

Meta-Analyse in der Soziologie: Bilanz der deutschen Scheidungsforschung oder „statistischer Fruchtsalat“?

Meta-Analysis in Sociology: Review of German Research on Divorce or “Statistical Fruit Salad”?

Josef Brüderl

Fakultät für Sozialwissenschaften, Universität Mannheim, A 5, D-68131 Mannheim

Vor kurzem haben Michael Wagner und Bernd Weiß (2003) in dieser Zeitschrift (Band 32, Heft 1) eine Meta-Analyse der deutschen Scheidungsforschung vorgelegt (im Folgenden WW). Mit dieser Arbeit wollen sie die bisher erzielten Befunde der deutschen Scheidungsforschung bilanzieren. Ihre zweite Zielsetzung ist es, die Verfahren der Meta-Analyse in der deutschen Soziologie bekannter zu machen, denn „Meta-Analysen sind insbesondere in der deutschen Soziologie weitgehend unbekannt“ (WW: 30).

Der zweiten Zielstellung kann man nur zustimmen: Meta-Analysen sind für kumulative Forschung höchst nützlich. Insofern muss man Wagner und Weiß dafür dankbar sein, diese Verfahren in die Diskussion gebracht zu haben. Ein „Abfallprodukt“ ihrer Meta-Analyse verdient ebenfalls Beachtung. Die Mehrzahl der verwendeten 42 Publikationen genügt elementaren Standards der empirischen Sozialforschung nicht: Standardfehler werden nicht berichtet, die Konstruktion der Variablen wird nicht genau beschrieben, eine explizite theoretische Begründung fehlt, usw.

Die erste Zielstellung aber – die Bilanzierung der deutschen Scheidungsforschung mittels Meta-Analyse – wurde nicht erreicht: Die Art und Weise wie Wagner und Weiß die Meta-Analyse anwenden, produziert teilweise nicht sinnvoll interpretierbare Ergebnisse. Diese Behauptung soll im Folgenden begründet werden.

Wie Wagner und Weiß bemerken, sind Meta-Analysen in der Medizin und Psychologie weit verbreitet. Die häufigste Anwendung in diesen Disziplinen ist die Zusammenfassung der Ergebnisse von Experimenten. Nehmen wir an, es wurden 10 Experimente zum Zusammenhang von Y und X mit jeweils 50 Fällen durchgeführt. Es liegen 10 Korrelationskoeffizienten r_{YX} vor. Aufgrund der geringen Fallzahl sind die Korrelationskoeffizienten

mit einem hohen Standardfehler behaftet. Eine Meta-Analyse fasst die Ergebnisse der 10 Experimente zusammen und liefert somit einen Schätzer für r_{YX} , der auf 500 Fällen basiert. Damit ist der meta-analytisch gewonnene Schätzer wesentlich präziser. Dies ist eine höchst nützliche Anwendung der Meta-Analyse. Voraussetzung ist allerdings, dass die 10 Experimente kontrolliert durchgeführt wurden (Randomisierung) und die Korrelationskoeffizienten somit den „kausalen“ Effekt abbilden.

Die Erkenntnisse der Soziologie basieren aber in den seltensten Fällen auf Experimenten. Die Daten der Soziologie sind meist mittels eines ex-post-facto Designs erhoben (Umfragen). Die Folge davon ist, dass eine bivariate Korrelation r_{YX} normalerweise konfundiert ist. Mittels elaborierter multivariater Analyseverfahren versucht man deshalb, dem „wahren“ („kausalen“) Effekt nahe zu kommen. Das gelingt aber nur, wenn das multivariate Modell korrekt spezifiziert ist (d.h. alle konfundierenden Größen berücksichtigt sind). Nun sagt dem Forscher aber die Statistik nicht, welche Größen konfundierend sind. Auch die soziologischen Theorien helfen hier nur begrenzt weiter. Insofern ist die soziologische Forschungsliteratur gekennzeichnet von einem Nebeneinander konkurrierender Spezifikationen. Ein Forscher berichtet den bivariaten Effekt, ein weiterer den Effekt aus einem multivariaten Modell. Ein dritter behauptet, dieses sei falsch spezifiziert und verwendet eine neue Spezifikation, usw.

Wagner und Weiß verwenden nun für ihre Meta-Analyse alle publizierten Befunde! Genau hier liegt das fundamentale Problem jeder Meta-Analyse: Verwendet man alle berichteten Effekte unbesehen, so kann es vorkommen, dass das Ergebnis nicht mehr sinnvoll interpretierbar ist. Ein Beispiel soll das verdeutlichen. Nehmen wir an, die bivariate Korrelation $r_{YX} = 0,5$ beruhe auf einer Scheinkorre-

lation. Ein erster (naiver) Forscher berichtet diese Korrelation dennoch in einem Artikel. Ein zweiter (cleverer) Forscher argumentiert, dass dieses Ergebnis auf einer antezedierenden Variablen Z beruht. Mit einem korrekt spezifizierten Modell findet er $r_{YX:Z} = 0$. Ein Meta-Analytiker – von Wagner und Weiß angespornt – wird nun beide Effekte für seine Meta-Analyse verwenden und eine mittlere Effektstärke von 0,25 berichten (ungefähr). Dieses Ergebnis der Meta-Analyse ist offensichtlich problematisch. Während beide ursprünglichen Effekte eine sinnvolle Interpretation haben (konfundierter bzw. „wahrer“ Effekt), hat der mittlere Effekt keinerlei sinnvolle Interpretation. Eine solche Meta-Analyse produziert „statistischen Fruchtsalat“.

Das Problem entsteht offensichtlich dadurch, dass man alle berichteten Effekte in eine Meta-Analyse einbezieht: bivariate Effekte, aus falsch spezifizierten Modellen resultierende Effekte und – wenn man Glück hat – auch „wahre“ Effekte aus einem korrekt spezifizierten Modell. Genau dies tun aber Wagner und Weiß in ihrer Bilanz der deutschen Scheidungsforschung. Da auch in diesem Feld unterschiedlichst spezifizierte Modelle nebeneinander verwendet werden, sind die von ihnen berichteten mittleren Effektstärken in vielen Fällen nicht sinnvoll interpretierbar.

Wagner und Weiß verschärfen dieses Problem noch dadurch, dass jeder in einer Publikation berichtete Effekt in die Meta-Analyse eingeht; sie sagen dies zwar nicht explizit, aber aus Abb. 1 und der Tatsache, dass aus 42 verwerteten Publikationen bis zu 90 Effektstärken eingehen, kann man dies folgern. Ein Autor, der seine Regressionsmodelle schrittweise aufbaut, um den Quellen der Konfundierung auf die Spur zu kommen, wird bivariat beginnend das Modell schrittweise um Variablengruppen erweitern, bis schließlich die volle Spezifikation erreicht wird. Es gehen somit mehrere Effektstärken in die Analyse ein. Hat der Autor beispielsweise fünf Modelle berichtet, dann trägt er vier mehr oder weniger konfundierte Effekte bei und einen „wahren“ Effekt. Da kann er im Text noch so überzeugend argumentieren, dass nur das fünfte Modell korrekt spezifiziert ist. Sein Beitrag für die „Bilanz der deutschen Scheidungsforschung“ enthält überwiegend konfundierte Effekte!

Diese Probleme sind nicht nur hypothetisch. Sie treten in WW ganz konkret auf. Dies will ich an einem Beispiel verdeutlichen. Es betrifft den Effekt der Kohabitations-Dummy (nichteheliche Lebensgemeinschaft). 74 Effekte gehen in die Meta-Analyse ein, 61 Effekte sind positiv, 13 sind negativ. Als mittlere Effektstärke ergibt sich +15 % (WW: Tab. 3), d. h.

als „Bilanz“ der jahrelangen Forschung zu diesem Effekt zeigt sich, dass eine der Ehe vorangehende Kohabitation die Scheidungsrate um 15 % erhöht. Dieses Ergebnis kommt dadurch zustande, dass alle publizierten Effekte verwendet werden: die meisten sind aber konfundiert. Verwendet man Standard-Scheidungsmodelle, so bewegt sich der Kohabitations-Effekt in der Größenordnung von etwa +40 % bis +80 % (siehe z. B. Brüderl et al. 1997). Dies ist ein erstaunliches Ergebnis, widerspricht es doch der höchst plausiblen „Probetheese“. Dennoch wird es von fast allen deutschen Scheidungsstudien so berichtet. Brüderl et al. (1999) argumentieren aber, dass Standard-Scheidungsmodelle fehlspezifiziert sind. Mit ihrer verbesserten Spezifikation erhalten sie einen Effekt von –6 %. Der „wahre“ Kohabitationseffekt ist negativ: eine vorangehende Probeehe senkt also doch das Scheidungsrisiko! Dies konnte inzwischen in weiteren Publikationen mit anderen Daten bestätigt werden (z. B. Hall 1999). In der Meta-Analyse von Wagner und Weiß findet dies aber keinen Niederschlag, weil die 61 konfundierten Effekte die „Bilanz“ dominieren.

Analoge Probleme gibt es mit vielen von WW berichteten Effekten. Einige Beispiele: In der neueren Literatur wird die korrekte Spezifikation der Suchkosten diskutiert. Ehespezifische Investitionen – so ein durchaus plausibles Argument – könnten endogen sein, was aber in Standard-Scheidungsmodellen nicht berücksichtigt wird (Brüderl/Kalter 2001). Ein analoges Argument wird bezüglich des Effektes der Erwerbstätigkeit der Ehefrau diskutiert (Beck/Hartmann 1999). Schließlich ist auch der Transmissionseffekt stark von der verwendeten Spezifikation abhängig (Esser 2002). Auch bezüglich dieser Determinanten macht es einfach keinen Sinn, über die Effekte unterschiedlichster Spezifikationen zu mitteln.

Wagner und Weiß sind sich dieser Problematik durchaus bewusst; ihnen zufolge „stellt die Meta-Analyse noch keine adäquaten Verfahren bereit, um Befunde zu vereinheitlichen, die durch multivariate Analyseverfahren gewonnen wurden“ (WW: 34); „Koeffizienten, die unterschiedlich spezifizierten Modellen entstammen, schätzen nicht denselben Parameter“ (WW: 35). Genau! Aber mit einem pragmatischen Argument wird dann doch über alle Effekte gemittelt: „Uns ist keine Meta-Analyse bekannt, die Befunde nicht berücksichtigt, nur weil sie beispielsweise einem multivariaten Modell entstammen“ (WW: 34). Auch wenn andere Meta-Analytiker dies so gemacht haben sollten, so ist dies dennoch keine sinnvolle Strategie!

Als Fazit kann man festhalten, dass die von Wagner und Weiß vorgelegte Bilanz der deutschen Scheidungsforschung mit Vorsicht zu genießen ist. Viele – wenn nicht gar die meisten – der von ihnen berichteten mittleren Effektstärken sind durch konfundierte Effekte dominiert. Man kann nur davor warnen, die von dieser „Bilanz“ berichteten Effekte unkritisch als den Stand der Forschung anzusehen.

Dies soll aber nicht heißen, dass die Meta-Analyse an sich für die Soziologie unbrauchbar ist. Sie ist nur dann unbrauchbar, wenn man unbesehen alle berichteten Effekte verwendet. Sie ist sehr wohl brauchbar, wenn der quantitativen Meta-Analyse ein „qualitativer Review“ vorausgeht. Dieser müsste die publizierten Effekte klassifizieren und bewerten. So sollten offensichtlich fehlerbehaftete Publikationen von vornherein ausgeschlossen werden. Allerdings sind die Bewertungskriterien explizit darzulegen, weil das Ganze sonst in pure Willkür ausartet. Basierend auf einer solchen Klassifizierung könnte man dann etwa die mittlere Effektstärke der bivariaten Effekte errechnen. Dies wäre ein sehr hilfreicher Beitrag zur Scheidungsforschung (siehe etwa die von WW erwähnten Meta-Analysen zu den Scheidungsfolgen). Noch hilfreicher wäre natürlich eine Meta-Analyse der „wahren“ Effekte (wie in den experimentellen Wissenschaften üblich). Aber was korrekt und was fehlerhaft spezifiziert ist, das kann uns natürlich auch die Meta-Analyse nicht sagen.

Literatur

- Beck, N. / Hartmann, J., 1999: Die Wechselwirkung zwischen Erwerbstätigkeit der Ehefrau und Ehestabilität unter Berücksichtigung des sozialen Wandels. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 51: 655–680.
- Brüderl, J. / Diekmann, A. / Engelhardt, H., 1997: Erhöht eine Probeehe das Scheidungsrisiko? *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 49: 205–222.
- Brüderl, J. / Diekmann, A. / Engelhardt, H., 1999: Artefakte in der Scheidungsursachenforschung? *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 51: 744–753.
- Brüderl, J. / Kalter, F., 2001: The Dissolution of Marriages: The Role of Information and Marital-Specific Capital. *Journal of Mathematical Sociology* 25: 403–421.
- Esser, H., 2002: Ehekrisen: Das (Re-)Framing der Ehe und der Anstieg der Scheidungsraten. *Zeitschrift für Soziologie* 31: 472–496.
- Hall, A., 1999: Drum prüfe wer sich ewig bindet. S. 119–141 in: T. Klein / J. Kopp (Hrsg.), *Scheidungsursachen aus soziologischer Sicht*. Würzburg: Ergon.
- Wagner, M. / Weiß, B., 2003: Bilanz der deutschen Scheidungsforschung: Versuch einer Meta-Analyse. *Zeitschrift für Soziologie* 32: 29–49.