischen Programmen wäre wünschenswert.

Es gibt Software, die nur in Sparten der Forschung wie Epidemiologie und Diagnostik verwendbar ist. Da große Programme aber auch diese Anforderungen abdecken können, ist zu erwarten, dass kleinere Programme speziell für diese Zwecke sich nicht am Markt halten können, vor allem, wenn sie kostenpflichtig sind.

Zu erwarten wäre, dass MIX oder ein anderer großer Hersteller für Meta-Analysen-Software in der Zukunft anbietet, Daten aus Primärstudien in gemischten Formaten einzugeben, da dies bislang nur in CMA möglich ist und für viele Anwender Vorteile bietet, da eine aufwändige Vorbereitung der Daten entfällt.

Programme wie CMA und MIX bieten schon jetzt eine Vielzahl von Funktionen, es ist aber gleichzeitig wichtig, dass verbesserte oder neue wichtige Rechenverfahren regelmäßig in die Software integriert werden. Beispielsweise ist die Berechnung einer Metaregression mit mehreren Prädiktorvariablen bislang nur mit sehr wenigen Programmen möglich, würde aber den realen Anforderungen durchaus entsprechen. Ebenso müssen die Programme dem Umgang mit immer größeren Datenmengen gewachsen bleiben. Für Publikationen sind hochauflösende Grafiken wichtig; in diese Richtung ist bereits zum jetzigen Zeitpunkt ein deutlicher Trend erkennbar.

Da Meta-Analysen zu den fortschrittlichsten und aussagekräftigsten Forschungsmethoden gezählt werden können, wäre zu erwarten, dass auch allgemeine Statistikprogramme, wie bei NCSS im Ansatz zu erkennen ist, mehr Möglichkeiten für meta-analytische Berechnungen in ihr Repertoire aufnehmen.

Aufgrund der raschen Weiterentwicklung im Bereich Software ist es sinnvoll, ähnliche Arbeiten in einem Abstand von wenigen Jahren erneut durchzuführen, da bereits in kurzer Zeit wieder Neuerungen bei den verfügbaren Programmen, aber auch weitere Fortschritte im Bereich der Usability zu erwarten sind. Interessant wäre auch, mit der entsprechenden Anwendergruppe einen Vergleich von befehlsgesteuerten

Programmen durchzuführen, da so vor allem der Aspekt der Individualisierbarkeit, der bei vorliegender Studie ausgeschlossen wurde, genauer betrachtet werden könnte.

3. Schlusswort

Diese Arbeit soll einen guten Überblick über das vorhandene Angebot an Software für Meta-Analysen geben. Die Entscheidung, welches Programm tatsächlich verwendet wird, wird allerdings sowohl von Verfügbarkeit und Kosten, aber auch von persönlichen Präferenzen und dem eigenen Wissensstand abhängen.

4. Abbildungsverzeichnis

Abbildung 1: Verteilung der Teilnehmerlevel
Abbildung 2: Mittelwerte mit Konfidenzintervall aller Programme
Abbildung 3: Mittelwerte mit Konfidenzintervall aller Skalen

5. Tabellenverzeichnis

Tabelle 1: Arithmetisches Mittel aller Programme über alle Skalen hinweg
Tabelle 2: Programmvergleich Aufgabenangemessenheit
Tabelle 3: Programmvergleich Selbstbeschreibungsfähigkeit
Tabelle 4: Programmvergleich Steuerbarkeit
Tabelle 5: Programmvergleich Erwartungskonformität
Tabelle 6: Programmvergleich Erlernbarkeit
Tabelle 7: Multiple Regression

6. Anhänge

Anhang 1: Programminformationen
Anhang 2: Programme Usability-Studie
Anhang 3: Paarweise Vergleiche Gesamtscore
Anhang 4: Paarweise Vergleiche Aufgabenangemessenheit
Anhang 5: Paarweise Vergleiche Selbstbeschreibungsfähigkeit
Anhang 6: Paarweise Vergleiche Steuerbarkeit
Anhang 7: Paarweise Vergleiche Erwartungskonformität
Anhang 8: Paarweise Vergleiche Erlernbarkeit
Anhang 9: Reliabilitäten der einzelnen Skalen

Anhang 1: Programminformationen

Programme	Aktuelle Version	Letzte Aktualisierung	Verfügbarkeit	Bezugsquelle	Systemvoraussetzung
Advanced BASIC Meta-Analysis (ABA)	1.10	1989	Buchbeilage	ISBN-13: 978-0805805024	DOS
BASIC Meta-Analysis		1985	Buchbeilage	ISBN-13: 978-0898596199	DOS
Comprehensive Meta-analysis (CMA)	2.2.064	2011-Aug	kostenpflichtig	www.meta-analysis.com	WINDOWS
DSTAT	.120	1997	auf Anfrage	http://meta-analysis.ning.com/group/softwaretocreategraphicsformetaanalysis/forum/topics/dstat-1x-legacy-version-1	DOS
DSTAT 2.x	Build #0166	2012	auf Anfrage	http://meta-analysis.ning.com/group/softwaretocreategraphicsformetaanalysis/forum/topics/dstat-2x-contemporary-version	Java
EasyMA	2001	2001-Sep	kostenlos	http://www.spc.univ-lyon1.fr/easyma.dos/	DOS
EpiMeta	1.2	1997-Apr	kostenlos	http://ftp.cdc.gov/pub/Software/epimeta/	DOS
Fast*Pro	1.7	1992	Buchbeilage	ISBN-13: 978-0122306211	DOS
HEpiMA	2.1.3	2000	kostenlos, auf Anfrage		WINDOWS
Hunter Schmidt MA Program Package		1993	Buchbeilage, auf Anfrage	ISBN-13: 978-1412904797	WINWOWS
Meta-Analysis	5.3	1993	kostenlos	http://userpage.fu-berlin.de/~health/meta_e.htm	DOS
Meta-Analysis Easy To Answer (META)	III	2003-Feb	kostenlos	http://davidakenny.net/meta.htm	DOS, (WINDOWS)
Meta-Analyst	ß 3.13	2009-Jun	kostenlos	http://tuftscaes.org/meta_analyst/	WINDOWS
Meta-DiSc	1.4	2006-Jun	kostenlos	ftp://ftp.hrc.es/pub/programas/metadisc/Metadisc_update.htm	WINDOWS
MetaDos		1996	kostenlos, auf Anfrage		DOS
MetAnalysis	1.2.0	2004-Mär	Buchbeilage	ISBN-13: 978-1405127332	WINDOWS
Meta-Stat	1.5	2002	kostenlos	http://ericae.net/meta/metastat.htm, http://echo.edres.org:8080/meta/metastat.htm	DOS
Meta-Test	0.6	1997-Apr	auf Anfrage		DOS
MetaWin	2.1.5.10	2007-Feb	kostenpflichtig	http://www.metawinsoft.com/	WINDOWS
MIX	2.0.1.4	2011-Jul	kostenpflichtig	http://www.meta-analysis-made-easy.com/	MS Excel 2007
NCSS	07.1.21	2011-Jun	kostenpflichtig	http://www.ncss.com/ncss.html	WINDOWS
R - Metafor Package	1.6-0	2011-Apr	kostenlos	http://www.metafor-project.org/	R
Review Manager (RevMan)	5.1.6	2011-Dez	kostenlos	http://ims.cochrane.org/revman	Java
SAS (Statistical Analysis System)	9.3	2011-Dez	kostenpflichtig	http://www.sas.com/software/sas9/	WINDOWS, UNIX,Linux
S+	8.2	2010-Nov	kostenpflichtig	http://spotfire.tibco.com/products/s-plus/statistical-analysis-software.aspx	WINDOWS, UNIX, Linux
Software for Permutation Methods	n.a.	2011-Nov	kostenlos	http://www.stat.colostate.edu/~mielke/permute.html	FORTRAN-77
SPSS - Metaanalyse (Rustenbach)	n.a.	2003	Buchbeilage	ISBN-13: 978-3456838021	SPSS
Stata	12	2011-Nov	kostenpflichtig	http://www.stata.com/stata12/	WINDOWS, Mac OS X, UNIX
Synthesizer	1.0	2009	kostenlos	http://www.psychology.iastate.edu/~zkrizan/Synthesizer	MS Excel 2007
TRUE EPISTAT	5.3	1997	n.a.		DOS
WEasyMA	2.5	2002	auf Anfrage	http://www.clininfo.fr/en/weasyma.html	WINDOWS

Anhang 2: Programme Usability-Studie

Name des Programms	CMA	MetaAnalyst	MetaWin	MIX	NCSS	RevMan
			Allgemeine Informationen			
Systemvoraussetzungen	WINDOWS	WINDOWS	WINDOWS	WINDOWS	WINDOWS	WINDOWS, MacOS, Linux
Aktuelle Version	2.2.057	Beta 3.13	2.1.5.10	2.0.1.2 Pro	07.1.21	5.1.1
Letzte Aktualisierung	2010	2009	2007	2011	2011	2011
Verfügbarkeit	kostenpflichtig	kostenlos	kostenpflichtig	kostenpflichtig	kostenpflichtig	kostenlos
Einzelplatzlizenz	195-1295	n.a.	60$	210$	795$	n.a.
Einzelplatzlizenz für Studenten	95-795	n.a.	n.a.	75$	495$	n.a.
Download- / Programmgröße	22MB/30MB	8MB/19MB	0.2MB/4MB	3MB/4MB	71MB/100MB	29MB/25MB
			Eingabeoptionen			
Manuelle Dateneingabe	x	x	x	x	x	x
Copy & Paste	x	x	x	x	x	x
Dichotome Daten	x	x	x	x	x	x
Kontinuierliche Daten	x	x	x	x	x	x
Eingabe gemischter Formate	x					
			Hilfe-/ Informationsmöglichkeiten			
Manual	x	x	x		x	x
Beschreibung der Methoden		x	x	x	x	x
Hilfe innerhalb des Programms	x	x	x	x	x	x
Gewartete Internetpräsenz	x	x	x	x	x	x
			Exportmöglichkeiten			
Output in Zwischenablage kopieren	x	x	x	x	x	x
Erstellen eines Reports	x	x	x	x	x	x
Output (als pdf) drucken	x	x	x	x	x	x
			Numerischer Output			
Kontinuierliche Daten	x	MD, HG, CD, andere	HG, andere	MD, HG, CD	MD	MD, HG
Dichotome Daten	x	OR, RR, RD	OR, RR, RD	RD, RR, OR	RD, RR, OR	RD, RR, OR
Andere Datentypen	x		FZ	CC, FZ		
Fixed-Effect-Modell	IV, MH, PETO	IV, MH, PETO	IV, MH, PETO	IV, MH, PETO	IV	IV, MH, PETO
Random-Effects-Modell	DL	DL	DL	DL, PM	DL	DL
Heterogenitätsanalysen	x		x	x	x	x
Publikationsbias	x		FSN, RC	FSN, RK, EGG, MAC, TF		
Kumulative Meta-Analyse	x	x	x	x		
Metaregression	x	x	x			
			Grafischer Output			
Forest Plot	x	x	x	x	x	x
- Punkte proportional zur Gewichtung	x	x		x	x	x
- Kumulative Darstellung	x	x	x	x		
Funnel Plot	x	x	x	x		x
Galbraith / Radial Plot			x	x	x	
"Trim and Fill"-Plot	x			x		
L'Abbe Plot		x	x	x	x	
Andere Plots	ScatterPlot		HIST, NQ, ScatterPlot	BOX, HIST, NQ, andere		SROC

Anhang 3: Paarweise Vergleiche Gesamtscore

Gesamtscore (I)	Gesamtscore (J)	Mittlere Differenz (I-J)	95% Konfidenzintervall für die Differenz	
			Untergrenze	Obergrenze
CMA	MetaAnalyst	-0.085	-0.330	0.159
	MetaWin	0.997*	0.736	1.258
	MIX	0.220	-0.057	0.498
	NCSS	0.352*	0.080	0.624
	RevMan	0.680*	0.364	0.995
MetaAnalyst	CMA	0.085	-0.159	0.330
	MetaWin	1.083*	0.778	1.388
	MIX	0.306	-0.045	0.657
	NCSS	0.438*	0.115	0.761
	RevMan	0.765*	0.400	1.130
MetaWin	CMA	-0.997*	-1.258	-0.736
	MetaAnalyst	-1.083*	-1.388	-0.778
	MIX	-0.777*	-1.126	-0.428
	NCSS	-0.645*	-0.923	-0.368
	RevMan	-0.318*	-0.621	-0.015
MIX	CMA	-0.220	-0.498	0.057
	MetaAnalyst	-0.306	-0.657	0.045
	MetaWin	0.777*	0.428	1.126
	NCSS	0.132	-0.205	0.468
	RevMan	0.459*	0.086	0.833
NCSS	CMA	-0.352*	-0.624	-0.080
	MetaAnalyst	-0.438*	-0.761	-0.115
	MetaWin	0.645*	0.368	0.923
	MIX	-0.132	-0.468	0.205
	RevMan	0.327	-0.007	0.662
RevMan	CMA	-0.680*	-0.995	-0.364
	MetaAnalyst	-0.765*	-1.130	-0.400
	MetaWin	0.318*	0.015	0.621
	MIX	-0.459*	-0.833	-0.086
	NCSS	-0.327	-0.662	0.007

*. Die mittlere Differenz ist auf dem .05-Niveau signifikant.

Anhang 4: Paarweise Vergleiche Aufgabenangemessenheit

Aufgabenange-messenheit (I)	Aufgabenange-messenheit (J)	Mittlere Differenz (I-J)	95% Konfidenzintervall für die Differenz	
			Untergrenze	Obergrenze
CMA	MetaAnalyst	-0.133	-0.520	0.255
	MetaWin	1.332*	0.970	1.694
	MIX	0.366	-0.005	0.736
	NCSS	0.752*	0.403	1.101
	RevMan	0.978*	0.587	1.369
MetaAnalyst	CMA	0.133	-0.255	0.520
	MetaWin	1.464*	1.059	1.870
	MIX	0.498*	0.082	0.914
	NCSS	0.884*	0.499	1.270
	RevMan	1.111*	0.662	1.559
MetaWin	CMA	-1.332*	-1.694	-0.970
	MetaAnalyst	-1.464*	-1.870	-1.059
	MIX	-0.966*	-1.331	-0.601
	NCSS	-0.580*	-0.925	-0.235
	RevMan	-0.354	-0.748	0.040
MIX	CMA	-0.366	-0.736	0.005
	MetaAnalyst	-0.498*	-0.914	-0.082
	MetaWin	0.966*	0.601	1.331
	NCSS	0.386	-0.022	0.794
	RevMan	0.612*	0.190	1.035
NCSS	CMA	-0.752*	-1.101	-0.403
	MetaAnalyst	-0.884*	-1.270	-0.499
	MetaWin	0.580*	0.235	0.925
	MIX	-.0386	-0.794	0.022
	RevMan	0.226	-0.164	0.616
RevMan	CMA	-0.978*	-1.369	-0.587
	MetaAnalyst	-1.111*	-1.559	-0.662
	MetaWin	0.354	-0.040	0.748
	MIX	-0.612*	-1.035	-0.190
	NCSS	-0.226	-0.616	0.164

*. Die mittlere Differenz ist auf dem .05-Niveau signifikant.

Anhang 5: Paarweise Vergleiche Selbstbeschreibungsfähigkeit

Selbstbeschrei-bungsfähigkeit (I)	Selbstbeschrei-bungsfähigkeit (J)	Mittlere Differenz (I-J)	95% Konfidenzintervall für die Differenz	
			Untergrenze	Obergrenze
CMA	MetaAnalyst	0.397*	0.067	0.726
	MetaWin	1.095*	0.715	1.475
	MIX	0.550*	0.184	0.915
	NCSS	0.516*	0.169	0.863
	RevMan	0.744*	0.437	1.051
MetaAnalyst	CMA	-0,397*	-0.726	-0.067
	MetaWin	0.698*	0.270	1.127
	MIX	0.153	-0.249	0.555
	NCSS	0.119	-0.295	0.533
	RevMan	0.347	-0.051	0.746
MetaWin	CMA	-1.095*	-1.475	-0.715
	MetaAnalyst	-0.698*	-1.127	-0.270
	MIX	-0.546*	-1.004	-0.087
	NCSS	-0.579*	-0.918	-0.241
	RevMan	-0.351*	-0.616	-0.086
MIX	CMA	-0.550*	-0.915	-0.184
	MetaAnalyst	-0.153	-0.555	0.249
	MetaWin	0.546*	0.087	1.004
	NCSS	-0.034	-0.410	0.343
	RevMan	0.194	-0.151	0.540
NCSS	CMA	-0.516*	-0.863	-0.169
	MetaAnalyst	-0.119	-0.533	0.295
	MetaWin	0.579*	0.241	0.918
	MIX	0.034	-0.343	0.410
	RevMan	0.228	-0.050	0.506
RevMan	CMA	-0.744*	-1.051	-0.437
	MetaAnalyst	-0.347	-0.746	0.051
	MetaWin	0.351*	0.086	0.616
	MIX	-0.194	-0.540	0.151
	NCSS	-0.228	-0.506	0.050

*. Die mittlere Differenz ist auf dem .05-Niveau signifikant.

Anhang 6: Paarweise Vergleiche Steuerbarkeit

Steuerbarkeit (I)	Steuerbarkeit (J)	Mittlere Differenz (I-J)	95% Konfidenzintervall für die Differenz	
			Untergrenze	Obergrenze
CMA	MetaAnalyst	-0.198	-0.532	0.137
	MetaWin	0.760*	0.392	1.128
	MIX	0.086	-0.180	0.352
	NCSS	0.076	-0.277	0.429
	RevMan	0.452*	0.056	0.849
MetaAnalyst	CMA	0.198	-0.137	0.532
	MetaWin	0.957*	0.609	1.305
	MIX	0.283	-0.072	0.639
	NCSS	0.274	-0.112	0.659
	RevMan	0.650*	0.263	1.037
MetaWin	CMA	-0.760*	-1.128	-0.392
	MetaAnalyst	-0.957*	-1.305	-0.609
	MIX	-0.674*	-1.073	-0.274
	NCSS	-0.683*	-1.032	-0.334
	RevMan	-0.307	-0.676	0.062
MIX	CMA	-0.086	-0.352	0.180
	MetaAnalyst	-0.283	-0.639	0.072
	MetaWin	0.674*	0.274	1.073
	NCSS	-0.010	-0.392	0.373
	RevMan	0.367	-0.074	0.808
NCSS	CMA	-0.76	-0.429	0.277
	MetaAnalyst	-0.274	-0.659	0.112
	MetaWin	0.683*	0.334	1.032
	MIX	0.010	-0.373	0.392
	RevMan	0.376	-0.084	0.837
RevMan	CMA	-0.452*	-0.849	-0.056
	MetaAnalyst	-0.650*	-1.037	-0.263
	MetaWin	0.307	-0.062	0.676
	MIX	-0.367	-0.808	0.074
	NCSS	-0.376	-0.837	0.084

*. Die mittlere Differenz ist auf dem .05-Niveau signifikant.

Anhang 7: Paarweise Vergleiche Erwartungskonformität

Erwartungs-konformität (I)	Erwartungs-konformität (J)	Mittlere Differenz (I-J)	95% Konfidenzintervall für die Differenz	
			Untergrenze	Obergrenze
CMA	MetaAnalyst	-0.074	-0.344	0.195
	MetaWin	0.836*	0.499	1.174
	MIX	0.080	-0.229	0.389
	NCSS	0.265	-0.102	0.631
	RevMan	0.503*	0.097	0.909
MetaAnalyst	CMA	0.074	-0.195	0.344
	MetaWin	0.911*	0.614	1.207
	MIX	0.155	-0.206	0.515
	NCSS	0.339	-0.001	0.680
	RevMan	0.577*	0.170	0.985
MetaWin	CMA	-0.836*	-1.174	-0.499
	MetaAnalyst	-0.911*	-1.207	-0.614
	MIX	-0.756*	-1.166	-0.346
	NCSS	-0.571*	-0.934	-0.209
	RevMan	-0.333	-0.730	0.064
MIX	CMA	-0.080	-0.389	0.229
	MetaAnalyst	-0.155	-0.515	0.209
	MetaWin	0.756*	0.346	1.166
	NCSS	0.185	-0.149	0.518
	RevMan	0.423*	0.017	0.828
NCSS	CMA	-0.265	-0.631	0.102
	MetaAnalyst	-0.339	-0.680	0.001
	MetaWin	0.571*	0.209	0.934
	MIX	-0.185	-0.518	0.149
	RevMan	0.238	-0.189	0.665
RevMan	CMA	-0.503*	-0.909	-0.097
	MetaAnalyst	-0.577*	-0.985	-0.170
	MetaWin	0.333	-0.064	0.730
	MIX	-0.423*	-0.828	-0.017
	NCSS	-0.238	-0.665	0.189

*. Die mittlere Differenz ist auf dem .05-Niveau signifikant.

Anhang 8: Paarweise Vergleiche Erlernbarkeit

Erlernbarkeit (I)	Erlernbarkeit (J)	Mittlere Differenz (I-J)	95% Konfidenzintervall für die Differenz	
			Untergrenze	Obergrenze
CMA	MetaAnalyst	-0.420*	-0.769	-0.070
	MetaWin	0.964*	0.564	1.364
	MIX	0.021	-0.432	0.473
	NCSS	0.152	-0.190	0.494
	RevMan	0.720*	0.190	1.250
MetaAnalyst	CMA	0.420*	0.070	0.769
	MetaWin	1.384*	0.949	1.819
	MIX	0.440	-0.069	0.950
	NCSS	0.571*	0.168	0.975
	RevMan	1.140*	0.634	1.646
MetaWin	CMA	-0.964*	-1.364	-0.564
	MetaAnalyst	-1.384*	-1.819	-0.949
	MIX	-0.943*	-1.453	-0.434
	NCSS	-0.813*	-1.213	-0.412
	RevMan	-0.244	-0.813	0.325
MIX	CMA	-0.021	-0.473	0.432
	MetaAnalyst	-0.440	-0.950	0.069
	MetaWin	0.943*	0.434	1.453
	NCSS	0.131	-0.368	0.630
	RevMan	0.699*	0.122	1.277
NCSS	CMA	-0.152	-0.494	0.190
	MetaAnalyst	-0.571*	-0.975	-0.168
	MetaWin	0.813*	0.412	1.213
	MIX	-0.131	-0.630	0.368
	RevMan	0.568*	0.025	1.112
RevMan	CMA	-0.720*	-1.250	-0.190
	MetaAnalyst	-1.140*	-1.646	-0.634
	MetaWin	0.244	-0.325	0.813
	MIX	-0.699*	-1.277	-0.122
	NCSS	-0.568*	-1.112	-0.025

*. Die mittlere Differenz ist auf dem .05-Niveau signifikant.

Anhang 9: Reliabilitäten der einzelnen Skalen

Programm	Aufgaben-angemessenheit	Selbstbeschreibungs-fähigkeit	Steuerbar-keit	Erwartungs-konformität	Erlernbar-keit
CMA	.894	.805	.890	.821	.808
MetaAnalyst	.922	.905	.842	.748	.844
MetaWin	.837	.798	.814	.771	.851
MIX	.910	.919	.878	.874	.903
NCSS	.885	.849	.882	.831	.865
RevMan	.866	.721	.900	.834	.865

Literaturverzeichnis

Bax, L., Yu, L.-M., Ikeda, N., & Moons, G. K. (2007). A systematic comparison of software dedicated to meta-analysis of causal studies. *BMC Medical Research Methodology*, 7, 40. doi:10.1186/1471-2288-7-40

Begg, C. B., & Berlin, J. A. (1988). Publication bias: A problem in interpreting medical data. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 151*, 419-463.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*, 1088-1101.

Binder, T. (2006). Usability-Evaluierung von Universitätswebsites. *Unpublizierte Diplomarbeit, Universität Wien*.

Borenstein, M. (2005). Software for publication bias. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis* (pp. 163-219). Chichester, West Sussex: Wiley.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). Chapter 44: Software. In M. Borenstein, L. V. Hedges, J. P. Higgins, & H. Rothstein, *Introduction to meta-analysis* (pp. 391-403). Chichester: Wiley.

Clintools. (2008, March 23). *Effect Size Generator*. Retrieved January 30, 2012, from http://www.clintools.com/products/esg/effect_size_generator.html

Cooper, H. M., & Rosenthal, R. (1980). Statistical versus traditional procedures for summarizing research findings. *Psychological Bulletin, 87*, 442-449.

Cooper, H. R. (1979). Statistically combining independent studies: A meta-analysis of sex differences in conformity research. *Journal of Personality and Social Psychology, 37*, 131-146.

Costa-Bouzas, J., Takkouche, B., Cadarso-Suarez, C., & Spiegelman, B. (2001). HEpiMA: Software for the identification of heterogeneity in meta-analysis. *Computer Methods and Programs in Biomedicine, 64*, 101–107.

Deutsches Cochrane Zentrum. (2011, August 3). *Cochrane Glossar*. Retrieved January 25, 2012, from http://www.cochrane.de/de/cochrane-glossar

Eddy, D. M., Hasselblad, V., & Shachter, R. (1992). *Meta-analysis by the confidence profile method: The statistical synthesis of evidence.* Boston: Academic Press.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629-634.

Eysenck, H. J. (1978). An exercise in mega-silliness. *American Psychologist, 33*, 517.

Franke, G. R. (May 1992). Review ABMA DSTAT. *Journal of Marketing Research, 29*, 276-279.

Freiman, J. A., Chalmers, T. C., Smith, H. J., & Kuebler, R. R. (1978). The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 "negative" trials. *The New England Journal of Medicine, 299*, 690-694.

Gediga, G., & Hamborg, K. C. (2002). Evaluation in der Software Ergonomie: Methoden und Modelle im Software-Entwicklungsprozess. *Zeitschrift für Psychologie, 210*, 40-57.

Gediga, G., Hamborg, K. C., & Willumeit, H. (2000). *Das IsoMetrics Manual.* Osnarbrück: Universität Osnarbrück.

Glass, G. V. (1976). Primary, secondary and meta-analysis of research. *Educational Researcher, 5*, 3-8.

Glass, G. V., & Smith, M. L. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32*, 752-760.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando: Academic Press.

Hunter, J., & Schmidt, F. L. (1990). *Methods of meta-analysis.* Newbury Park: Sage.

Johnson, B. T. (2010, April 7). *DSTAT 2.x (Contemporary Version)*. Retrieved January 30, 2012, from http://meta-analysis.ning.com/group/softwaretocreategraphicsformetaanalysis/forum/topics/dstat-2x-contemporary-version?xg_source=activity

Jordan, P. W. (1998). *An introduction to usability.* Boca Raton: CRC Press.

Kaspar, K., Gediga, K.-C., Sackmann, T., & Hesselmann, J. (2010). Die Effektivität formativer Evaluation bei der Entwicklung gebrauchstauglicher Software: eine Fallstudie. *Zeitschrift für Arbeits- und Organisationspsychologie, 54*, 29-38.

Krueger, R. A., & Casey, M. A. (2000). *Focus groups. A practical guide for applied research.* Thousand Oaks: Sage.

Krug, S. (2002). *Don't make me think! Web Usability: Das intuitive Web.* Bonn: mitp.

Kuss, O., & Koch, A. (1996). Meta-analysis macros for SAS. *Computational Statistics & Data Analysis, 22*, 325-333.

Lau, J., Schmid, C. H., & Chalmers, T. C. (1995). Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care. *Journal of Clinical Epidemiology, 48*, 45-57.

Leandro, G., & Gallus, G. (2005). *Meta-analysis in medical research.* Malden, Oxford, Carlton: Blackwell.

Lewis, S., & Clarke, M. (2001). Forest plots: Trying to see the wood and the trees. *British Medical Journal, 322*, 1479-1480.

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research.* Cambridge, Massachusetts: Harvard University Press.

Lipsey, M. W., & Wilson, D. (2009). *Practical meta-analysis.* Tousand Oaks, California: Sage.

Nielsen, J. (1994). Heuristic evaluation. In J. Nielsen, & R. L. Mack, *Usability inspection methods* (pp. 25-62). New York: Wiley.

Nielsen, J. (1995). *Multimedia and hypertext: The internet and beyond.* Boston: AP Professional.

Nielsen, J. (2000). *Usability engineering.* San Diego: Morgan Kaufmann.

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. *Proceedings ACM CHI'90 Conference*, (pp. 249-256). Seattle.

Normand, S.-L. T. (August 1995). Meta-analysis software: A comparative review. *American Statistician, 49*, 298-309.

Oster, R. A. (1998). An examination of five statistical software packages for epidemiology. *American Statistician, 52*, 267-280.

Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *British Medical Journal, 2*, 1243-1246.

Rampl, H. (2007). *Handbuch Usability*. Retrieved January 30, 2012, from http://www.handbuch-usability.de/

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86*, 638-641.

Rosenthal, R., & Rubin, D. B. (1982). Comparing effect sizes of independent studies. *Psychological Bulletin, 92*, 500-504.

Rosenthal, R., & Rubin, D. B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin, 99*, 400-406.

Roth, P. L. (2008). Software review: Hunter-Schmidt meta-analysis programs 1.1. *Organizational Research Methods, 11*, 192-196.

Rudner, L. M. (2003, July 5). *Meta-Stat: A tool for the meta-analysis of research studies*. Retrieved January 30, 2012, from http://echo.edres.org:8080/meta/metastat.htm

Rustenbach, S. J. (2003). *Metaanalyse: Eine anwenderorientierte Einführung.* Bern: Huber.

Sarodnick, F., & Brau, H. (2011). *Methoden der Usability Evaluation.* Bern: Huber.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62*, 529-540.

Schwarzer, R. (1997, December 22). *Meta-Analysis*. Retrieved January 30, 2012, from http://userpage.fu-berlin.de/~health/meta_e.htm

Schweibenz, W., & Thissen, F. (2003). *Qualität im Web: Benutzerfreundliche Webseiten durch Usability Evaluation.* Berlin: Springer.

Sharp, S., & Sterne, J. (1997). sbe16: Meta-analysis. *Stata Technical Bulletin, 38*, 9–14.

Sharpe, D. (1997). Of apples and oranges, file drawers and garbage: Why validity issues in meta-analysis will not go away. *Clinical Psychology Review, 17*, 881-901.

Stauffer, J. M. (August 1996). A graphical user interface psychometric meta-analysis program for DOS. *Educational and Psychological Measurement, 56*, 675-677.

Sterne, J. A., Bradburn, M. J., & Egger, M. (2001). Meta-analysis in Stata. In M. Egger, G. D. Smith, & D. Altman, *Systematic reviews in health care. Meta-analysis in context.* (pp. 347-369). London: BMJ Books.

Sterne, J. A., Egger, M., & Sutton, A. J. (2001). Meta-analysis software. In M. Egger, G. D. Smith, & D. Altman, *Systematic reviews in health care. Meta-analysis in context.* (pp. 336-346). London: BMJ Books.

The Cochrane Collaboration. (2011, März 25). *About RevMan*. Retrieved Januar 30, 2012, from The Cochrane IMS: http://ims.cochrane.org/revman/about-revman-5

Viechtbauer, W. (2010, October 8). *The metafor package: A meta-analysis package for R*. Retrieved January 30, 2012, from http://www.metafor-project.org/

Völkel, M. (2007, August 21). *Usability inside*. Retrieved January 30, 2012, from http://www.web-blog.net/?wwparam=1328002472

Wang, M. C., & Bushman, B. (1999). *Integrating results through meta-analytic review using SAS software.* Cary: SAS Publishing.

Yuan, Y., & Little, R. J. (2009). Meta-analysis of studies with missing data. *Biometrics, 65*, 487-496.

Zamora, J., Abraira, V., Muriel, A., Khan, K., & Coomarasamy, A. (2006). Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Medical Research Methodology*, 6, 31. doi:10.1186/1471-2288-6-31

Curriculum Vitae

Persönliche Angaben:

Name:	Sandra Geislinger
Geburtstag:	01.06.1980
Geburtsort:	Straubing, Deutschland
Nationalität:	Deutschland
Familienstand:	Ledig

Ausbildung:

2005 – 2012	**Universität Wien** Studium der Psychologie
2001 – 2003	**Techno Color GmbH, Bogen (DE)** Ausbildung zur Industriekauffrau
1991 – 2001	**Anton-Bruckner-Gymnasium, Straubing (DE)**

Berufliche Erfahrung:

08/2006 – dato **Western Union International Bank, Wien**

Projekt- und Prozessmanagement

- Erstellung und Verbesserung von Policies, Arbeitsanweisungen und Flussdiagrammen, Betreuung von Neuprojekten und –produkten, Datenbankbetreuung, Auditvorbereitungen

09/2003 – 09/2005 **Techno Color GmbH, Bogen (DE)**

Projekt- und Qualitätsmanagement

- Durchführung und Dokumentation von Erstbemusterungen, Analyse und Verbesserung auftretender Probleme während der Projektanlaufphase, Erstellung von Arbeitsanweisungen und Schulungen, Freigabe von Serienteilen

Eidesstattliche Erklärung

Ich bestätige, dass ich die vorliegende Diplomarbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe und dass die Arbeit in gleicher oder ähnlicher Form noch bei keiner anderen Prüfungsbehörde vorlag. Alle Ausführungen der Arbeit, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.

Wien, im März 2012 _____

Sandra Geislinger