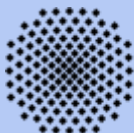


SISS
Schriftenreihe des Instituts für Sozialwissenschaften
der Universität Stuttgart
No. 1 / 2014

Meta-Analyse mit Mehrebenenmodellen unter Verwendung von HLM

Joachim Fiebig
Dieter Urban



Universität Stuttgart
Abteilung für Soziologie und empirische Sozialforschung

**SISS:
Schriftenreihe
des Instituts für Sozialwissenschaften
der Universität Stuttgart:
No. 1 / 2014**

Meta-Analysen mit Mehrebenenmodellen
unter Verwendung von HLM

Joachim Fiebig
Dieter Urban

ISSN 0945-9197

**Universität Stuttgart
Institut für Sozialwissenschaften
Abt. f. Soziologie u. empirische Sozialforschung
70174 Stuttgart**

Joachim Fiebig und Dieter Urban

Meta-Analysen mit Mehrebenenmodellen unter Verwendung vom HLM.

SISS–Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart. No. 1/2014.
(<http://www.uni-stuttgart.de/soz/institut/forschung/2014.SISS.1.pdf>)

Z U S A M M E N F A S S U N G: Dieser Beitrag beschreibt, wie Mehrebenenmodelle zur Meta-Analyse (=MEMA-Technik) eingesetzt werden können. Dabei werden insbesondere die Vorteile von ergänzenden Moderatoranalysen herausgestellt. Der konkrete Analyseablauf einer solchen Analyse wird anhand eines Beispiels und unter Verwendung der Statistik-Software "HLM" vorgestellt.

S C H L A G W O R T E: Meta-Analyse, Mehrebenenanalyse, Mehrebenen-Meta-Analyse, Intraklassenkorrelationskoeffizient (ICC), Moderatoreffekte, Odds-Ratio, HLM, Statistiksoftware.

Joachim Fiebig and Dieter Urban

Meta-Analysis with Multilevel Models using HLM.

SISS–Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart. No. 1/2014.
(<http://www.uni-stuttgart.de/soz/institut/forschung/2014.SISS.1.pdf>)

A B S T R A C T: This paper describes how to use multi-level-models for meta-analysis (=MEMA-technique). A special focus is also set on additional moderator analyses. The report presents an example of an applied meta-analysis within a multi-level-model approach using the HLM software package.

K E Y W O R D S: meta-analysis, multi-level-analysis, multi-level-models for meta-analysis, intraclass correlation coefficient (ICC), moderator analysis, odds-ratio, HLM, statistic software package.

1 Themenstellung¹

Meta-Analysen beschäftigen sich mit den Forschungsergebnissen einer größeren Anzahl von systematisch ausgewählten, themengleichen oder themenähnlichen Primärstudien. Sie versuchen, die Forschungsergebnisse dieser Studien zusammenzufassen und zu bewerten. Insbesondere interessieren sich Meta-Analysen für folgende drei Fragestellungen:

1. Lassen sich die einzelnen Studienergebnisse zu einem Gesamtergebnis zusammenfassen, und wie groß ist z.B. die mittlere Effektstärke, wenn in den Einzelstudien analytisch vergleichbare Einflussbeziehungen zwischen einer abhängigen und einer unabhängigen Variablen ermittelt wurden?
2. Wie stark unterscheiden sich die einzelnen Studien hinsichtlich der von ihnen ermittelten Effektstärken (s.o.), und wie zuverlässig ist die Schätzung von mittleren bzw. durchschnittlichen Effektstärken?
3. Von welchen Studienmerkmalen (z.B. differierende Untersuchungsdesigns und differierende Messmethoden) hängt es ab, dass sich die einzelnen Studienergebnisse von vergleichbaren Primärstudien deutlich voneinander unterscheiden? Welche Studiencharakteristika wirken wie stark als Moderatorvariablen für die einzelnen, unterschiedlich stark geschätzten Effektbeziehungen?

In diesem Beitrag wird die Durchführung von Meta-Analysen mittels Konstruktion und Schätzung von Mehrebenenmodellen unter Verwendung der Statistik-Software "HLM" erläutert.² Dieses Verfahren wird im Text als "MEMA"-Technik bezeichnet (MEMA = Mehrebenen-Meta-Analyse). Ziel der Darstellung ist es, dem Leser die wichtigsten Informationen zur selbstständigen Durchführung einer MEMA zu vermitteln.

Es gibt verschiedene Varianten von MEMA-Verfahren. Im Folgenden werden die Voraussetzungen aufgelistet, die erfüllt sein müssen, damit eine MEMA, so wie sie hier beschrieben wird, praktiziert werden kann. Dabei gehen wir davon aus, dass für die Meta-Analyse nur die aggregierten Resultate der jeweiligen Studien, so wie sie z.B. in Zeitschriftenaufsätzen zu finden sind, zur Verfügung stehen (z.B. absolute oder relative Häufigkeitsverteilungen oder Mittelwerte) und die personenbezogenen Individualdaten der Primärstudien nicht zugänglich sind. Voraussetzungen für die hier beschriebene MEMA sind:

¹ Verschiedene Textblöcke dieses SISS-Beitrags stammen aus dem Skript der Autoren "Quantitative Meta-Analysen zur Überprüfung sozialwissenschaftlicher Hypothesen - dargestellt an einem Beispiel aus der Delinquenzforschung." Stuttgart (SOWI) 2014.

² Informationen zur Statistiksoftware "HLM" finden sich in Raudenbush et al. (2004) und Raudenbush/Bryk (2002) sowie auf der Homepage von SSI-Central: <http://www.ssicentral.com/hlm/examples.html>.

- A)** Es muss eine möglichst große Anzahl von Primärstudien (Forschungsberichte, Zeitschriftenaufsätze, Monographien u.a.) vorliegen, in denen Informationen über die im Folgenden genannten Punkte (B bis E) enthalten sind. Und da stets davon auszugehen ist, dass die zu analysierenden Primärstudien eine Auswahl darstellen, sollte diese Auswahl nicht systematisch verzerrt sein (z.B. dadurch, dass wichtige Literaturdatenbanken bei der Auswahl unberücksichtigt bleiben). Zudem darf die Auswahl auch keinen "publication bias" aufweisen (vgl. dazu Fiebig et al. 2012).

Die Auswahl von Primärstudien, die im hier benutzten Anwendungsbeispiel metaanalytisch untersucht wird, besteht aus 15 Studien (mehr wären besser gewesen, ließen sich aber nicht auffinden), die unter den veröffentlichten Aufsätzen der 13 thematisch wichtigsten Fachzeitschriften und in den 4 umfangreichsten sozialwissenschaftlichen Literaturdatenbanken (u.a.: SSCI) zu ermitteln waren. Für diese Studien konnte kein "publication bias" nachgewiesen werden (s.o).

- B)** Die ausgewählten Primärstudien sollten quantitative Daten zu einem thematisch eng umrissenen, empirisch untersuchten Zusammenhang zwischen einer bestimmten, unabhängigen Prädiktorvariablen und einer abhängigen Kriteriumsvariablen enthalten, sodass die Effektstärke dieses Zusammenhangs mit einem über alle Studien einheitlichen Effektstärkemaß zu berechnen ist.

Im hier benutzten Anwendungsbeispiel aus der Delinquenzforschung interessierte der Zusammenhang zwischen einer pädosexuellen Viktimisierung, die befragte Personen in ihrer Kindheit erlitten haben, und einer späteren, im Erwachsenenalter dieser Personen aufgetretenen, pädosexuellen Delinquenz (These von der pädosexuellen Opfer-Täter-Transition).³ In diesem Beispiel sind die Prädiktorvariable "pädosexuelle Viktimisierung" und die Kriteriumsvariable "pädosexuelle Delinquenz" binomial skaliert (j/n), sodass als Effektstärkemaß das "Odds-Ratio" berechnet werden konnte.

- C)** Um eine MEMA mit Moderatoranalyse durchführen zu können, müssen die ausgewählten Primärstudien hinsichtlich der berechneten Effektstärke (vgl. Pkt. B) ein deutliches Ausmaß an Heterogenität aufweisen.

In unserem Beispiel konnte hinsichtlich der ermittelten Effektstärken eine Studienvarianz von 0,29 und ein Intraklassenkoeffizient (ICC) von 0,43 ermittelt werden (mehr dazu im Folgenden). Beides verweist auf eine genügend große Heterogenität der ausgewählten Primärstudien, um deren Abhängigkeit von bestimmten Studienmerkmalen (z.B. von den jeweiligen

³ Zur These von der pädosexuellen Opfer-Täter-Transition vgl. Urban/Fiebig 2010, 2011.

Stichprobengrößen) untersuchen zu können. Nach einer (eher strengen und deshalb umstrittenen) Daumenregel, sollte der ICC für eine MEMA oberhalb von 0,25 liegen, was in unserem Beispiel deutlich erreicht wird.

- D)** Möglichst alle ausgewählten Primärstudien sollten die gleichen qualitativen und quantitativen Informationen über verschiedenste Eigenschaften der durchgeführten Forschung bereitstellen (z.B. über die Art der Auswahl der untersuchten Probanden, über die Form der Operationalisierung zentraler Studienvariablen, über den Umfang der verwendeten Stichproben usw.).

Im hier verwendeten Beispiel werden zwei binomiale Studienmerkmale benutzt: Zum einen wird die Form der in den einzelnen Primärstudien benutzten Operationalisierung von "pädosexueller Delinquenz" (OP-DEL) benutzt, die "eng" (mit Körperkontakt zwischen Täter und Opfer, OP-DEL=1) oder "weit" ("both hands-on and hands-off"-Täter, OP-DEL=0) sein kann. Und zum anderen wird die Rekrutierung der in den einzelnen Studien befragten Täter berücksichtigt, die über therapeutische Einrichtungen (THERAPIE=1) oder über Justizvollzugsanstalten erfolgt sein kann (THERAPIE=0).

- E)** Die ausgewählten Primärstudien sollten für die hier beschriebene MEMA-Variante nach einem Fall-Kontroll-Design durchgeführt worden sein, d.h. es sollten die unter Pkt. B erwähnten Informationen sowohl für eine Untersuchungsgruppe als auch für eine (oder mehrere) Kontrollgruppen zur Verfügung stehen.

Im Beispiel bilden pädosexuelle Delinquenten (verurteilte Straftäter) die Untersuchungsgruppe und adultsexuelle Delinquenten (ebenso verurteilte Straftäter) die Kontrollgruppe. Andere Kontrollgruppen können nicht-sexuelle Delinquenten oder Personen aus der "Normalbevölkerung" umfassen.

Folgende MEMA-Themen werden in den nachstehenden Gliederungsabschnitten behandelt:

- 2.1 Die Logik der MEMA-Technik
- 2.2 Effektstärken und Heterogenität
- 2.3 Interpretation von MEMA-Schätzungen (anhand eines Beispiels)
- 2.4 Methodische Erläuterungen zur MEMA-Modellierung
- 2.5 Programmtechnische Informationen zur MEMA mittels HLM

2.1 Die Logik der MEMA-Technik

Einfach gesagt ist eine Mehrebenenanalyse (MEA) eines von vielen möglichen statistischen Verfahren, die für multivariate Analysen eingesetzt werden können. Ähnlich wie bei einer Regressionsschätzung wird bei der MEA getestet, in welcher Stärke verschiedene unabhängige Variablen mit einer abhängigen Variablen in Beziehung stehen. Auch in der MEA werden die Effekte aller unabhängigen Variablen durch deren gleichzeitige (simultane) Berücksichtigung in einer Modellschätzung gegenseitig kontrolliert, sodass bereinigte Effekte ermittelt werden können. Das bedeutet, dass mit multivariaten Statistikverfahren im Unterschied zu bivariaten Analysen einzelne Effekte isoliert und Aussagen darüber getroffen werden können, welche Variablen den stärksten bzw. den "entscheidenden" Einfluss auf eine abhängige Variable ausüben.

Insofern unterscheidet sich die MEA kaum von den übrigen multivariaten Analyseverfahren. Der zusätzliche Vorteil, den dieses Verfahren mit sich bringt, besteht in der Möglichkeit, statistische Effekte auf unterschiedlichen Analyseebenen zu unterscheiden und auch schätzen zu können. So kann getestet werden, inwiefern Effekte auf der Individualebene (z.B. der Einfluss individueller pädosexueller Viktimisierung auf die Straftat einzelner Akteure) in Abhängigkeit von Einheiten auf einer übergeordneten Analyseebene (z.B. von unterschiedlichen Methodiken, die in den einzelnen Primärstudien eingesetzt werden) variieren.

In diesem Falle wäre die Methodik kein Merkmal der befragten Personen, sondern sie wäre ein Merkmal der einzelnen Studien und somit als Variable auf einer übergeordneten, zweiten Analyseebene zu untersuchen. So wäre es bspw. denkbar, dass in Studien, in denen eine weite Definition von pädosexueller Viktimisierung verwendet wird (in denen z.B. auch Exhibitionismus als eine Form von Viktimisierung betrachtet wird), andere Zusammenhänge identifiziert werden, als in Studien, in denen eine engere Definition von Viktimisierung zur Anwendung kommt (in denen z.B. nur eine vollzogene Penetration als Viktimisierung betrachtet wird). Wenn diese Variation in einer Analyse nicht berücksichtigt werden würde, könnte dies zu Verzerrungen bei der Schätzung der Effektkoeffizienten führen.

In solchen Fällen erlaubt die MEA die Modellierung von sogenannten "random coefficients", die solche Variationen abbilden können. Dies kann u.a. dadurch geschehen, dass im Verfahren eine Schätzung der Effektvariationen zwischen den Einheiten der zweiten Ebene vorgenommen wird. Auf diese Weise lässt sich erschließen, wie stark die Variation auf der zweiten Ebene ausfällt (hier: auf der Ebene der Studien). Zudem hat diese Schätzung auch den Vorteil, dass damit die Schätzungen der Effektkoeffizienten auf der ersten Ebene (hier: auf der Ebene der Probanden) um den Einfluss von Effekten auf der zweiten Ebene kontrolliert bzw. bereinigt werden können.

Auf diese Weise können unterschiedlichste Variablen, welche die Merkmale von Einheiten höherer Ordnung erfassen, in die jeweiligen Modelle eingebunden werden und können dadurch die

Schätzungen von Effekten auf der Individualebene kontrolliert werden. Solche Variablen einer zweiten Analyseebene können z.B. bei einer Befragung von Straftätern der Typ von Haftanstalt, in denen die Straftäter einsitzen, oder die Größe der Haftanstalt (entsprechend der Zahl der dortigen Insassen) sein. Alle diese Variablen auf der zweiten Analyseebene können benutzt werden, um zu untersuchen, inwiefern der eigentlich interessierende Zusammenhang auf der Individualebene von weiteren Variablen auf einer zweiten Analyseebene beeinflusst wird.

Mit dieser besonderen Form der Modellspezifizierung ist es also möglich, in der statistischen Analyse nicht nur die Wechselwirkungen zwischen Merkmalen einer, sondern auch Wechselwirkungen zwischen mehreren Ebenen zu kontrollieren. Folglich lässt sich mit dem Verfahren der MEA ein viel genaueres Bild von den Zusammenhängen zwischen den verwendeten unabhängigen und der zu erklärenden abhängigen Variablen zeichnen als mit herkömmlichen Ein-Ebenen-Statistikmodellen.

Zusammenfassend lässt sich also sagen, dass die Mehrebenenanalyse ein multivariates Analyseverfahren ist, das im Unterschied zu "konventionellen" Regressionsanalysen die explizite Berücksichtigung von Einheiten auf Analyseebenen höherer Ordnung erlaubt. Sie ermöglicht es, dass die Variation der Effekte auf der Individualebene mittels Schätzung von Effekten auf einer höheren Analyseebene kontrolliert werden kann. Die MEA ist folglich eine Regressionsanalyse, die nicht nur mit fixierten Koeffizientenschätzungen arbeiten muss, sondern auch variable Koeffizientenschätzungen ermöglicht ("random coefficients"). Zudem können Merkmale auf einer zweiten oder höheren Analyseebene explizit in den Modellierungen berücksichtigt werden.

Welche Vorteile hat die MEA nun bei der Durchführung von Meta-Analysen, mit denen die Resultate wissenschaftlicher Primärstudien evaluiert werden sollen?

Wie zuvor bereits geschildert, erlaubt das Verfahren der MEA im Unterschied zu anderen statistischen Analyseverfahren die explizite Berücksichtigung einer hierarchischen Datenstruktur. Hierarchische Datenstrukturen liegen immer dann vor, wenn die Untersuchungseinheiten der Stichprobe nicht aus einer gemeinsamen Population stammen, sondern aus vielen unterschiedlichen Subpopulationen bzw. Clustern gezogen werden. Recht einfach lässt sich dies am Beispiel von Untersuchungen über inhaftierte Straftäter verdeutlichen. Diese Personen sind Untersuchungseinheiten, die aus der Gesamtpopulation aller Sträflinge in der BRD stammen, jedoch kommen sie auch wiederum aus verschiedenen Subpopulationen. Subpopulationen sind z.B. alle Haftinsassen bestimmter Bundesländer oder (eine Ebene tiefer) alle Insassen bestimmter JVA's oder (noch eine Ebene tiefer) alle Insassen bestimmter Zellentakte. Alle diese unterschiedlichen Ebenen "sorgen" dafür, dass sich die Untersuchungseinheiten aus gleichen Subpopulationen aufgrund der dort vorliegenden Randbedingungen ähnlicher sein sollten, als die Untersuchungseinheiten aus

verschiedenen Subpopulationen (z.B. deshalb, weil unterschiedliche Bundesländer bzw. Justizministerien unterschiedliche Gesetzesauslegungen vornehmen oder unterschiedliche Haftanstaltstypen unterschiedliche Haftbedingungen aufweisen, etc.).

Die logische Konsequenz daraus ist, dass die Stichprobendaten, die aus solchen Populationen gezogen werden, nicht als unabhängig zu verstehen sind, was wiederum bedeutet, dass bei Nichtberücksichtigung dieses Sachverhalts in der statistischen Analyse die Möglichkeit besteht, dass die statistischen Schätzungen verzerrt und Rückschlüsse fehlerbehaftet sein können. Daher lohnt sich in solchen Fällen generell die Anwendung einer MEA, mit der verschiedene Variableneffekte auf unterschiedlichen Ebenen berechnet und kontrolliert werden können.

Eine ähnliche Datenstruktur wie im genannten Häftlingsbeispiel ergibt sich auch für die Beobachtungseinheiten von Meta-Analysen. In Meta-Analysen sollen Effekte evaluiert werden, die mit den Daten von individuell befragten/beobachteten Probanden (erste Ebene) errechnet wurden, wobei jede Auswahl und Befragung im Kontext einer bestimmten Forschungsstudie (zweite Ebene) durchgeführt wurde. Aufgrund dieser Zwei-Ebenen-Struktur ist die Wahrscheinlichkeit relativ groß, dass die Ergebnisse bzw. Effekte, die in den einzelnen Studien ermittelt wurden, zumindest in Teilen durch Merkmale der jeweiligen Studie beeinflusst werden. Denn es lässt sich sicherlich annehmen, dass es zwischen den verschiedenen Studien deutliche Unterschiede z.B. hinsichtlich der Reliabilität der darin eingesetzten Erhebungsinstrumente gibt. Folglich ist davon auszugehen, dass die in den Studien erzielten Ergebnisse schon alleine aufgrund dieser methodischen Unterschiede (die auf der zweiten Analyseebene liegen) zumindest zu einem gewissen Ausmaß zwischen den einzelnen Studien variieren "müssen" (vgl. Hox 2010: 206ff).

Anders gesagt: die Annahme, dass die in jeder Studie ermittelten Ergebnisse gänzlich unabhängig vom Design der dort vorgenommenen Untersuchung, der verwendeten Verfahren und anderer methodischer Spezifika sind, ist höchst unwahrscheinlich. Folglich sind in einer Meta-Analyse mehrerer Primärstudien die Probanden, die in jeder Studie untersucht wurden, als Untersuchungseinheiten nichtunabhängiger Teilstichproben zu verstehen und folglich würde eine "konventionelle" eindimensionale Analyse dieser Daten mit gravierenden Problemen behaftet sein (vgl. Ditton 1998: 14ff).

In einem solchen Falle ist es sinnvoll, für Meta-Analysen auf die Konstruktionslogik von MEA-Verfahren zurückzugreifen. Denn bzgl. der Beobachtungseinheiten auf der zweiten Ebene (Primärstudien) lassen sich Merkmale identifizieren, die die einzelnen Studienergebnisse bzw. die einzelnen Effektschätzungen beeinflussen könnten. Zu diesen Studienmerkmalen gehören u.a.: die Breite von Operationalisierungen (z.B. enge vs. weite Fassung von pädosexuellem Missbrauch und pädosexueller Delinquenz), die Festlegung von Stichprobenclustern (Therapieeinrichtungen vs.

Haftanstalten etc.) aber auch die von Studie zu Studie variierende Stichprobengröße.⁴

Wie bei traditionellen Mehrebenenanalysen gilt auch für Mehrebenenanalysen im Rahmen von Meta-Analysen, dass dafür gewisse Anwendungsvoraussetzungen erfüllt sein müssen. Dazu gehört zunächst einmal das Vorhandensein einer hierarchischen Struktur in den zu analysierenden Daten, welche die in den Meta-Analysen zu beobachtenden Ereignisse (hier: Effektschätzungen) systematisch beeinflusst. Dass eine solche hierarchische Struktur gegeben ist, wenn Primärstudien mittels Meta-Analysen ausgewertet werden sollen, wurde in den vorangegangenen Erläuterungen begründet.

Ob die hierarchische Struktur der Daten in und über Primärstudien auch relevant zur Beurteilung der Effektschätzung ist, lässt sich durch den sog. "Intraklassenkorrelationskoeffizienten" (ICC) ermitteln (dazu auch mehr in Abschnitt 2.3). Dieser Koeffizient setzt die in den Studien einer Meta-Analyse beobachtete Gesamtvarianz in Relation zur beobachteten Varianz der Studien auf der zweiten Ebene. Je größer der Koeffizient ist, desto mehr Studienvarianz wird von Studienmerkmalen der zweiten Ebene bestimmt. Nach einer eher strengen Daumenregel⁵ sind Mehrebenenschätzungen in Meta-Analysen dann sinnvoll, wenn der ICC größer als 0,25 ist. Dementsprechend wäre erst dann, wenn auf Studienebene (=zweite Ebene) mehr als 25% der Gesamtvarianz aller Studien statistisch erklärt werden kann, eine MEA zur Durchführung von Meta-Analysen sinnvoll.

Allerdings ergab eine Simulationsstudie von Schulze (2008), dass die Auswahl gerade dieser Schwelle nicht eindeutig zu rechtfertigen ist und sie deshalb nicht empfohlen werden kann (vgl. Hox 2010: 211). Hox selbst ist der Auffassung, dass auch wenn der ICC unterhalb von 0,25 liegt, eine MEA durchgeführt werden sollte, da in Meta-Analysen immer mit unerklärter Varianz zu rechnen sei, für deren Analyse und Kontrolle spezifische "random-effects" (s.o.) geschätzt werden könnten. Zudem könnten mit der Berechnung der "random-effects" in einer MEA die Standardfehler der Effektschätzungen unverzerrt geschätzt werden, was in aller Regel zu einem realistischeren bzw. "konservativeren" Ergebnis führe, bei dem die Schätzwerte für Standardfehler und damit auch die Grenzwerte für Konfidenzintervalle nach oben hin korrigiert werden (Hox 2010: 211; Pötschke 2006: 167).

Eine weitere wichtige Anwendungsvoraussetzung für MEA-gestützte Meta-Analysen sind ausreichend umfangreiche Stichprobengrößen sowohl bei der Anzahl zu untersuchender Primärstudien (Abk.: "K") als auch bei der Anzahl von Probanden, die in jeder Primärstudie

⁴ Je größer eine Stichprobe, desto geringer ist der zu erwartende Stichprobenfehler; was jedoch auch durch Gewichtung tlw. zu berücksichtigen ist (vgl. Hox 2010: 213).

⁵ Diese Daumenregel wurde von Hunter/Schmidt 2004 vorgeschlagen.

untersucht werden (Abk.: "N"). Denn im Unterschied zu Analyseverfahren, die nicht von einer Schichtung der Stichprobendaten ausgehen, taucht bei hierarchischen Daten die Problematik ausreichender Stichprobengrößen für jede einzelne der analysierten Datenebenen auf. So kann z.B. eine Stichprobe von Primärstudien, die insgesamt mehrere tausend Beobachtungen auswerten, für eine Mehrebenenanalyse ungeeignet sein, wenn diese mehrere tausend Beobachtungen in nur zwei oder drei ausgewählten Primärstudien ausgewertet wurden.

Welche Anzahl von Primärstudien für eine Mehrebenen-Meta-Analyse mindestens zur Verfügung stehen sollte, wird in der Literatur unterschiedlich diskutiert. Nach Hox (ders. 2010: 233) sollten nicht deutlich weniger als 50 Fälle auf der zweiten Ebene (Ebene der Primärstudien-Merkmale) für eine Meta-Analyse mit Mehrebenenmodellierung vorhanden sein. Ansonsten sei mit einer Verzerrung der geschätzten Standardfehler zu rechnen.

Es gibt allerdings auch methodische Hinweise darauf, dass das Fallzahlenproblem von Mehrebenenschätzungen nur dann gravierend ist, wenn gleichzeitig die Anzahl von Probanden pro Primärstudie kleiner als 30 ist (vgl. Maas/Hox 2004: 136). Dies ist bei den von uns ausgewerteten Primärstudien bei weitem nicht der Fall. Denn die für unsere Meta-Analyse identifizierten Primärstudien operieren mit deutlich höheren Fallzahlen, sodass wir die o.g. Anwendungsvoraussetzungen für eine MEA-basierte Meta-Analyse erfüllen können.

Die Problematik ausreichend großer Fallzahlen steht in Verbindung mit dem Ausmaß von Spezifität, das in einer Meta-Analyse angestrebt wird. Je spezifischer die Fragestellung einer Meta-Analyse ist, umso größer wird in der Regel die Anzahl an bestimmten Variablen sein, die je Studie benötigt werden, um die interessierende Variablenkombination zu erhalten, und umso weniger Studien wird es geben, die diese spezifische Kombination von Variablen auch aufweisen. Dies hat zur Folge, dass sich bei hoher Spezifität die Fallzahl an zu verwendenden Studien und damit auch der zu betrachtenden Probanden deutlich reduzieren kann.

Um die Fallzahlen (K) von Primärstudien, die in einer MEMA berücksichtigt werden sollen, zu erhöhen, ist es zudem möglich, solche Studien, die nicht in exakt gleicher Weise durchgeführt wurden, für eine Analyse zu poolen. Wenn z.B. die Mitglieder der in den einzelnen Studien verwendeten Untersuchungsgruppen unterschiedlich rekrutiert worden sind, kann diese Unterschiedlichkeit in Form von Dummy-Variablen festgehalten werden, so dass diese Dummy-Variablen als zusätzliche Kontrollvariablen in der jeweiligen Modellschätzung benutzt werden können.

Auf diese Weise kann der Datenpool auszuwertender Primärstudien um solche Studien erweitert werden, die ansonsten aufgrund ihrer Besonderheiten (geringe methodische Güte, unklare Operationalisierung von unabhängiger und abhängiger Variablen etc.) nicht zusammen mit anderen, "konventionell" durchgeführten Studien evaluiert werden könnten. Erst im multivariaten Design

von Mehrgruppen-Modellierungen lässt sich relativ genau schätzen, wie groß (sofern überhaupt gegeben) der Einfluss solcher besonderen Merkmale auf die Effektstärke ist.

Somit lässt sich auf diese Weise nicht nur der Pool von meta-analytisch auszuwertenden Primärstudien vergrößern, sondern lässt sich auch die Selektionsproblematik bei der Auswahl der zu analysierenden Studien entschärfen. Denn jede als relevant identifizierte Studie, die aus einer Meta-Analyse ausgeschlossen wird, könnte eine Studie sein, die ein zusätzliches, möglicherweise sehr wichtiges Licht auf die untersuchten Zusammenhänge wirft. Und möglicherweise könnte sich gerade ein Ausschluss von wenigen, unorthodox durchgeführten Studien verzerrend auf die Analyseergebnisse auswirken.

Weitere ausführliche methodische und programmtechnische Hinweise zur Durchführung einer MEMA finden sich in den Gliederungsabschnitten 2.3 bis 2.5.

2.2 Effektstärken und Heterogenität

Bevor die Forschungsergebnisse der ausgewählten Primärstudien mittels einer MEMA untersucht werden können, müssen die in den einzelnen Studien oftmals unterschiedlich konstruierten Maßzahlen zur Beschreibung der Effektbeziehungen (hier: bzgl. des Zusammenhangs zwischen pädosexueller Viktimisierung und pädosexueller Delinquenz) in die Werte eines einheitlichen, gemeinsam gültigen Effektstärkemaßes überführt werden. Die dabei generierten Effektstärken ergeben dann die Werte der abhängigen Variablen für alle folgenden Meta-Analysen.

In der Literatur wird eine Fülle von unterschiedlichen Effektstärkemaßen vorgestellt. In der folgenden Darstellung orientieren wir uns vor allem an den Ausführungen von Borenstein et al. in ihrem sehr guten Lehrbuch "Introduction to Meta-Analysis" (dies. 2009). Danach können Effektstärkemaße in drei große Gruppen unterteilt werden. Diese sind:

1. Effektstärken auf der Basis von Mittelwerten und Standardabweichungen
(dazu gehören u.a. rohe und standardisierte Mittelwertdifferenzen),
2. Effektstärken auf der Basis von Mittelwerten binärer Daten
(dazu gehören u.a. "Risk-Ratio"- und "Odds-Ratio"-Maße sowie Risikodifferenzen),
3. Effektstärken auf der Basis von Korrelationen
(dazu gehört u.a. das in einen standardisierten z-Wert transformierte "Pearson's r").

Für unsere MEMA, in der die Effektstärken der Beziehung zwischen zwei binomialen Variablen untersucht werden sollen (vgl. Pkt. B in Textabschnitt 1.0), kann ein Effektstärkemaß aus der zweiten, oben benannten Gruppe gewählt werden. Dies ist immer dann möglich, wenn die Summendaten der Primärstudien auf binären Ergebnissen basieren (vgl. Littell et al. 2008: 80). Die

Effektstärkenmaße können dann aus den in den Primärstudien angegebenen Anzahlen von Ereignissen und Nicht-Ereignissen (zumeist in Form einer 2x2-Tabelle) errechnet werden.

Im Konkreten entscheiden wir uns für das Effektstärkemaß: "Odds-Ratio" oder "Chancenverhältnis". Zur Erläuterung dieses Maßes dient die im Folgenden erörterte 2x2-Tabelle (Tabelle 1), die vier Zellen umfasst. Die Anzahl von Fällen, die in jeder Zelle enthalten sind, werden mit den Buchstaben "A, B, C, D" benannt (wie in Tabelle 1 dargestellt). In solch einer Tabelle kann die Verteilung von Fällen (Personen) mit pädosexuellen Viktimisierungserfahrungen (Ereignis/Nicht-Ereignis) über zwei verschiedene Straftätergruppen erfasst werden, wobei die Mitglieder der Untersuchungsgruppe die pädosexuellen Delinquenten und die Mitglieder der Kontrollgruppe nicht pädosexuelle Delinquenten sein könnten.

Tabelle 1: Benennung binärer Daten in einer 2x2 Tabelle.

	Ereignis	Nicht-Ereignis	N
Untersuchungsgruppe	A	B	n_1
Kontrollgruppe	C	D	n_2

Das Odds-Ratio entspricht dem Verhältnis zweier Chancen. Eine Chance ergibt sich aus der Wahrscheinlichkeit dafür, dass ein bestimmtes Ereignis (Y) eintritt, dividiert durch die Wahrscheinlichkeit dafür, dass dieses Ereignis nicht eintritt: " $P(Y)/[1-P(Y)]$ ".⁶ In unserer Meta-Analyse berichtet das Chancenverhältnis bzw. das Odds-Ratio die Chance eines gesuchten Ereignisses in der Untersuchungsgruppe relativ zu der Chance des gleichen Ereignisses in der Kontrollgruppe. Dieses Odds-Ratio eignet sich insbesondere dann als Effektstärkemaß, wenn die ausgewählten Primärstudien die Häufigkeiten und/oder Prozentanteile der von ihnen untersuchten abhängigen Variablen berichten.

Das Odds-Ratio lässt sich berechnen nach der Formel (vgl. dazu Tabelle 1):

$$\text{Odds-Ratio} = \frac{AD}{BC}$$

Häufig wird das Odds-Ratio auch in seiner logarithmierten Form in Meta-Analysen benutzt:

$$\text{LogOddsRatio} = \ln(\text{OddsRatio})$$

Die (approximative) Varianz des Odds-Ratios wird nach folgender Formel berechnet:

$$V_{\text{LogOddsRatio}} = \frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D}$$

⁶ Vgl. dazu Urban/Mayerl 2011: 340ff.

Gleichermaßen ergibt sich der diesbezügliche (approximative) Standardfehler aus der Quadratwurzel der zuvor gebildeten Varianz:

$$SE_{\text{LogOddsRatio}} = \sqrt{V_{\text{LogOddsRatio}}}$$

Logarithmierte Odds-Ratios können natürlich auch wieder in rohe Odds-Ratios überführt werden (mit UL als oberer Grenze und LL als unterer Grenze):

$$\text{OddsRatio} = \exp(\text{LogOddsRatio}),$$

$$LL_{\text{OddsRatio}} = \exp(LL_{\text{LogOddsRatio}})$$

$$UL_{\text{OddsRatio}} = \exp(UL_{\text{LogOddsRatio}})$$

Der Wertebereich von Odds-Ratios reicht von null bis plus unendlich. Ein Wert von 1,00 bedeutet, dass die Chancen für den Eintritt eines bestimmten Ereignisses (z.B. dass eine Variable einen Wert von 1 und nicht von 0 annimmt) in der Untersuchungs- und der Kontrollgruppe gleich groß sind. Ein Odds-Ratio von größer 1,00 bedeutet, dass die Chance in der Untersuchungsgruppe höher ist als in der Kontrollgruppe. Und ein Odds-Ratio von kleiner 1,00 besagt, dass die Chance für die Ausprägung 1 beim interessierenden Merkmal in der Untersuchungsgruppe geringer ist als in der Kontrollgruppe.

Allgemein gilt die Daumenregel, dass Odds-Ratios nahe 1,00 eine sehr schwache Beziehung zwischen Variablen anzeigen, während Odds-Ratios über 3,00 eine positiv starke Beziehung indizieren. Für eine stark negative Beziehung muss das Odds-Ratio einen Wert von 0,33 (Kehrwert von 3,00) unterschreiten.

Bei der Aggregation der einzelnen Effektstärken zu einer durchschnittlichen bzw. mittleren Effektstärke besteht die Möglichkeit, die unterschiedlichen Stichprobengrößen der einzelnen Primärstudien durch einen **Gewichtungsfaktor** (w) zu berücksichtigen. Ein gängiger Gewichtungsfaktor ist die Stichprobengröße (vgl. Hunter/Schmidt 2004: 100f; Shadish/Haddock 2009: 260) oder die inverse Effektstärkenvarianz ("inverse variance weight").

Bei einer Gewichtung mittels Stichprobengröße leisten Primärstudien mit einer größeren Stichprobe einen größeren Beitrag zur mittleren Effektstärke als Primärstudien mit einer kleineren Stichprobe.

Bei einer Gewichtung mittels der inversen Effektstärkenvarianz werden zunächst die primärstudien-spezifischen Gewichte über die inverse Effektstärkenvarianz ermittelt:

$$w = \frac{1}{SE_i^2}$$

Dann wird die gewichtete mittlere Effektstärke entsprechend folgender allgemeinen Formel berechnet (Hedges 1994a: 287):

$$\overline{ES} = \frac{\sum(w \cdot ES)}{\sum w}$$

Zudem kann ggf. ein Gewichtungsfaktor (q) berücksichtigt werden, um potenzielle Messverzerrungen mittels Erfassung der Reliabilitäten der gemessenen Konstrukte in der Analyse zu berücksichtigen (Hunter/Schmidt 2004: 100f).

In jedem Fall folgt jedoch nach Berechnung der mittleren Effektstärke ein Test des sog. Populationsnulleffekts. Dabei wird getestet, ob sich die mittlere Effektstärke signifikant von 0,00 unterscheidet. Dies kann nach Berechnung der mittleren Effektstärke und ihres Stichprobenfehlers mittels Schätzung eines Konfidenzintervalls erfolgen. Unter Verwendung dieses Intervalls kann eine globale Aussage darüber getroffen werden, ob sich der mittlere Effekt über die einzelnen Primärstudien hinweg signifikant von 0,00 unterscheidet: sofern das Konfidenzintervall den Wert 0,00 einschließt, wird die Nullhypothese beibehalten und kann angenommen werden, dass der Effektstärkenmittelwert nicht signifikant von 0,00 verschieden ist. Wenn aber das Konfidenzintervall den Wert 0,00 nicht enthält, wird die Nullhypothese zurückgewiesen. In diesem Fall unterscheidet sich der Effektstärkenmittelwert signifikant von 0,00.

Alternativ kann auch der z -Wert der Standardnormalverteilung berechnet werden. Wenn dieser bei einer angenommenen Irrtumswahrscheinlichkeit von unter 5% größer als der kritische Wert von 1,96 ist, kann auf einen Populationseffekt geschlossen werden (Littell et al. 2008: 81f).

Nachdem aus den Effektstärken der ausgewählten Primärstudien ein Effektstärkenmittelwert bestimmt worden ist, ist in einer Meta-Analyse zu prüfen, in welchem Ausmaß sich die Effektstärken der Primärstudien voneinander unterscheiden. Dazu ist das Ausmaß von deren Variabilität zu ermitteln. Denn der Gesamteffekt ist nur dann ein annehmbarer Schätzer des wahren Populationseffekts, wenn die einzelnen Effektstärken der Primärstudien homogen sind, d.h. nicht in besonderem Maß streuen. Denn wenn die Effektstärken stark voneinander abweichen, ist der mittlere Gesamteffekt nur ein sehr grober Schätzwert für den "wahren" Effektstärkenparameter in der Population.

Sollte keine Homogenität der Effektstärken vorliegen, muss die Varianz der einzelnen Effektstärken untersucht werden. Denn signifikante Heterogenität bedeutet, dass bestimmte Studienmerkmale (z.B. der Umfang der jeweiligen Stichprobe von Primärstudien) eine Überschussvariabilität ("excess variability") in den Effektstärken produzieren könnten (Rosenthal/DiMatteo 2001: 69).

Da eine allzu große Heterogenität unter den Effektstärken auf die Wahrscheinlichkeit

hindeutet, dass bestimmte Studiencharakteristika in Form von Moderatorvariablen für die Variabilität der Effektstärken verantwortlich sein können, sollten in diesem Falle im Kontext von Meta-Analysen immer Moderatoranalysen durchgeführt werden. Dazu sind Moderatorvariablen zu definieren und auch inhaltlich zu messen, die, abstrakt betrachtet, die Primärstudien entsprechend ihrer Ausprägungen in verschiedene Subgruppen aufteilen können. Denkbar sind z.B. methodische Moderatorvariablen (wie bspw. die methodische Güte der Primärstudien) oder auch inhaltliche Moderatorvariablen (wie bspw. das durchschnittliche Alter der in jeder Studie untersuchten Probanden). Es wird dann in einer Moderatoranalyse angenommen, dass sich eine im Vergleich zu der Gesamtvarianz geringere Varianz in den Subgruppen zeigen sollte, sofern die Moderatorvariablen eine Erklärungskraft für die Varianz der Effektstärken besitzen.

Allerdings können auch signifikante Moderatoren für eine nicht signifikant heterogene Streuung der Effektstärken gefunden werden. Somit wäre in einer Meta-Analyse auch unabhängig vom Ergebnis des Heterogenitätstests eine Überprüfung von Moderatoren sinnvoll (Hall/Rosenthal 1995: 399).

Natürlich gilt auch für die Interpretation der Ergebnisse von Moderatoranalysen das gleiche Risiko wie für die Interpretation anderer multivariater Analyseverfahren: Sofern signifikante Moderatoren gefunden werden ist bzgl. der Ziehung kausaler Schlüsse stets eine große Vorsicht geboten. Beispielsweise kann in einem Fall, in dem Studien mit "geringer Qualität" größere bzw. stärkere Effekte produzieren als Studien mit "hoher Qualität", die Schlussfolgerung, dass Studien mit "geringer Qualität" stets überhöhte oder verzerrte Ergebnisse produzieren, nicht korrekt sein. Denn die Studien mit unterschiedlichen Qualitäten könnten sich auch bzgl. anderer Merkmale unterscheiden, die mit der "Studienqualität" (gering vs. hoch) korrelieren und die nicht als Drittvariablen im entsprechenden Analysemodell berücksichtigt werden.

2.3 Interpretation von MEMA-Schätzungen (anhand eines Beispiels)

Aufgrund der vorangegangenen Erläuterungen dürfte deutlich geworden sein, was den Kern einer Meta-Analyse ausmacht. Eine jede Meta-Analyse zielt auf die Schätzung der mittleren Effektstärke von Einflussbeziehungen, die in einer möglichst großen Auswahl/Stichprobe von Primärstudien untersucht worden sind sowie auf die Ermittlung der Varianz dieser Effektstärken über die verschiedenen Studien hinweg. Wird dann festgestellt, dass die Varianz der Effektstärken signifikant ist und somit nicht von einer Homogenität der Studienergebnisse auszugehen ist, kann die Meta-Analyse mittels Moderatoranalysen ermitteln, ob die Varianz der Effektstärke von Merkmalen der untersuchten Primärstudien (wie z.B. von den unterschiedlichen Operationalisierungen zentraler Konzepte) abhängt.

Der Anteil der systematischen Varianz zwischen den Studien kann im Kontext einer

Mehrebenen-Meta-Analyse ermittelt werden (ausf. methodische Informationen zum Folgenden finden sich in Textabschnitt 2.4). So wird in der MEMA der Anteil der systematischen Varianz an der Gesamtvarianz mittels eines sog. "Intraklassenkorrelationskoeffizienten" (ICC) nach der Formel $\rho = \sigma_u^2 / (\sigma_u^2 + \sigma_e^2)$ geschätzt (σ_u^2 steht für die systematische Varianz der Effektstärke, σ_e^2 steht für die Varianz des zufälligen Fehlers e_j).

Um also den ICC zu berechnen, wird neben der systematischen Varianz auch die Varianz des zufälligen Fehlers benötigt. Diese Varianz wird auch als Studienvarianz σ_e^2 bezeichnet. Bei ihr handelt es sich um die Varianz der Effektkoeffizienten innerhalb jeder untersuchten Studie. Diese kann im Rahmen einer MEMA nicht direkt geschätzt werden, da dort nur Aggregatdaten (Studienmerkmale) und keine Individualdaten zur Verfügung stehen (mehr dazu im Folgenden). Deshalb wird die Studienvarianz in der MEMA in einem gesonderten Verfahrensschritt modellextern berechnet.

Wenn die Studienvarianz ermittelt wurde, kann eine MEMA durchgeführt werden. Bei dieser Analyse wird zunächst ein Linearmodell mit der abhängigen Variablen "Effektstärke" geschätzt, das als "Nullmodell" spezifiziert wird. In diesem Nullmodell wird nur ein Interceptwert und dessen Varianz unter Ausschluss unabhängiger Variablen geschätzt.⁷ Der Anteil der systematischen Varianz zwischen den Studien kann dann, wie oben erläutert, mittels des Intraklassenkorrelationskoeffizienten (ICC) bestimmt werden. Können sodann mittels Schätzung weiterer Linearmodelle, in die unterschiedlichste Studienmerkmale als unabhängige Variablen einbezogen werden, bedeutende Anteile systematischer Varianz der Effektstärke gebunden werden, sodass diese Merkmale einen Einfluss auf das Ausmaß und die Richtung von Veränderungen der Effektstärke haben, können auf diese Weise wichtige Moderatorvariablen aufgedeckt werden.

Der ICC verringert sich in dem Maße, in dem die im Modell berücksichtigten erklärenden (Moderator-)Variablen die beobachtete systematische Varianz der abhängigen Variablen statistisch erklären bzw. binden. Eine starke Verringerung der systematischen Varianz (und damit des ICC) indiziert somit, welche Studienmerkmale für die Variation der Effektstärken zwischen den Studien in welcher Weise verantwortlich sind. Somit interessiert in der MEMA nicht nur die Signifikanz der geschätzten Regressionskoeffizienten bestimmter Moderatorvariablen, sondern auch das Ausmaß mit dem diese Variablen den ICC zu reduzieren vermögen.

Ein besonderer Vorteil der MEMA ist, dass die Schätzung der mittleren Effektstärke über alle Studien und die Schätzung von deren Varianz (zwischen den Studien) nicht in separaten Schritten erfolgen muss. In der MEMA kann eine kombinierte Schätzung der mittleren Effektstärke und der Zwischen-Studienvarianz vorgenommen werden. Auch können in der MEMA mehrere

⁷ Dieses Nullmodell einer MEMA entspricht dem random-effects model der kMA nach Hedges und Olkin 1985 (vgl. Hox 2010: 209). Auch darin wird durch die Bestimmung der Varianz von σ_u^2 die Variation der Effektstärke ermittelt und durch die Bestimmung einer Irrtumswahrscheinlichkeit p ein indirekter Heterogenitätstest vorgenommen.

Moderatorvariablen simultan in einer multivariaten Modellschätzung eingebunden werden, womit deren partieller (bzw. drittvariablen-kontrollierter) Einfluss sowohl auf die mittlere Effektstärke als auch auf die Zwischenstudien-Varianz getestet werden kann.

Im Folgenden werden die Ergebnisse einer MEMA-basierten Schätzung von mittlerer Effektstärke und deren Variation sowie die Ergebnisse der Schätzung von Moderatoreffekten anhand eines Beispiels erörtert.

Die folgende Tabelle 2 zeigt in ihrem linken Teil die Ergebnisse der MEMA bei der Schätzung eines Nullmodells (s.o). In diesem Nullmodell wurden als Werte der abhängigen Variablen die Effektstärken in Form von Odds-Ratios (s.o. und Abschnitt 2.4) für die Untersuchungsgruppe "pädosexuelle Straftäter" und für die Kontrollgruppe "adultsexuelle Straftäter" benutzt.

Die Ergebnisse zur mittleren Effektstärke lassen sich aus der Zeile mit der Beschriftung "Konstante" und der Spalte, die mit "exp(g)" überschrieben ist, entnehmen. In der Untersuchungsgruppe ist die Chance, dass eine pädosexuelle Viktimisierung vorliegt, exakt 2,68mal größer als die entsprechende Chance in der Kontrollgruppe. Somit bezeichnet der Wert von 2,68 die mittlere Effektstärke, wenn diese mit der Maßzahl "Odds-Ratio" gemessen wird.

Das Ausmaß der Variation der Effektstärken zwischen den Studien bzw. deren Heterogenität lässt sich aus der Zeile "Level 2" und den Spalten "VAR" ablesen. Die Irrtumswahrscheinlichkeit, die jeweils in der dritten Spalte einer jeden Teiltabelle enthalten ist, informiert über die Irrtumswahrscheinlichkeit bezüglich der Annahme, dass die Variation zwischen den Studien nicht zufällig ist. In unserem Nullmodell beträgt diese Varianz 0,29 und ist hochgradig signifikant ($p=0,02$).

Die Zeile "Level 1" enthält die Studienvarianz (Varianz der Effektkoeffizienten bzw. des Odds-Ratios) innerhalb der jeweiligen Studien. Diese sind, wie oben bereits erwähnt, vor Durchführung der MEMA mit dem EDV-Programm HLM "händisch" zu berechnen und einzulesen (vgl. dazu die folgenden Abschnitte 2.4 und 2.5). Diese Varianz beträgt im Nullmodell von Tabelle 2 exakt 0,39.

Neben der ermittelten Varianz auf Studienebene (Level 2) wird in Tabelle 2 noch ein weiteres Kriterium zur Beurteilung der vorliegenden Studienvarianz ausgewiesen. Gemeint ist der Wert des oben erläuterten Intraklassenkorrelationskoeffizienten (ICC). Wie erinnerlich, sollte der ICC nach der sehr strengen Vorgabe von Hunter/Schmidt (2004) oberhalb von 0,25 liegen, damit von einem substantiellen Ausmaß an Studienvarianz gesprochen werden kann. Mit einem Wert von 0,43 liegt der Wert in Tabelle 2 beträchtlich oberhalb dieses Grenzwerts und verlangt somit nach einer

Moderatoranalyse.⁸

Als Moderatorvariablen werden in Modell 1 (rechter Teil von Tabelle 2) die beiden binomialen Variablen "Operationalisierung der pädosexuellen Delinquenz" (OP-DEL) und "Rekrutierung über eine therapeutische Einrichtung" (THERAPIE) eingesetzt.⁹ Für die beiden Prädiktorvariablen in Modellschätzung 1 ergibt sich nun folgender Befund: Der Effekt der Moderatorvariablen "Operationalisierung der pädosexuellen Delinquenz" wird signifikant ($p=0,05$) und auch der Effekt der Variablen "Rekrutierungsort" verfehlt die 5%-Signifikanzgrenze nur knapp ($p=0,06$). Insbesondere die Variable OP-DEL hat einen substantiell und inferenzstatistisch sehr bedeutsamen Moderatoreffekt. Wenn Studien eine enge Messung der pädosexuellen Delinquenz vornehmen, erhöht das noch einmal die mittlere Effektstärke um das 2,27-Fache, sodass sich eine mittlere Effektstärke von $3,20 \times 2,27 = 7,26$ ergibt. Somit ist unter dieser Bedingung die Chance für eine pädosexuelle Delinquenz in der Untersuchungsgruppe um das 7,27-Fache größer als in der Kontrollgruppe.

Wie den Zahlen im rechten Teil von Tabelle 2 weiterhin zu entnehmen ist, ist auch in diesem Schätzmodell die Konstante signifikant ($p<0,01$) und größer als 1,00, sodass auch mit dieser Modellschätzung die POTT-Hypothese (s.o.) nicht verworfen werden kann.

Die Passung von Modell 1 ist gemessen an der Differenz der FML-Devianzwerte (vgl. dazu Abschnitt 2.4) besser als die des Nullmodells. Folglich ist Modell 1 dem Nullmodell vorzuziehen.

Da zudem die Level 2-Varianzkomponente bestenfalls noch marginal signifikant ist ($p=0,08$), ergibt sich in diesem Modell ein nicht unerheblicher Erfolg bei der Varianzaufklärung (die Varianz der Konstanten reduziert sich von 0,29 im Nullmodell auf 0,15 in Modell 1, was eine Reduktion um etwa die Hälfte bedeutet). Gleiches gilt auch für die Reduktion des ICC.

Interpretiert man die Ergebnisse inhaltlich, lässt sich mit Blick auf die Konstante festhalten, dass pädosexuell viktimisierte Straftäter eine deutlich erhöhte Chance haben, ein pädosexuelles Delikt (anstatt eines adultsexuellen Delikts) zu begehen. Und der Einfluss der Variablen zur Operationalisierung indiziert, dass Studien mit einer "engen" Definition pädosexueller Delinquenz einen stärkeren diesbezüglichen Zusammenhang aufdecken als Studien, die eine "weite" Definition

⁸ Zu weiteren Validierungszwecken wurden alle Schätzungen auch mit dem FML-Algorithmus durchgeführt (dazu mehr in Abschnitt 2.4). Auch diese Schätzungen zeigen in der Substanz die gleichen Befunde und unterscheiden sich nur geringfügig hinsichtlich der absoluten Koeffizientenschätzwerte. Damit lässt sich aus diesen Ergebnissen insgesamt zweifelsfrei ableiten, dass unter Verwendung der MEMA zusätzliche Moderatoranalysen durchgeführt werden sollten.

⁹ Zur Erinnerung: Im hier verwendeten Beispiel werden zwei binomiale Studienmerkmale benutzt: Zum einen wird die Form der in den einzelnen Primärstudien benutzten Operationalisierung von "pädosexueller Delinquenz" (OP-DEL) benutzt, die "eng" (mit Körperkontakt zwischen Täter und Opfer, OP-DEL=1) oder "weit" ("both hands-on and hands-off"-Täter, OP-DEL=0) sein kann. Und zum anderen wird die Rekrutierung der in den einzelnen Studien befragten Täter berücksichtigt, die über therapeutische Einrichtungen (THERAPIE=1) oder über Justizvollzugsanstalten erfolgt sein kann (THERAPIE=0).

verwenden.

Tabelle 2: Multivariate MEMA-Moderatoranalyse (nach RML-Schätzmethode)

	Nullmodell			Modell 1		
	exp(g)		p	exp(g)		p
Konstante	2,68		0,00	3,20		0,00
OP-DEL (eng)	-		-	2,27		0,05
THERAPIE (ja)	-		-	0,50		0,06
	VAR	df / chi ²	p	VAR	df / chi ²	p
Level 2	0,29	14 / 28,27	0,02	0,15	12 / 19,21	0,08
Level 1	0,39			0,39		
ICC	0,43			0,28		
Deviance _{RML}	53,37			48,44		
Deviance _{FML}	50,13			44,86		
Fallzahlen	K=15 N=2002			K=15 N=2002		

2.4 Methodische Erläuterungen zur MEMA-Modellierung

Eine Mehrebenenmodellierung ist im Rahmen einer Meta-Analyse deshalb möglich, weil auch bei der Auswertung von Daten aus bzw. über Primärstudien von einer hierarchischen Datenstruktur auszugehen ist: Auf der ersten Ebene der Daten sind Eigenschaften der in den einzelnen Studien beobachteten Probanden anzusiedeln (wie z.B. deren jeweiliges Alter und deren jeweilige Viktimisierungserfahrung), und auf der zweiten, übergeordneten Datenebene befinden sich studienspezifische Merkmale der ausgewerteten Primärstudien (wie z.B. die in jeder Studie analysierte Anzahl von Fällen und die in jeder Studie benutzte Definition einer Vergleichs-/Kontrollgruppe).

Da in einer MEMA u.a. die Effekte der Studienmerkmale (zweite Datenebene) auf eine abhängige Kriteriumsvariable geschätzt werden sollen, ist für die Analyse zunächst diese abhängige Variable (Y) zu definieren. In unserer MEMA, die wir im vorangegangenen Gliederungsabschnitt 2.3 erläutert haben, ist Y die Stärke des Effekts, mit dem eine pädosexuelle Viktimisierung zu einer pädosexuellen Delinquenz führt und die in den einzelnen Primärstudien untersucht wurde.

Als Maß für die Effektstärke des Einflusses von pädosexueller Viktimisierung auf pädosexuelle Delinquenz benutzten wir das für jede Einzelstudie separat kalkulierte Odds-Ratio (in logarithmierter Form).

Das Odds-Ratio als abhängige Variable

Als "Odds-Ratio" (OR) wird das Verhältnis zweier "Chancen" verstanden. Jede Chance (engl. "odds") lässt sich als Quotient aus der Wahrscheinlichkeit, dass ein bestimmtes Ereignis eintritt und der Wahrscheinlichkeit dafür, dass es nicht eintritt, berechnen (odds: $P(Y)/[1-P(Y)]$). In unserem Zusammenhang bedeutet eine Chance von 1,00, dass die Wahrscheinlichkeit für eine pädosexuelle Delinquenz mit vorausgegangener pädosexueller Viktimisierung gleich groß der Wahrscheinlichkeit für eine pädosexuelle Delinquenz ohne vorausgegangene pädosexuelle Viktimisierung ist. Alle Werte größer 1,00 signalisieren, dass die Wahrscheinlichkeit für eine pädosexuelle Delinquenz bei vorausgegangener Viktimisierung größer ist als die Wahrscheinlichkeit für eine pädosexuelle Delinquenz ohne vorausgegangene Viktimisierung. Im Gegensatz dazu bedeuten dann Werte kleiner als 1,00, dass die Wahrscheinlichkeit für eine Delinquenz ohne Viktimisierung größer ist als die Wahrscheinlichkeit für eine Delinquenz mit Viktimisierung.

Nun lässt sich nicht nur die Chance für eine pädosexuelle Delinquenz als Quotient berechnen, sondern ein weiterer Quotient kann auch aus zwei Chancen berechnet werden, nämlich aus der oben erläuterten Chance, die mit den Daten von pädosexuellen Delinquenten ermittelt wird, und aus einer weiteren Chance, die mit den Daten von Mitgliedern einer Kontrollgruppe berechnet wird. In den Studien, die wir hier ausgewertet haben, bestand die Kontrollgruppe aus adultsexuellen Straftätern. Auch unter diesen Tätern gab es Probanden, die in ihrer Kindheit pädosexuell missbraucht worden sind. Somit lässt sich auch für adultsexuelle Straftäter die Chance berechnen, dass sie vor ihrer Straftat pädosexuell viktimisiert worden sind.

Wenn nun die Chance einer pädosexuellen Viktimisierung unter pädosexuellen Delinquenten (=Untersuchungsgruppe, UG) in Relation gesetzt wird zur Chance einer pädosexuellen Viktimisierung unter adultsexuellen Delinquenten (=Kontrollgruppe, KG) ergibt sich die von uns in unserer MEMA benutzte abhängige Variable in Form eines Odds-Ratio:

$$OR = [P(Y)_{UG}/1-P(Y)_{UG}] / [P(Y)_{KG}/1-P(Y)_{KG}]$$

Mit diesem Odds-Ratio wird gemessen, um wie viel größer bzw. kleiner die Chance einer pädosexuellen Viktimisierung unter pädosexuellen Delinquenten (= Mitglieder der Untersuchungsgruppe) im Verhältnis zur Chance einer pädosexuellen Viktimisierung unter adultsexuellen Delinquenten (= Mitglieder der Kontrollgruppe) ist.¹⁰

Da das Odds-Ratio asymmetrisch um 1,00 verteilt ist (OR: $0 \rightarrow 1,00 \rightarrow \infty$), wird in der MEMA das logarithmierte Odds-Ratio als abhängige Variable benutzt: (ln(OR): $-\infty \rightarrow 0,00 \rightarrow$

¹⁰ Eine Bestätigung der POTT-Hypothese wird dementsprechend durch ein Odds-Ratio signifikant größer als 1,00 erreicht.

$+\infty$). Wenn also im Nullmodell (s.u.) einer MEMA-Regression mit der abhängigen Variablen "ln(OR)" ein Regressionskoeffizient "g₀" (=Konstante auf der zweiten Modellebene, s.u.) von größer als 0,00 geschätzt wird, bedeutet dies in unseren Analysen, dass nach den Ergebnissen der von uns ausgewerteten Primärstudien die durchschnittliche Chance für eine pädosexuelle Viktimisierung unter den Mitgliedern der Untersuchungsgruppe größer ist als unter den Mitgliedern der Kontrollgruppe. Alle Werte oberhalb von 0,00 würden somit die POTT-Hypothese bestätigen.

Zu beachten ist allerdings, dass in unserer Tabelle 2 (Abschnitt 2.3) nicht "g₀" sondern "exp(g₀)" ausgewiesen wird (s.u.) und exp(g₀) muss zur Bestätigung der POTT-Hypothese nicht größer als 0,00 sondern größer als 1,00 sein (mehr dazu im Folgenden).

In unserer Meta-Analyse wird das Odds-Ratio für jede Primärstudie, die in die Untersuchung aufgenommen werden soll, zunächst nach der folgenden Formel (1) berechnet (vorausgesetzt die Daten sind in Form von Tabelle 3 gegeben, s.u.):

$$OR = \frac{AD}{BC} \quad (1)$$

Sodann muss das Odds-Ratio für die MEMA noch logarithmiert werden. In SPSS geschieht dies z.B. mittels der Anweisung:

COMPUTE LOR = ln(OR).

Tabelle 3: Zweidimensionale Tabelle zur Berechnung des Odds-Ratio

	pädosex. Vikt.: ja	pädosex. Vikt.: nein	N
pädosexuelle Delinquenten (UG)	A	B	n_1
adultsexuelle Delinquenten (KG)	C	D	n_2

Wenn die abhängige Variable "ln(OR)" bzw. "LOR" (s.o.) berechnet und auch die Studienmerkmale als Werte von unabhängigen Z-Variablen (binär oder metrisch) abgespeichert wurden (z.B.: Z₁= Studie verwendet die enge Operationalisierung von Delinquenz: j=1/n=0; Z₂=Anzahl untersuchter Fälle in jeder Studie), dann sind die ersten Voraussetzungen erfüllt, um eine MEMA durchzuführen.

Bevor wir aber zeigen, wie die weiteren Input-Größen für die MEMA berechnet werden, wollen wir zunächst das formale Grundmodell der Mehrebenenmodellierung in einer Meta-Analyse verdeutlichen (die programmtechnische Umsetzung der MEMA mit dem Softwarepaket HLM wird dann in Gliederungsabschnitt 2.5 erläutert).

Die Mehrebenenmodellierung der Meta-Analyse¹¹

Die formale Logik von Mehrebenenmodellen zur Meta-Analyse lässt sich wie folgt skizzieren: Für jede j -te Primärstudie ergibt sich der Wert der abhängigen Variablen "Y" bzw. der Wert der logarithmierten Effektstärke "ln(OR)" aus seinem wahren Wert in der Population (β_0) und einem Stichprobenfehler (e), was in der folgenden Gleichung (2) verdeutlicht wird:

$$Y_j = \beta_{0j} + e_j \quad (2)$$

In Gleichung (2) sind keine Prädiktoreffekte von X-Variablen (βX) enthalten. Dies unterscheidet die spezielle Mehrebenenmodellierung der Meta-Analyse von der allgemeinen Mehrebenenmodellierung, bei der in die Modellgleichung (2) auch X-Prädiktoren bzw. Level-1-Prädiktoren aufgenommen werden können. Solche X-Variablen wären z.B. das jeweilige Alter oder die jeweilige Ausbildung einzelner Probanden, die in den diversen Primärstudien untersucht worden sind. Solche Daten stehen uns allerdings für unsere Meta-Analyse nicht zur Verfügung, da es uns annahmegemäß (vgl. Gliederungsabschnitt 1) nicht möglich ist, die Roh-Datensätze der einzelnen Primärstudien zu erhalten und diese erneut sekundäranalytisch auszuwerten.

Bezüglich Gleichung (2) wird angenommen, dass der Stichprobenfehler " e_j " normalverteilt ist und eine bestimmte Varianz (σ^2_j) aufweist. Diese Varianz wird auch als "Studienvarianz" oder als "Level-1 Varianz" (entsprechend der inhaltlichen Logik von MEMA) bezeichnet:

$$\text{Var}(e_j) = \text{Studienvarianz} = \text{Level-1 Varianz} = \sigma^2_j \quad (2a)$$

Wie Gleichung (2) zeigt, kann die abhängige Variable "Y" bzw. die Effektstärke "ln(OR)" über die verschiedenen j -ten Primärstudien hinweg variieren. Die Variation der Effektstärken wird dadurch ausgelöst, dass β_0 von verschiedenen Merkmalen (Z) der Primärstudien abhängt. So ist z.B. die Anzahl der Probanden, die in einer Primärstudie untersucht werden, eine Variable (Z), die einen Einfluss (γ) auf β_0 ausübt. Denn alle Z-Variablen sind Prädiktoren auf Ebene 2 (Level-2) des Mehrebenenmodells, da sie sich auf Merkmale der einzelnen Primärstudien (Ebene 2) und nicht auf Merkmale von einzelnen Probanden (Ebene 1) beziehen. Die folgende Gleichung (3) zeigt diesen Zusammenhang.

$$\beta_{0j} = \gamma_0 + \gamma_1 Z_{1j} + \dots + \gamma_p Z_{pj} + u_j \quad (3)$$

Wie Gleichung (3) verdeutlicht, haben die Z-Variablen keinen deterministischen Einfluss auf β_0 . Es gibt immer noch ein Residuum " u ", das zur Größenbestimmung von β_0 beiträgt. Wie auch schon bei den e -Fehlern (s.o.) wird auch für die u -Fehler angenommen, dass sie normalverteilt sind und eine

¹¹ Bei dieser Darstellung folgen wir im Wesentlichen den Ausführungen von Hox 2010: 205-230.

bestimmte Varianz " σ_u^2 " aufweisen:

$$\text{Var}(u_j) = \text{Level-2 Varianz} = \sigma_u^2 \quad (3a)$$

Wenn nun Gleichung (3) in Gleichung (2) eingefügt wird, ergibt sich die folgende Gleichung (4):

$$Y_j = \gamma_0 + \gamma_1 Z_{1j} + \dots + \gamma_p Z_{pj} + u_j + e_j \quad (4)$$

Falls die Level-2 Varianz " σ_u^2 " signifikant ist, beschreibt die Gleichung (4) auch ein "random coefficient"-Modell, da die durch "Y" abgebildeten Zusammenhänge von Primärstudie zu Primärstudie in Abhängigkeit von "Z" variieren.

Sollte es in einem Modell keine Z-Variablen zur Erklärung der Variation von Y geben, so reduziert sich die Modellgleichung (4) zu folgendem Modell (5), das ebenfalls ein "random-coefficient"-Modell ist:

$$Y_j = \gamma_0 + u_j + e_j \quad (5)$$

Das Modell (5) wird auch als "Nullmodell" bezeichnet. Jede Mehrebenenanalyse beginnt immer mit der Schätzung des Nullmodells. Diese Schätzung (mit einer abh. Variablen, aber ohne Prädiktorvariablen) dient als Referenz für die Beurteilung des Schätzerfolgs von vollständigen Modellen (Modelle mit Prädiktor/en). Der Schätzwert " γ_0 " für die Konstante (bzw. für das Intercept) " γ_0 " im Nullmodell lässt sich als Schätzung des Erwartungswerts oder (praxisnäher:) des Mittelwerts der abhängigen Variablen interpretieren. In unserem Fall informiert also die Konstante über die durchschnittliche Effektstärke (bzw. das mittlere Odds-Ratio) von allen Primärstudien, die in die MEMA einbezogen wurden.

Die Schätzung der Varianzkomponente " u " im Nullmodell liefert ein Maß für die Variation der Effektstärke zwischen den einzelnen Studien. Je geringer die u -Varianz (σ_u^2 bzw. Level-2-Varianz) ist, d.h. je weniger Variation für das Intercept zwischen den Einheiten der zweiten Ebene vorliegt (u_{0j}), desto geringer ist auch der empirische Anlass, ein Mehrebenenmodell zu schätzen. In Modell (4) ist " σ_u^2 " die Varianz von " u " zwischen den Studien, nachdem die erklärenden Z-Variablen in das Modell eingebracht wurden. Somit ist die Differenz zwischen " σ_u^2 " im Nullmodell und im Prädiktorenmodell als derjenige Varianzanteil zu interpretieren, der im Modell von den Z-Variablen statistisch erklärt werden kann.

Die Koeffizienten " γ " der oben erläuterten Gleichung (4) werden in der MEMA mittels eines bestimmten Schätzverfahrens (s.u.) als " γ -Koeffizienten" geschätzt. Sind es γ -Koeffizienten, die zu einer bestimmten Z-Variablen gehören, so indizieren sie die Einflussstärke dieser unabhängigen Z-Variablen (z.B. von der "Fallzahl einer jeden Studie") auf die Effektstärke " $\ln(\text{OR})$ ".

Die Gleichung A10.4 kann auch entlogarithmiert werden. Dann wird aus Y bzw. aus $\ln(OR)$ die neue abhängige Variable "OR" und aus den g -Koeffizienten werden die neuen Koeffizienten $\exp(g)$. Sie lassen sich dann im Sinne eines Multiplikationsfaktors interpretieren. So wird z.B. aus einem g von "1,25" der neue Wert $\exp(1,25)=3,50$, sodass ein geschätzter Interceptwert von $\exp(0,69)=2,00$ auf einen neuen Wert von $2,00 \times 3,50=7,00$ ansteigt, wenn sich die dazugehörige Z -Variable um eine empirische Einheit vergrößert. Dies könnte im hier gegebenen Anwendungsbeispiel bedeuten, dass bei einem Wechsel der Z -Prädiktorvariablen von $Z=0$ auf $Z=1$ das Chancenverhältnis zwischen der Chance, nach einer pädosexuellen Viktimisierung auch selbst pädosexuell delinquent zu handeln, und der Chance, nach einer pädosexuellen Viktimisierung ausschließlich adultsexuelle Delinquenzfolgen zu zeigen, vom 2-Fachen auf das 7-Fache zugunsten einer späteren pädosexuellen Delinquenz ansteigen kann (zur Logik der hier berichteten Einflusskoeffizienten vgl. auch Urban/Mayerl 2011: 340-345).

MEMA-Schätzmethode (RML vs. FML) und Moderatoranalyse¹²

Zur Schätzung der MEMA-Modellparameter können zwei verschiedene Varianten des Maximum-Likelihood-Schätzverfahrens eingesetzt werden: die RML- und die FML-Methode. Generell betrachtet wird bei einer MEA die Varianz des Intercepts mit dem "restricted maximum likelihood"-Algorithmus (RML) exakter geschätzt als mit dem "full maximum likelihood"-Algorithmus (FML).¹³ Da sich gerade bei einer MEMA (im Unterschied zu einer konventionellen MEA) das Interesse besonders auf das Ausmaß der zu schätzenden Varianz richtet (s.o.), ist dafür die Verwendung der RML-Methode zu empfehlen.

Der FML-Algorithmus sollte hingegen benutzt werden, wenn verschiedene, ineinander geschachtelte Modelle, die z.B. verschiedene Untergruppen von Moderatorvariablen verwenden, miteinander zu vergleichen sind. Denn wenn dabei die FML-Methode benutzt wird, können die Schätzwerte der Regressionskoeffizienten und die Anpassungswerte ("Deviances") von mehreren geschachtelten Modellen mittels χ^2 -Tests gegeneinander getestet werden.

Wenn also die Konsequenzen von Moderatoreffekten statistisch untersucht und dazu Devianz-Tests interpretiert werden sollen, ist es vorteilhaft, beide genannten ML-Schätzmethoden einzusetzen. Zudem empfiehlt es sich gerade zur Bestimmung der Signifikanz der Intercept-Varianz, eher Devianz- als Wald-Tests zu benutzen, welche eine größere Teststärke aufweisen und deshalb bei kleinen Fallzahlen von Vorteil sein können.¹⁴

¹² Bei dieser Darstellung folgen wir im Wesentlichen den Ausführungen von Hox 2010: 41, 215, 217.

¹³ Die Schätzergebnisse nach der RML-Methode entsprechen auch eher den Ergebnissen der klassischen Meta-Analyse (kMA).

¹⁴ Zum Prinzip und der Bedeutung der Teststärke vgl. Urban/Mayerl 2011: 138ff.

Mit Blick auf die Bestimmung der Signifikanzen einzelner Prädiktoreffekte ist zu beachten, dass bei kleinen Fallzahlen die Berechnung von Konfidenzintervallen unter Annahme der Standardnormalverteilung zu einer Unterschätzung der Irrtumswahrscheinlichkeit führen kann. Daher bietet sich für diese Signifikanztests die Verwendung der Student-t-Verteilung an, bei welcher die Freiheitsgrade nach "k-p-1" berechnet werden (k=Zahl der Studien; p=Zahl der Prädiktorvariablen im Modell). Im von uns verwendeten Statistikprogramm "HLM 7.0" ist dies der Standardtest für Regressionskoeffizienten (Hox 2010: 229).

Für die Überprüfung des Einflusses verschiedener Moderatorvariablen ist in der MEMA zu berücksichtigen, dass die dabei verwendeten Variablen oft hochgradig untereinander korrelieren. Das kann bei multivariaten Modellschätzungen zu Multikollinearitätsproblemen führen. Daher sollten in der MEMA möglichst sparsame Modelle geschätzt und nur die "bedeutsamsten" Moderatorvariablen bei der Gesamtschätzung verwendet werden (Hox 2010: 221). Ein Multikollinearitätstest lässt sich unter Verwendung der von Lipsey/Wilson (2001) vorgeschlagenen WLS-Methode und mittels Bestimmung von Toleranzwerten durchführen (z.B. unter Verwendung von SPSS; vgl. dazu Urban/Mayerl 2011: 231ff).

Berechnung der Studienvarianz¹⁵

Wie zuvor bereits erwähnt, ist ein Hauptunterschied zwischen einer klassischen Mehrebenenanalyse und einer speziellen Mehrebenenanalyse für eine Meta-Analyse, dass bei einer MEMA die Daten auf der ersten Ebene (Level-1 Variablen) nicht in roher Form vorliegen, sodass Informationen über statistische Kennwerte, die auf Ebene 1 liegen, aus den aggregierten Studienberichten, also aus Level-2 Kennwerten abgeleitet werden müssen. Das führt u.a. dazu, dass die Statistikprogramme, die für eine MEMA verwendet werden, die Varianz der abhängigen Variablen auf der Individualebene nicht automatisiert berechnen können. Folglich muss auch die Information über die Studienvarianz einer jeden Primärstudie der jeweiligen Software manuell (d.h. als Dateninput) zur Verfügung gestellt werden. Dazu ist eine individuelle Berechnung von jeder Studienvarianz notwendig (s.u.).

Die Bestimmung der Studienvarianz kann dadurch erfolgen, dass die Inverse der Ziehungsvarianz (der quadrierte Standardfehler), welche auch als die Präzision (engl.: "precision") bezeichnet wird, als Gewicht auf der ersten Ebene der Mehrebenenanalyse verwendet wird.¹⁶ Diese

¹⁵ Bei dieser Darstellung folgen wir im Wesentlichen den Ausführungen von Borenstein et al. 2009: 33.

¹⁶ Zudem muss in diesem Falle bei der Programmsteuerung von HLM die Varianz der abhängigen Variablen auf der niedrigsten Ebene auf einen Wert von 1,0 fixiert werden, damit das Programm auf diese Weise erfährt, dass die Varianz dieser Ebene nicht geschätzt, sondern bereitgestellt wird. Dazu muss im Datensatz für die erste Ebene die Variable zur Studienvarianz enthalten sein (hier: LOR_var). Diese kann dann in HLM unter dem Menüpunkt "andere Einstellungen" ("Other Settings") und dort unter "Einstellungen zur Schätzung" ("estimation settings") und dem Menübutton

Methode ist die am häufigsten eingesetzte Technik zur Durchführung einer MEMA. Sie wird auch im von uns benutzten EDV-Programm "HLM" eingesetzt (mehr dazu im folgenden Gliederungsabschnitt).

Exemplarisch wird nachfolgend erläutert, in welcher Weise die Studienvarianz, die Präzision und der Intraklassenkorrelationskoeffizient (ICC) berechnet werden können.

Zur Berechnung der Studienvarianz gehen wir von folgendem Beispiel aus: In einer Primärstudie ($j=1$) wären die in Tabelle 4 gezeigten Daten erhoben worden. Nach der Formel (1) (s.o.) ließe sich daraus ein Odds-Ratio als Effektstärkemaß berechnen. Das Odds-Ratio für die Zahlen in Tabelle 4 beträgt "2,1" bzw. " $\ln(2,1)=0,74$ ". Folglich ist die Chance, von Mitgliedern der Untersuchungsgruppe, eine pädosexuelle Viktimisierung erlebt zu haben, um das 2,1-Fache höher als die gleiche Chance von Mitgliedern der Kontrollgruppe.

Tabelle 4: Zweidimensionale Tabelle zur Berechnung der Studienvarianz

	pädosex. Vikt.: ja	pädosex. Vikt.: nein	N
pädosexuelle Delinquenten (UG)	46 (a)	136 (b)	n_1
adultsexuelle Delinquenten (KG)	5 (c)	31 (d)	n_2

Für die oben berechnete Effektstärke von 2,1 (in der Primärstudie $j=1$) lässt sich die Studienvarianz nach der folgenden Formel (6) ermitteln:

$$\sigma_j^2 = 1/a + 1/b + 1/c + 1/d \quad (6)$$

$$\sigma_1^2 = 1/46 + 1/136 + 1/5 + 1/31 = 0,26135$$

Somit weist die auf das Odds-Ratio bezogene Studienvarianz in Primärstudie $j=1$ einen Wert von 0,26 (gerundet) auf.

In einer MEMA ist diese Berechnung für jede berücksichtigte Studie durchzuführen und in einer Variable abzuspeichern.

Wird die Studienvarianz auf diese Weise berechnet, gehen auch die Fallzahlen der Primärstudien als Gewichte in die MEMA ein. Das heißt, dass die Meta-Analysen durch Verwendung der Studienvarianz nicht nur auf Grundlage der Anzahl berücksichtigter

"Gewichtung" ("Weighting") in einem separaten Menü per Dropdownbutton bei "Studienvarianz" bzw. "known Variance" ausgewählt werden. Wenn die Auswahl erfolgreich war, wird dies dadurch indiziert, dass die Level-1 Varianz nun auf 1.00000 fixiert ist. Danach kann in HLM die MEMA wie jede andere MEA vorgenommen werden. Vgl. dazu auch die Ausführungen zur praktischen Steuerung von HLM in Gliederungsabschnitt 2.5.

Primärstudien, sondern auch auf Grundlage der dabei insgesamt berücksichtigten Probandenzahl durchgeführt werden.

Für die Berechnung des Intraklassenkorrelationskoeffizienten (ICC) muss noch von allen Studienvarianzen aller Primärstudien der arithmetische Mittelwert gebildet und in der entsprechenden Formel berücksichtigt werden. Mehr dazu im folgenden Textabschnitt.

Berechnung von "Präzision" und "Intraklassenkorrelationskoeffizient" (ICC)

Wenn die Variable "LOR_var" (also die Studienvarianz, s.o.) gebildet wurde, kann auch die Variable zur Messung der Präzision (s.o.) erstellt werden. In SPSS erfolgt dies mittels der folgenden Syntax:

$$\text{COMPUTE precision} = 1/\text{LOR_var}.$$

Bildet man den Mittelwert der Variablen zur Studienvarianz "LOR_var" (bspw. in SPSS mittels: *FREQUENCIES LOR_var /STAT MEAN.*), kann unter Verwendung dieses Mittelwerts und der über die MEMA geschätzte Varianz für die zweite Ebene der Intraklassenkorrelationskoeffizient (ICC) berechnet werden. Das funktioniert unter Verwendung der folgenden Formel:

$$\rho = \sigma_u^2 / (\sigma_u^2 + \sigma_e^2) \quad (7)$$

Wenn z.B. für die Studienvarianz (= Varianz auf der ersten Ebene) ein Wert von " $\sigma_e^2 = 0,38$ " ermittelt wurde, und in einem MEMA-Nullmodell (s.o.) für die Varianz auf der zweiten Ebene ein Wert von " $\sigma_u^2 = 0,24$ " geschätzt wurde, ergibt sich nach Gleichung (7) ein ICC von:

$$\rho = 0,24 / (0,24 + 0,38) = 0,38662$$

Damit läge der ICC bei einem Wert von 0,39 (gerundet) und somit wären im analysierten Datensatz rund 40 Prozent der Effektstärke-Varianz auf Merkmale der zweiten Datenebene, also auf Studienmerkmale, zurückzuführen.

2.5 Programmtechnische Informationen zur MEMA mittels HLM

Die nachfolgenden Ausführungen veranschaulichen Schritt für Schritt die Durchführung einer Mehrebenen-Meta-Analyse unter Verwendung der Software HLM 7.0. Im ersten Teil dieser Beschreibung wird die Erstellung eines Datensatzes und im zweiten Teil die Spezifikation von Modellen sowie die Auswahl von Algorithmen und die "Gewichtung" der Beobachtungen (im Fall der MEMA also die Eingabe der Studienvarianz, vgl. Gliederungsabschnitt 2.4) erläutert.¹⁷

Erstellung eines HLM-Datensatzes

Für eine MEMA müssen zunächst die zu analysierenden Daten in HLM eingelesen werden. Dazu wird der Punkt "Make new MDM file" ausgewählt und im ausklappenden Drop-Down-Menü angegeben, ob der Datensatz im ASCII-Format oder dem Format eines Statistikprogramms ("Stat package input") vorliegt (standardmäßig ist dies SPSS) (siehe Abb.1).

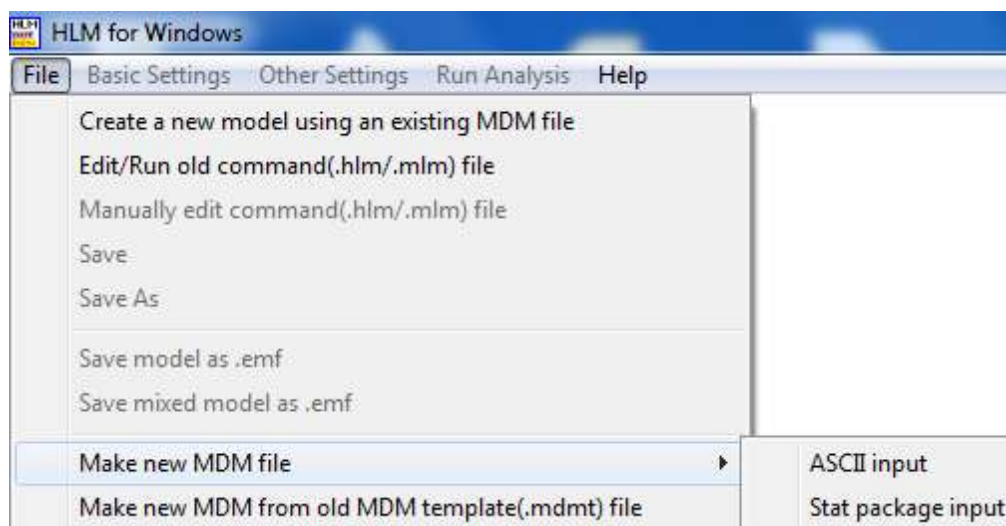


Abbildung 1

Nachdem diese Auswahl getroffen wurde, ist vor dem Einlesen der Daten noch die hierarchische Struktur dieser Daten anzugeben. Im Fall der MEMA handelt es sich um sogenannte "Nested Models" und in unserem Fall um Modelle mit zwei Ebenen. Daher ist der "MDM type" "HLM2" auszuwählen und diese Auswahl mit "OK" zu bestätigen (siehe Abb. 2).

Wurde diese Auswahl getroffen, kann der Datensatz bzw. können die Datensätze, welche die Informationen für die unterschiedlichen Ebenen enthalten, eingelesen werden. Dabei kann ein und derselbe Datensatz die Informationen zu beiden Ebenen enthalten. Ist das der Fall, muss dieser

¹⁷ Weitere Informationen zu HLM finden sich in Raudenbush et al. (2004) und Raudenbush/Bryk (2002) sowie auf der Homepage von SSI-Central: <http://www.ssicentral.com/hlm/examples.html>.

Datensatz einmal als Level-1 File und einmal als Level-2 File eingelesen werden. Dazu ist jeweils der "Browse"-Button anzuklicken (siehe Abb. 3).

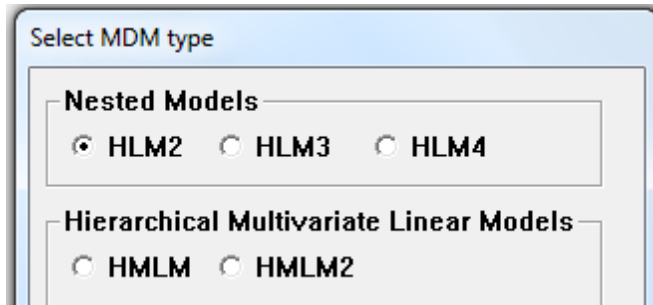


Abbildung 2

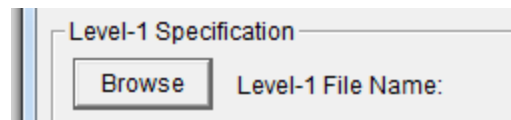
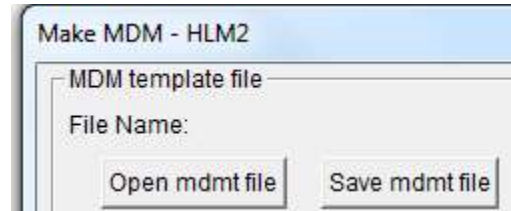


Abbildung 3

Danach kann der betreffende Datensatz ausgewählt und in den Speicher eingelesen werden (hier handelt es sich um den Datensatz "MAPS_N26") (siehe Abb. 4). Wurden die Datensätze (oder wie hier der gleiche Datensatz zweimal) eingelesen, können die Variablen für die unterschiedlichen Ebenen ausgewählt werden. Dazu ist der Button "Choose Variables" anzuklicken (siehe Abb. 5, rechts unten). In dem sich öffnenden Fenster werden alle Variablen angezeigt, die im Datensatz enthalten sind.

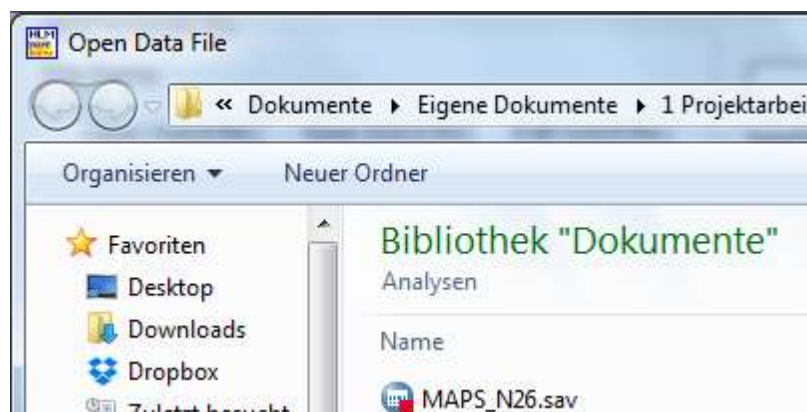


Abbildung 4

Make MDM - HLM2

MDM template file

File Name:

Open mdmt file Save mdmt file Edit mdmt file

MDM File Name (use .mdm suffix)

Input File Type SPSS/Windows

Structure of Data - this affects the notation only!

cross sectional (persons within groups) measures within groups

longitudinal (occasions within persons)

Level-1 Specification

Browse Level-1 File Name: F:\Users\Uoachim Fiebig\Documents\1 Projektarbei Choose Variables

Abbildung 5

Unabhängig von der Ebene, die ausgewählt wurde, muss die Variable, welche die Identifikation der zweiten bzw. höchsten Ebene ermöglicht, mittels Klick als "ID" ausgewiesen werden. Dazu wird neben der betreffenden Variablen (hier "ID") der Haken in das betreffende Kästchen gesetzt. Alle Haken, die neben den Variablen in die Kästchen mit der Abkürzung "MDM" ("Multivariate Data Matrix") gesetzt werden, weisen diese als Variablen der betreffenden Ebene aus (hier, der Ebene 1; z.B. "LOR_ESVA" als die Studienvarianz, die auf der Individualebene vorliegt) (siehe Abb. 6, linker Kasten). Die gleiche Prozedur wird mit dem Datensatz zur Ebene 2 wiederholt, wobei hier dann eben die Variablen der Ebene 2 ausgewählt werden, wie etwa die Fallzahl oder die methodische Güte (siehe Abb. 6, rechter Kasten).

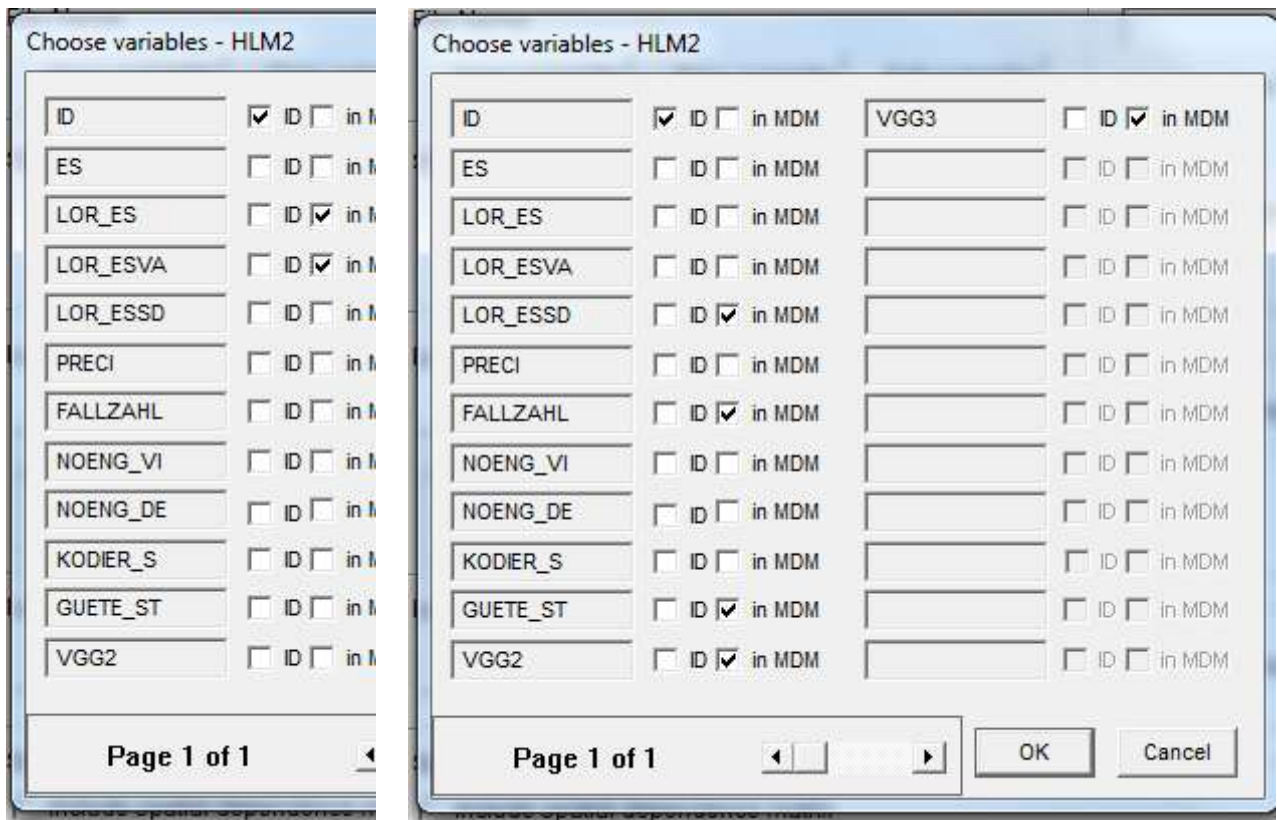


Abbildung 6

Wurden alle Variablen ausgewählt, muss die in diesem Vorgang erstellte multivariate Datenmatrix abgespeichert werden. Dazu ist zunächst ein Name für diese zu vergeben, wofür in das freie Feld unter "MDM File Name (use .mdm suffix)" zu klicken ist (siehe Abb. 7). Dieser Schritt ist nicht gesondert zu bestätigen, etwa durch den Klick auf einen weiteren Button. Im Anschluss lässt sich die MDM-Template-File ("*.mdmt") nach Klick auf "Save mdmt file" in einem gewählten Ordner und mit gewähltem Namen abspeichern (siehe Abb. 7).



Abbildung 7

Ist auch dieser Schritt erfolgt, kann die multivariate Datenmatrix erstellt werden. Hierfür ist ein Klick auf den Button "Make MDM" nötig (siehe Abb. 8, links unten). Wenn beim verwendeten Datensatz "listwise-deletion" durchgeführt wurde (was im Fall von MEMA eine durchaus

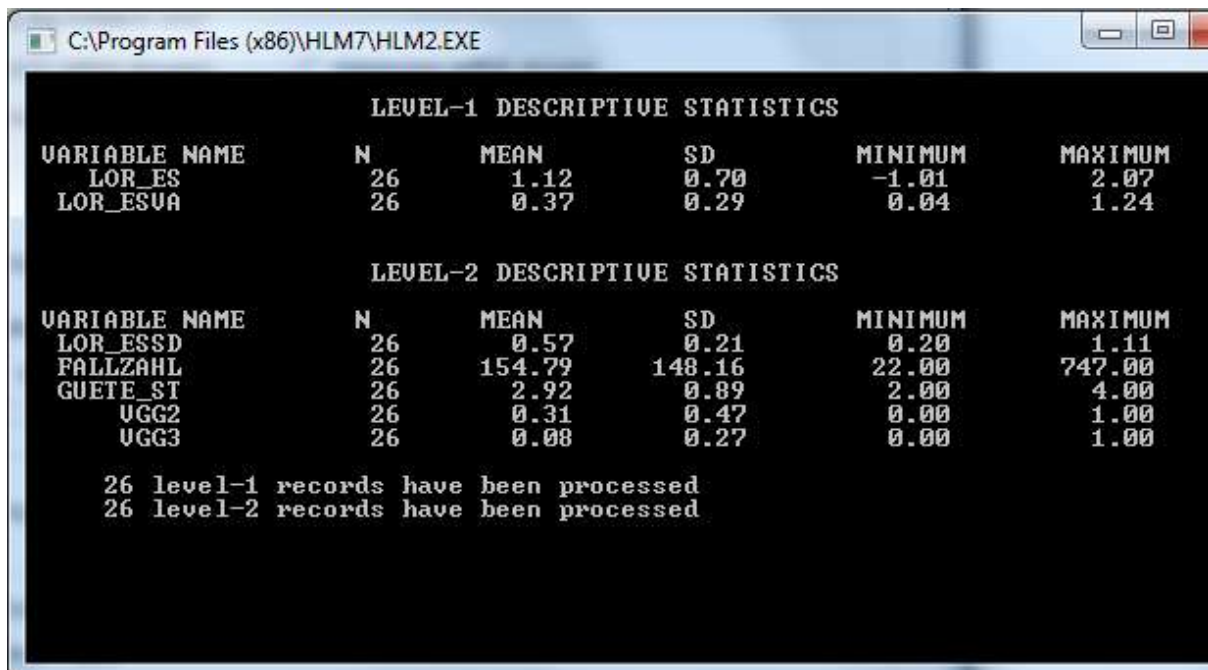
angemessene Vorgehensweise sein kann), können dazu alle übrigen Einstellungen auf den voreingestellten Werten bleiben.

Abbildung 8

Wurde die Datenmatrix erstellt, was durch das Erscheinen eines schwarzen DOS-Fensters indiziert wird (siehe Abb. 9), müssen die deskriptiven Statistiken erneut und dieses Mal in einer *.txt Datei, die in einem Editor-Fenster angezeigt wird, nach dem Klick auf den Button "Check Stats" überprüft werden (siehe Abb. 8 bzw. 10). In beiden Fenstern sind dabei Informationen zu den Variablen der unterschiedlichen Ebenen enthalten. Bei diesen handelt es sich um Angaben zur Fallzahl, dem Mittelwert, der Standardabweichung sowie zur Minimal- und zur Maximalausprägung der betreffenden Variablen. Im Editorfenster werden darüber hinaus die Namen der "mdm-" und der "mdmt-Dateien" sowie das Bearbeitungsdatum angezeigt.

Erst nach dieser Prüfung kann mit einem Klick auf den neben dem "Make MDM" befindlichen Button "Done" die Erstellung des Datensatzes abgeschlossen werden (siehe Abb. 8).

Das Programm wechselt gleichzeitig in die Anzeige der Modellierungsoptionen. Im nachfolgenden Abschnitt wird erläutert, wie die Modellspezifizierung und die Einstellung der Schätzalgorithmen etc. vorgenommen werden.



C:\Program Files (x86)\HLM7\HLM2.EXE

LEVEL-1 DESCRIPTIVE STATISTICS

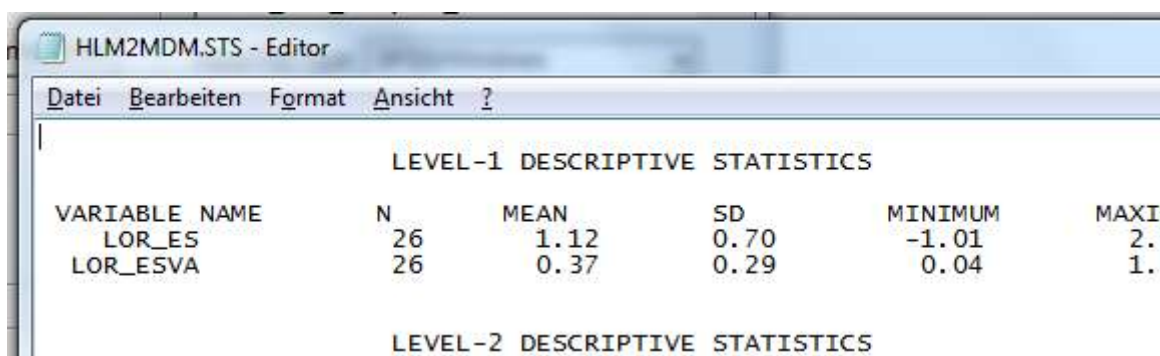
VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
LOR_ES	26	1.12	0.70	-1.01	2.07
LOR_ESVA	26	0.37	0.29	0.04	1.24

LEVEL-2 DESCRIPTIVE STATISTICS

VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
LOR_ESSD	26	0.57	0.21	0.20	1.11
FALLZAHL	26	154.79	148.16	22.00	747.00
GUETE_ST	26	2.92	0.89	2.00	4.00
UGG2	26	0.31	0.47	0.00	1.00
UGG3	26	0.08	0.27	0.00	1.00

26 level-1 records have been processed
26 level-2 records have been processed

Abbildung 9



HLM2MDM.STS - Editor

Datei Bearbeiten Format Ansicht ?

LEVEL-1 DESCRIPTIVE STATISTICS

VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXI
LOR_ES	26	1.12	0.70	-1.01	2.
LOR_ESVA	26	0.37	0.29	0.04	1.

LEVEL-2 DESCRIPTIVE STATISTICS

Abbildung 10

Modellierung, Algorithmen und Gewichtung

Der Bildschirm mit den Modellierungsoptionen ist in zwei Fenster unterteilt: in ein schmales, vertikales Band auf der linken Seite, welches die abhängige und die unabhängige/n Variable/n der unterschiedlichen Ebenen zeigt, und in ein breites Fenster auf der rechten Seite, in welchem die Regressionsgleichungen für die unterschiedlichen Ebenen abgebildet werden, sobald das Modell

spezifiziert wurde (siehe Abb. 11 bis 14). Die Buttons "Level 1" und "Level 2" erlauben das Hin- und Herschalten zwischen den Ansichten der Variablen bzgl. der betreffenden Ebenen.

Um mit der Modellierung zu beginnen, ist als erstes die abhängige Variable (AV) zu bestimmen. Im Fall unserer Modelle ist dies immer das logarithmierte Odds Ratio bzw. die logarithmierte Effektstärke (hier: "LOR_ES"). Um diese als AV zu deklarieren, muss mit einem Rechtsklick auf die betreffende Variable das Kontextmenü geöffnet und die Option "Outcome variable" ausgewählt werden (wenn noch keine Variable als AV bestimmt wurde, sind die übrigen Optionen nicht verfügbar und daher ausgegraut) (siehe Abb. 12).

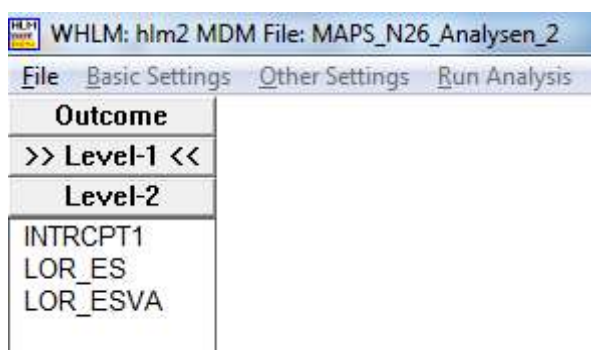


Abbildung 11

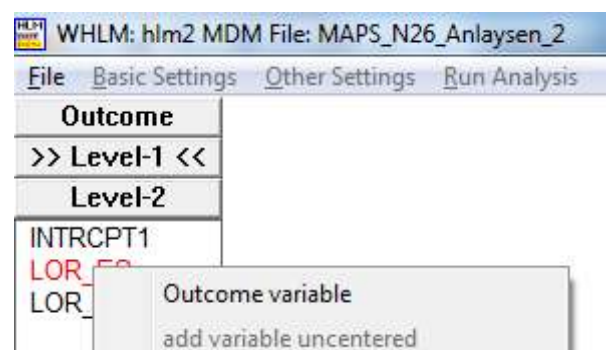


Abbildung 12

Wurde eine AV deklariert, dann erscheinen die Gleichungen für die beiden Ebenen auf der rechten Bildschirmfläche (siehe Abb. 13). Danach lassen sich die Modelle spezifizieren, indem nach und nach die interessierenden unabhängigen Variablen (UV) dem Modell hinzugefügt werden. Dies funktioniert wie bei der AV über das Öffnen des Kontextmenüs mittels Rechtsklick und einer entsprechenden Auswahl.

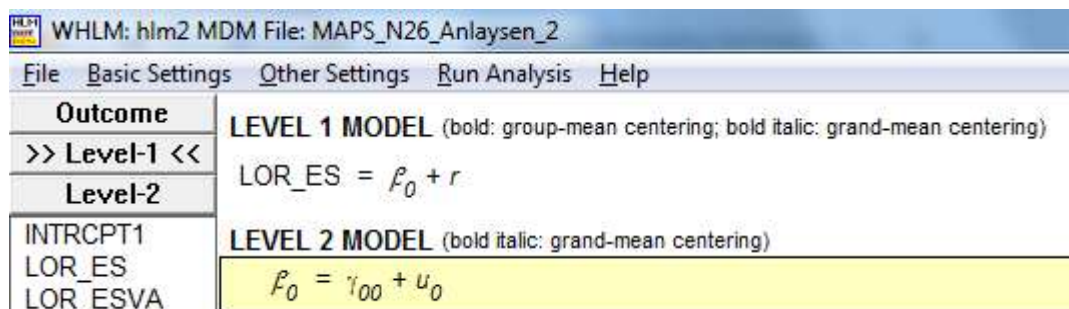


Abbildung 13

Wenn kein bestimmtes analytisches Forschungsinteresse vorliegt, empfiehlt es sich, die Variablen

in der rohen Form einzufügen ("add variable uncentered"). Unsere Abbildung 14 zeigt das vollständige Modell P4a mit den UV "Standardfehler der Effektstärke" (SE-ES), "Fallzahl" (Fallzahl), "methodische Studiengüte" (GUETE_ST) und die Kontrollvariablen zu den untersuchten Vergleichsgruppen, wobei hier als Referenz diejenigen Studien dienen, die adultsexuelle Straftäter als Kontrollgruppe benutzen. Damit weisen die Kontrollvariablen diejenigen Studien aus, die pädosexuelle mit "nichtsexuellen Straftätern" (VG2) bzw. mit "Nichtstraftätern" (VG3) vergleichen.

WHLM: hlm2 MDM File: MAPS_N26_Analysen_2	
File Basic Settings Other Settings Run Analysis Help	
Outcome	LEVEL 1 MODEL (bold: group-mean centering; bold italic: grand-mean centering)
Level-1	LOR_ES = β_0
>> Level-2 <<	$\sigma^2 = \text{LOR_ESVA}$ (Known)
INTRCPT2	LEVEL 2 MODEL (bold italic: grand-mean centering)
LOR_ESSD	$\beta_0 = \gamma_{00} + \gamma_{01}(\text{LOR_ESSD}) + \gamma_{02}(\text{FALLZAHL}) + \gamma_{03}(\text{GUETE_ST}) + \gamma_{04}(\text{VGG2}) + \gamma_{05}(\text{VGG3}) + u_0$
FALLZAHL	
GUETE_ST	
VGG2	
VGG3	

Abbildung 14

Bevor die Modellschätzungen durchgeführt werden können, müssen zunächst noch Anpassungen für das MEMA-Design sowie eine formale Änderung für die Ausspeicherung des Dateiformats für das Ergebnisfenster vorgenommen werden.

Beim Dateiformat empfiehlt sich, statt dem defaultmäßig eingestellten HTML-Format das TEXT-Format zu verwenden, da hiermit eine Konvertierung in andere Formate, die eine automatisierte Verwendung der vorhandenen Daten wie etwa *.xml oder *.xlsx, erleichtern, einfacher vorgenommen werden können. Um das Dateiformat des Ergebnisoutputs umzustellen, muss in der Menüleiste am oberen Rand des Fensters die Option "Basic Settings" angeklickt werden. Beim sich daraufhin öffnenden Fenster kann der Dateiname in die Zeile rechts neben der Bezeichnung "Output file name" geschrieben und die betreffende Dateierdung von *.html in *.txt geändert werden (siehe Abb. 15).

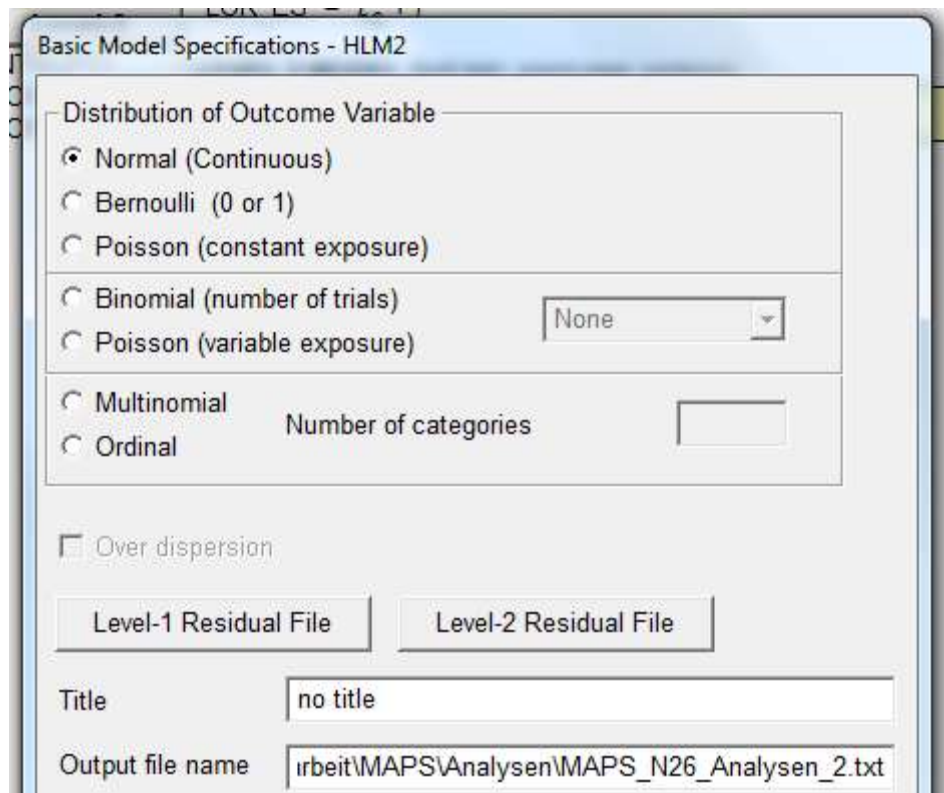


Abbildung 15

Um nun die MEMA-Modellschätzungen vornehmen zu können, muss dem Programm noch die Studienvarianz zur Verfügung gestellt werden (vgl. Abb. 10). Dazu ist der Punkt "Other Settings" aus der Menüzeile auszuwählen. Ein Klick öffnet ein Drop-Down-Menü. Beim Unterpunkt "Estimation Settings" können die betreffenden Einstellungen vorgenommen werden (siehe Abb. 16). Mit dessen Auswahl öffnet sich ein weiteres Fenster. Bei diesem lässt sich bei Bedarf auch der Algorithmus von "Restricted maximum likelihood"(RML) auf "Full maximum likelihood" (FML) umstellen (vgl. dazu die Ausführungen in Gliederungsabschnitt 2.4).

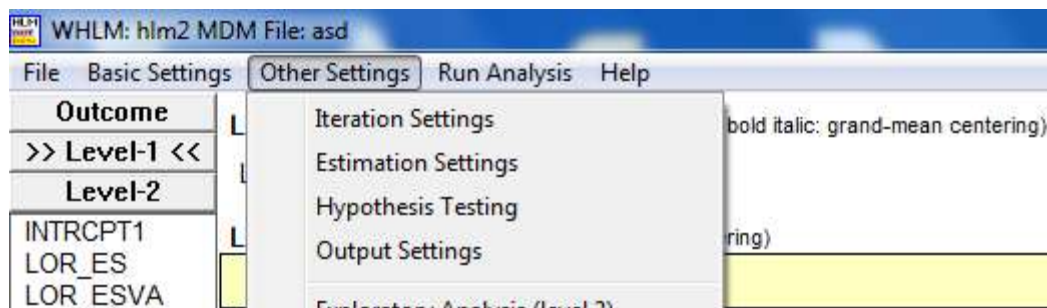


Abbildung 16

Für die Bereitstellung der Studienvarianz ist der etwas irreführende Button "Weighting"

anzuklicken (siehe Abb. 17), was dazu führt, dass sich ein weiteres Fenster öffnet. Bei diesem ist das unter der Bezeichnung "Known variance" befindliche Drop-Down-Menü zu öffnen und diejenige Variable auszuwählen, welche die Information zur Studienvarianz enthält (hier "LOR_ESVA", für die Varianz der logarithmierten Effektstärken) (siehe Abb. 18). Die Auswahl ist danach mit "Ok" zu bestätigen. War dies erfolgreich, sollte sich der Wert rechts neben "Fix sigma² to specific value" von "computed" zu "1.000000" verändert haben (vgl. Abb. 17 und 19). Das bedeutet, dass die Varianz auf Ebene 1 nicht durch HLM geschätzt wird, da sie nun bekannt ist.

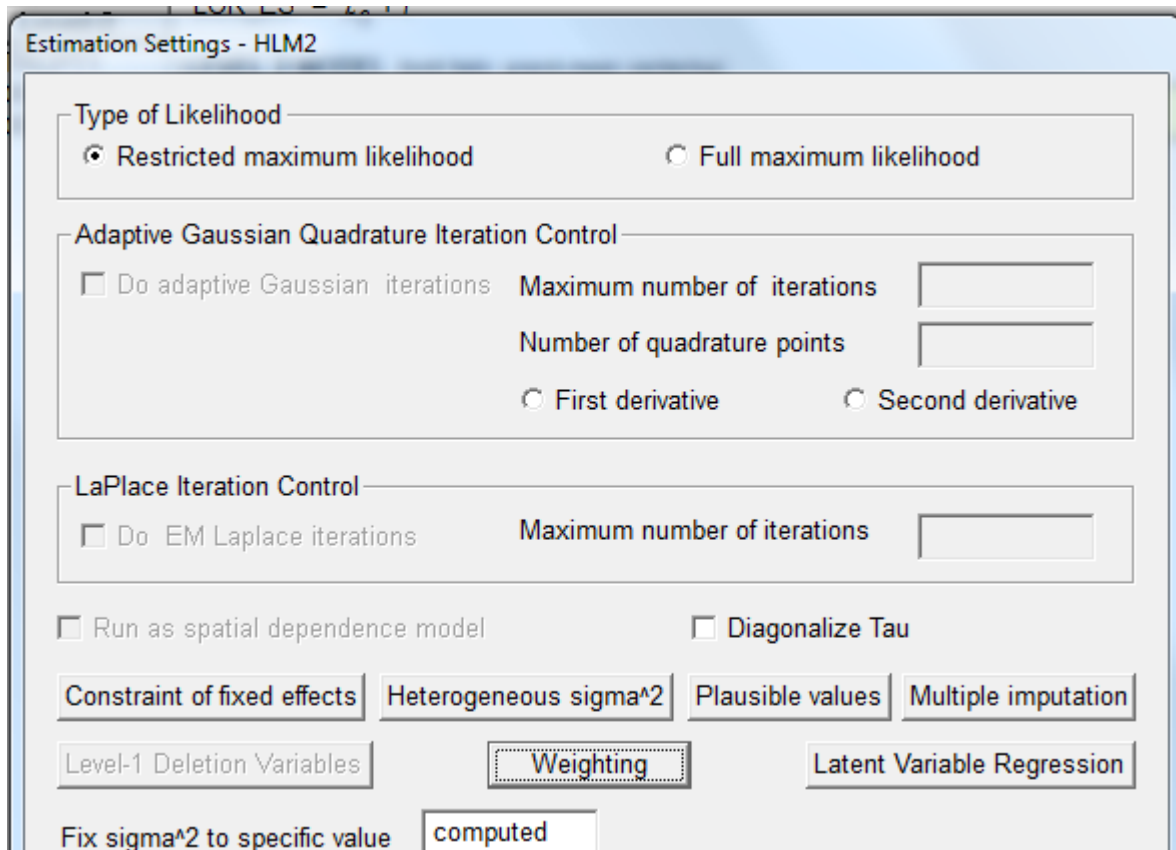


Abbildung 17

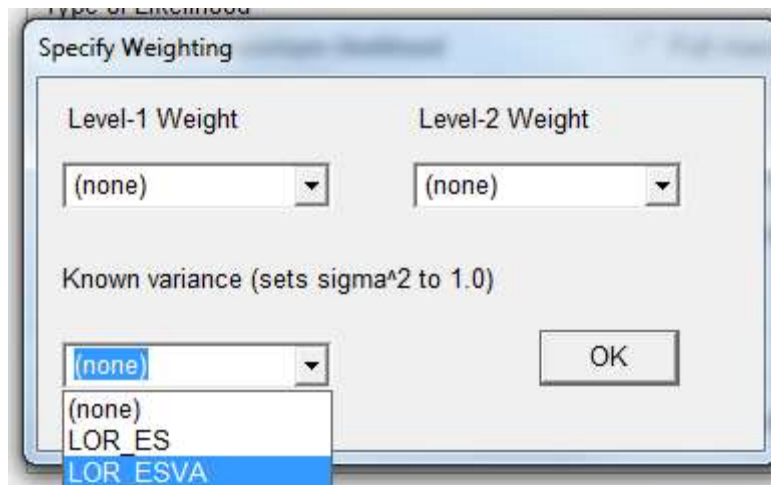


Abbildung 18

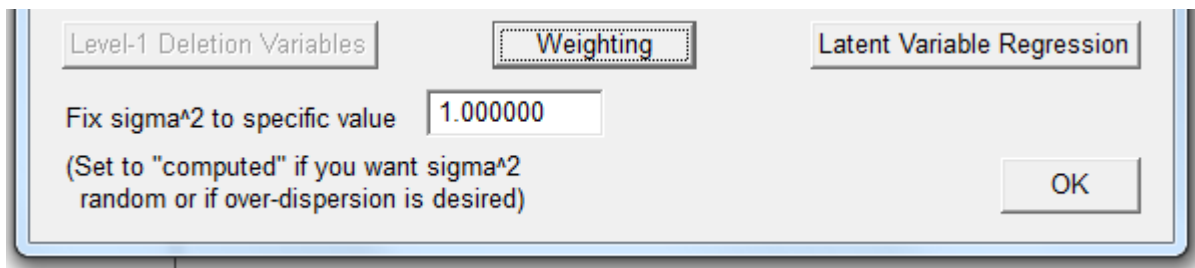


Abbildung 19

Anschließend kann die Modellschätzung mittels Klick auf "Run Analysis" in der oberen Menüzeile (vgl. z.B. Abb. 16) vorgenommen werden. Daraufhin öffnet sich wieder ein neues Fenster, bei welchem schließlich nur noch der Button "Run the model shown" anzuklicken ist (siehe Abb. 20). Ist dieser Klick gesetzt, öffnet sich erneut ein DOS-Fenster (siehe Abb. 21) und die Iterationen laufen entweder bis zur erfolgreichen Konvergenz der Schätzung oder bis zum Erreichen einer im Voraus festgelegten Iterationszahl (defaultmäßig ist das 100, was unter → "Other Settings" → "Iteration Settings" → "Number of (micro) iterations" verändert werden kann; vgl. Abb. 16).

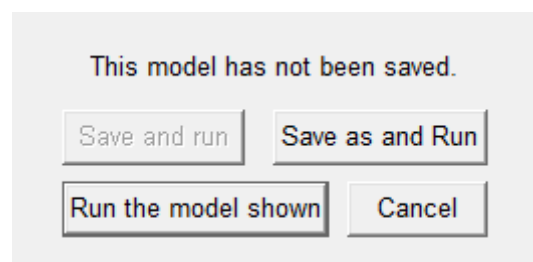
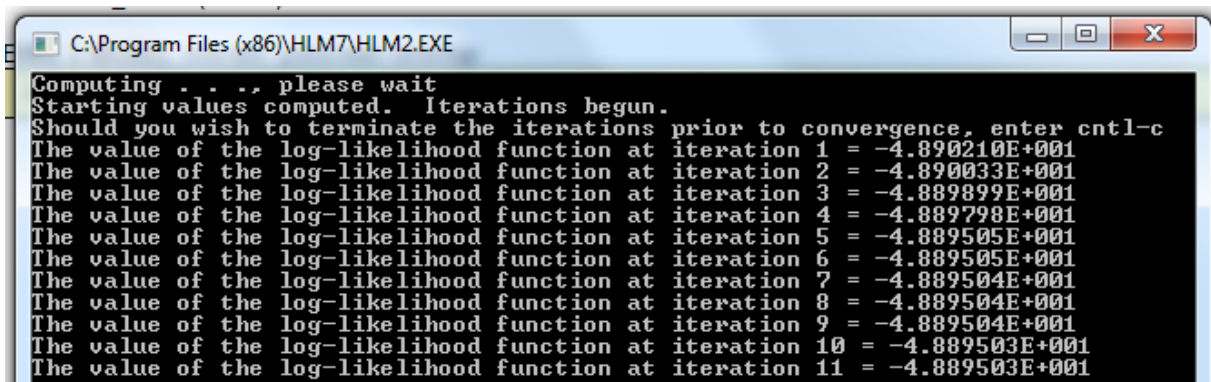


Abbildung 20

Wenn die Modellschätzung konvergiert, schließt sich nach einem kurzen Moment zunächst das DOS-Fenster. Anschließend öffnet sich das Output-Fenster. Wurde das Format des Outputs nicht vom HTML- auf das TXT-Format geändert, öffnet sich der Output im als Standard definierten Web-Browser. Für den Fall, dass die Dateieindung umgestellt wurde, öffnet sich ein neues Editor-Fenster mit der Output-Datei, welche Angaben zum Datum, der Fallzahl, der AV, der verwendeten "Gewichtungsvariable", dem spezifizierten Modell in Gleichungsform, dem ICC, den Reliabilitätskoeffizienten, den Ergebnissen der Modellschätzungen und viele weitere Informationen enthält (die Abbildung 22 zeigt dies nur auszugsweise, da hier ansonsten eine Vielzahl von Screenshots abgebildet werden müsste).



```
C:\Program Files (x86)\HLM7\HLM2.EXE
Computing . . . , please wait
Starting values computed. Iterations begun.
Should you wish to terminate the iterations prior to convergence, enter cntl-c
The value of the log-likelihood function at iteration 1 = -4.890210E+001
The value of the log-likelihood function at iteration 2 = -4.890033E+001
The value of the log-likelihood function at iteration 3 = -4.889899E+001
The value of the log-likelihood function at iteration 4 = -4.889798E+001
The value of the log-likelihood function at iteration 5 = -4.889505E+001
The value of the log-likelihood function at iteration 6 = -4.889505E+001
The value of the log-likelihood function at iteration 7 = -4.889504E+001
The value of the log-likelihood function at iteration 8 = -4.889504E+001
The value of the log-likelihood function at iteration 9 = -4.889504E+001
The value of the log-likelihood function at iteration 10 = -4.889503E+001
The value of the log-likelihood function at iteration 11 = -4.889503E+001
```

Abbildung 21

Um weitere Modellschätzungen mit veränderten Einstellungen vorzunehmen, sind die hier aufgezeigten Schritte ab Beginn des Textabschnitts "Modellierung, Algorithmen und Gewichtung" erneut zu durchlaufen und entsprechend anzupassen.

```

MAPS_N26_Analysen_2.txt - Editor
Datei Bearbeiten Format Ansicht ?

Program:                HLM 7 Hierarchical Linear and Nonlinear Modeling
Authors:                Stephen Raudenbush, Tony Bryk, & Richard Congdon
Publisher:              Scientific Software International, Inc. (c) 2000
                        techsupport@ssicentral.com
                        www.ssicentral.com

-----
Module:                HLM2.EXE (7.00.21106.30001)
Date:                  21 August 2014, Thursday
Time:                  17:24:14
-----

Specifications for this HLM2 run

Problem Title: no title

The data source for this run = MAPS_N26_Analysen_2
The command file for this run = C:\Users\JOACHI~1\AppData\Local\Temp\whlmtmp.hl
Output file name       = F:\Users\Joachim Fiebig\Documents\1 Projektarbei
The maximum number of level-1 units = 26
The maximum number of level-2 units = 26
The maximum number of iterations = 100

Method of estimation: restricted maximum likelihood

weighting specification
-----
Level 1      weighting?      weight
Level 2      weighting?      variable
Precision    yes                Name           Normalized?

Level 1      no                LOR_ESVA       no
Level 2      no
Precision    yes

The outcome variable is  LOR_ES

summary of the model specified (in hierarchical format)
-----
Level-1 Model
      LOR_ES = B0 + r

Level-2 Model
      B0 = G00 + G01*(LOR_ESSD) + G02*(FALLZAHL) + G03*(GUETE_ST) + G04*(VGG2)
          + G05*(VGG3) + u0

```

Abbildung 22

Literaturverzeichnis

- Borenstein, M., Hedges, L.V., Higgins, J.P.T., Rothstein, H.R., 2009: Introduction to meta-analysis. Chichester: Wiley.
- Ditton, H., 1998: Mehrebenenanalyse – Grundlagen und Anwendungen des Hierarchisch Linearen Modells. Weinheim/München: Juventa.
- Fiebig, J., Rüdebusch, G., Urban, D., 2012: Meta-Analyse und "publication bias". Eine empirisch-statistische Fallstudie. Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart (SISS), 2012/1.
- Hall, J.A., Rosenthal, R., 1995: Interpreting and evaluating meta-analysis. *Evaluation & the Health Professions* 18: 393-407.
- Hedges, L.V., 1994: Fixed effect models. S. 285-299 in: Cooper, H., Hedges, L.V. (Hrsg.), *The handbook of research synthesis*. New York: Sage.
- Hedges, L.V., Olkin, I., 1985: *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.
- Hox, J.J., 2010: *Multilevel analysis: techniques and applications* (2nd edition). New York: Routledge.
- Hunter, J.E., Schmidt, F.L., 2004: *Methods of meta-analysis. Correcting for bias and error in research findings* (2nd Edition). Thousand Oaks: Sage.
- Lipsey, M.W., Wilson, D.T., 2001: *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Littell, J.H., Corcoran, J., Pillai, V., 2008: *Systematic reviews and meta-analysis*. Oxford: University Press.
- Maas, C.J.M., Hox, J.J., 2004: Robustness issues in multilevel regression analysis. *Statistica Neerlandica* 58: 127-137.
- Pötschke, M., 2006: Mehrebenenanalyse. S. 167-179 in: Behnke, J., Gschwend, T., Schindler, D., Schnapp, K. (Hrsg.), *Methoden der Politikwissenschaft. Neuere qualitative und quantitative Analyseverfahren*. Baden-Baden: Nomos.
- Raudenbush, S.W., Bryk, A.S., 2002: *Hierarchical linear models. Applications and data analysis methods* (2. Auflage). Thousand Oaks: Sage.
- Raudenbush, S.W., Bryk, A.S. & Congdon, R. 2004: *HLM 6 for Windows* [Computer software]. Skokie, IL: Scientific Software International, Inc.
- Rosenthal, R., DiMatteo, M.R., 2001: Meta-analysis: recent developments in quantitative methods for literature reviews. *Annual Reviews of Psychology* 25: 59-82.
- Schulze, R. (2008). *Meta-analysis, a comparison of approaches*. Göttingen: Hogrefe & Huber.
- Shadish, W.R., Haddock, C.K., 2009: Combining estimates of effect size. S. 257-276 in: Cooper, H., Hedges, L.V., Valentine, J.C. (Hrsg.), *The handbook of research synthesis and meta-analysis* (2nd edition). New York: Russell Sage.

- Urban, D., Fiebig, J., 2010: Wird pädosexuelles Verhalten in jungen Jahren erlernt? Eine empirische Fall-Kontroll-Studie. *Soziale Probleme* 21: 226-248.
- Urban, D., Fiebig, J., 2011: Pädosexueller Missbrauch: Wenn Opfer zu Tätern werden. Eine empirische Studie. *Zeitschrift für Soziologie* 40: 42-61.
- Urban, D., Mayerl, J., 2011: *Regressionsanalyse: Theorie, Technik und Anwendung* (4. überarbeitete Auflage). Wiesbaden: VS-Verlag.

SISS: Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart
<http://www.uni-stuttgart.de/soz/institut/forschung/siss.html>

bisher sind erschienen:

- No. 1/1994 "Vertrauen" - soziologisch betrachtet. Ein Beitrag zur Analyse binärer Interaktionssysteme.
Peter Antfang, Dieter Urban
- No. 2/1994 Report on the German Machine Tool Industry.
Frank C. Englmann, Christian Heyd, Daniel Köstler, Peter Paustian
with the assistance of Susanne Baur and Peter Bergmann
- No. 3/1994 Neue württembergische Rechtstatsachen zum Unternehmens- und Gesellschaftsrecht.
Udo Kornblum
- No. 4/1994 Rechtstatsachen zum Unternehmens- und Gesellschaftsrecht aus den neuen Bundesländern.
Udo Kornblum
- No. 1/1995 Die Bedeutung Neuronaler Netze in der Ökonomie.
Hermann Schnabl
- No. 2/1995 Regionale Strukturprobleme.
Sammelband der Beiträge zum Symposium vom 13. und 14. Oktober 1994.
Frank C. Englmann (Hrsg.)
- No. 3/1995 Latent Attitude Structures Directing the Perception of New Technologies.
An Application of SEM-Methodology to the Construction of Attitude
Measurement Models Related to Technologies of Prenatal Genetic
Engineering and Testing.
Dieter Urban
- No. 4/1995 Handbuch zur empirischen Erhebung von Einstellungen/Kognitionen zur Bio- und Gentechnologie (inklusive Diskette)
(zweite, überarbeitete und erweiterte Auflage)
Uwe Pfenning, Dieter Urban, Volker Weiss
- No. 5/1995 Social Indicators in a Nonmetropolitan County: Testing the Representativeness of a Regional Nonrandom Survey in Eastern Germany.
Dieter Urban, Joachim Singelmann
- No. 1/1996 Jugend und Politik im Transformationsprozeß. Eine Fallstudie zur Stabilität und Veränderung von politischen Einstellungen bei ostdeutschen Jugendlichen zwischen 1992 und 1995.
Dieter Urban, Joachim Singelmann, Helmut Schröder
- No. 2/1996 Einstellungsmessung oder Einstellungsgenerierung? Die Bedeutung der informationellen Basis bei Befragten für die empirische Rekonstruktion von Einstellungen zu gentechnischen Anwendungen.
Martin Slaby
- No. 1/1997 Gentechnik: „Fluch oder Segen“ versus „Fluch und Segen“.
Bilanzierende und differenzierende Bewertungen der Gentechnik in der öffentlichen Meinung.
Dieter Urban und Uwe Pfenning

(Fortsetzung ...)

SISS: Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart
<http://www.uni-stuttgart.de/soz/institut/forschung/siss.html>

bisher sind erschienen (Fortsetzung):

- No.2/1997 Die soziale Vererbung von Ausländer"feindlichkeit". Eine empirische Längsschnittanalyse der intra- und intergenerativen Transmission von sozialen Einstellungen.
Dieter Urban und Joachim Singelmann
- No. 3/1997 Politische Sozialisation im Transformationsprozeß: Die Entwicklung demokratiebezogener Einstellungen von ostdeutschen Jugendlichen und deren Eltern zwischen 1992 und 1996.
Barbara Schmidt, Dieter Urban, Joachim Singelmann
- No.1/1998 Bewertende Einstellungen zur Gentechnik: ihre Form, ihre Inhalte und ihre Dynamik. Kurzbericht zu Ergebnissen des Forschungsprojektes „Einstellungen zur Gentechnik“.
Dieter Urban, Uwe Pfenning, Joachim Allhoff
- No.2/1998 Technikeinstellungen: gibt es die überhaupt? Ergebnisse einer Längsschnittanalyse von Bewertungen der Gentechnik.
Dieter Urban
- No.3/1998 Zur Interaktion zwischen Befragten und Erhebungsinstrument. Eine Untersuchung zur Konstanz des Meinungsurteils von Befragten im Interviewverlauf.
Martin Slaby
- No.1/1999 Role Models and Trust in Socio-Political Institutions: A Case Study in Eastern Germany, 1992-96.
Joachim Singelmann, Toby A. Ten Ayck, Dieter Urban
- No.1/2000 Die Zufriedenheit von Stuttgarter Studierenden mit ihrer Lebens- und Wohnsituation. Erste deskriptive Ergebnisse einer sozialwissenschaftlichen Studie zu allgemeinen und bereichsspezifischen Zufriedenheiten der Studierenden des Campus Vaihingen und des Campus Hohenheim.
Projektgruppe Campus: Slaby, M.; Grund, R.; Mayerl, J.; Noak, T.; Payk, B.; Sellke, P.; Urban, D.; Zudrell, I.
- No.2/2000 Längsschnittanalysen mit latenten Wachstumskurvenmodellen in der politischen Sozialisationsforschung.
Dieter Urban
- No.1/2001 Unser „wir“ - ein systemtheoretisches Modell von Gruppenidentitäten.
Jan A. Fuhse
- No.2/2001 Differentielle Technikakzeptanz, oder: Nicht immer führt die Ablehnung einer Technik auch zur Ablehnung ihrer Anwendungen.
Eine nutzentheoretische und modell-statistische Analyse.
Martin Slaby, Dieter Urban
- No.3/2001 Religiosität und Profession. Longitudinale Analysen zur Entwicklung des religiösen Selbstbildes bei Erzieherinnen.
Heiko Lindhorst

(Fortsetzung ...)

bisher sind erschienen (Fortsetzung):

- No.4/2001 Ist Glück ein affektiver Sozialindikator subjektiven Wohlbefindens?
Dimensionen des subjektiven Wohlbefindens und die Differenz
zwischen Glück und Zufriedenheit.
Jochen Mayerl
- No.1/2002 Risikoakzeptanz als individuelle Entscheidung.
Zur Integration der Risikoanalyse in die nutzentheoretische
Entscheidungs- und Einstellungsforschung.
Martin Slaby, Dieter Urban
- No.2/2002 Vertrauen und Risikoakzeptanz. Zur Relevanz von Vertrauen
bei der Bewertung neuer Technologien.
Martin Slaby, Dieter Urban
- No.3/2002 Probleme bei der Messung individueller Veränderungsraten.
13 empirisch und methodisch induzierte Effekte, die es schwierig machen,
Veränderungen von generalisierten Bewertungen zu ermitteln.
Dieter Urban
- No.1/2003 Systeme, Netzwerke, Identitäten. Die Konstitution sozialer Grenzziehungen
am Beispiel amerikanischer Straßengangs.
Jan A. Fuhse
- No.2/2003 Können Nonattitudes durch die Messung von Antwortreaktionszeiten ermittelt werden?
Eine empirische Analyse computergestützter Telefoninterviews.
Jochen Mayerl
- No.1/2004 Erhöht ein Opfer-Täter-Zyklus das Risiko, Sexualstraftaten als pädosexuelle Straftaten zu
begehen? Ergebnisse einer ereignisanalytischen Pilotstudie
Dieter Urban, Heiko Lindhorst
- No.1/2005 Persönliche Netzwerke in der Systemtheorie
Jan A. Fuhse
- No.2/2005 Analyzing cognitive processes in CATI-Surveys with response latencies:
An empirical evaluation of the consequences of using different
baseline speed measures.
Jochen Mayerl, Piet Sellke, Dieter Urban
- No.1/2006 Ist Bildung gleich Bildung? Der Einfluss von Schulbildung auf ausländerablehnende
Einstellungen in verschiedenen Alterskohorten.
Klaus Hadwiger
- No.2/2006 Zur soziologischen Erklärung individuellen Geldspendens.
Eine Anwendung und Erweiterung der Theory of Reasoned Action unter Verwendung von
Antwortlatenzzeiten in einem Mediator-Moderator-Design.
Jochen Mayerl
- No.1/2007 Antwortlatenzzeiten in TRA-Modellen. Zur statistischen Erklärung von (Geld)-
Spendenverhalten.
Dieter Urban, Jochen Mayerl

SISS: Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart
<http://www.uni-stuttgart.de/soz/institut/forschung/siss.html>

bisher sind erschienen (Fortsetzung):

- No.1/2008 Berufseintritt und Berufssituation von Soziologieabsolventen der Universität Stuttgart.
Deskriptive Ergebnisse einer Absolventenbefragung aus dem Jahr 2007.
Jochen Mayerl, Dieter Urban
- No.1/2010 Der Bystander-Effekt in alltäglichen Hilfsituationen:
Ein nicht-reaktives Feldexperiment.
Katrin Alle, Jochen Mayerl
- No.2/2010 Das Working-Poor-Problem in Deutschland.
Empirische Analysen zu den Ursachen von Armut trotz Erwerbstätigkeit.
Leonie Hellmuth, Dieter Urban
- No.3/2010 Binär-logistische Regressionsanalyse.
Grundlagen und Anwendung für Sozialwissenschaftler.
Jochen Mayerl, Dieter Urban
- No.1/2011 Pädosexuelle Viktimisierung und pädosexuelle Straffälligkeit. Ein „harter“ empirischer
Hypothesentest.
Dieter Urban, Joachim Fiebig
- No.1/2012 Meta-Analyse und "publication bias". Eine empirisch-statistische Fallstudie.
Joachim Fiebig, Gerrit Rüdibusch, Dieter Urban
- No.1/2013 Panelanalyse mit Mehrebenenmodellen. Eine anwendungsorientierte Einführung.
Thomas Krause, Dieter Urban
- No.1/2014 Meta-Analyse mit Mehrebenenmodellen unter Verwendung von HLM.
Joachim Fiebig, Dieter Urban

