

Karin Bammann, Nina Wawro  
Bremer Institut für Präventionsforschung und Sozialmedizin (BIPS)  
Linzer Strasse 10  
28359 Bremen  
Telefon: 0421-5959656; Fax: 0421-5959668  
e-mail: bammann@bips.uni-bremen.de

## **Die Einbeziehung genetischer Faktoren in Studien der Epidemiologie**

### **Zusammenfassung**

Neben der klassischen Epidemiologie hat sich in den letzten beiden Jahrzehnten die genetische Epidemiologie als neue Disziplin fest etabliert. Der Beitrag liefert einen kurzen Überblick der Epidemiologie genetischer Faktoren. Zunächst wird die Natur der komplexen Erkrankungen diskutiert, die durch nicht-genetische Faktoren und durch das Zusammenspiel mehrerer Gene beeinflusst werden. Es werden die Probleme erläutert, die mit der geeigneten Modellierung von Gen-Gen- und Gen-Umwelt-Interaktionen und ihrer Integration in kausale Wirkkreise verbunden sind. Mit der Abschätzung der Beteiligung am gesamten Morbiditätsgeschehen wird die Bedeutung genetischer Faktoren für die Ätiologie komplexer Erkrankungen unterstrichen. Dann werden klassische und neuere Studiendesigns vorgestellt, die eine adäquate Einbeziehung genetischer Daten ermöglichen. Die Einführung der letzteren wird durch die Erläuterung der potenziellen Verzerrung von Ergebnissen genetisch-epidemiologischer Studien aufgrund unberücksichtigter genetischer Heterogenität (Populationsstratifikation) motiviert. Anschließend wird die Problematik der mangelnden Replizierbarkeit von Studienergebnissen aufgegriffen und die Möglichkeiten, die sich durch das Konzept der so genannten Mendelschen Randomisierung eröffnen, diskutiert.

*Schlüsselwörter: Epidemiologie, genetische Faktoren, komplexe Erkrankungen, Studiendesign, Populationsstratifikation*

## **The integration of genetic factors into epidemiological studies**

### **Abstract**

During the last two decades, genetic epidemiology has been established in parallel to the area of classical epidemiology. The paper presents some essentials of the epidemiology of genetic factors. It begins with a discussion of complex diseases that are characterized by an involvement of several genes. The paper elucidates the problems that are attached to modeling gene-gene- and gene-environment-interactions and their integration into causal pathways. The role of genetic factors in the etiology of complex diseases is investigated. Classical and new epidemiological study designs that allow an integration of genetic data are introduced. The introduction is partly motivated by the danger of bias due to genetic heterogeneity (population stratification) in classical designs. The problem of replication of study results is discussed and the concept of Mendelian randomization is presented.

*Keywords: epidemiology, genetic factors, complex diseases, study design, population stratification*

Die gebündelte Anstrengung durch multizentrische Großforschungsvorhaben, wie dem internationalen *Humane Genome Project (HGP)* oder dem deutschen *Nationalen Genomforschungsnetz (NGFN)*, haben entscheidend zu einer rasanten wissenschaftlichen und technologischen Entwicklung auf dem Gebiet der Genetik beigetragen und wichtige Impulse für die angrenzenden Fachgebiete gegeben. Dies erstreckt sich auch auf die Gesundheitsforschung und die Epidemiologie. Hier befasst sich insbesondere die genetische Epidemiologie mit der Identifizierung von Genen, die an der Entstehung von Krankheiten bzw. dem Erhalt von Gesundheit beteiligt sind. Das Forschungsgebiet der genetischen Epidemiologie etablierte sich Mitte der 80-er Jahre - trotz der etwas irreführenden Denomination - weitgehend unabhängig von der klassischen Epidemiologie. Insbesondere der Bevölkerungsbezug, das zentrale Element der klassischen Epidemiologie, wurde auf dem Gebiet der genetischen Epidemiologie wenig beachtet. Umgekehrt nahm die klassische Epidemiologie wenig Notiz von den Besonderheiten genetischer Faktoren, deren Abhängigkeitsstrukturen (auf intraindividuell sowie auf Familien- und Bevölkerungsebene) besondere Methoden verlangen. Die Ursachen hierfür könnten unter anderem darin liegen, dass sich die genetische Epidemiologie aus der Humangenetik und nicht aus der Epidemiologie entwickelt hat, während sich andererseits die Epidemiologie lange Zeit nur wenig Erkenntnisgewinn von diesem neuen Forschungsbereich versprach. Seit einigen Jahren gibt es jedoch vermehrte Bemühungen, beide Gebiete stärker miteinander zu verknüpfen (siehe das Plädoyer von [1] und auch [2]).

Im folgenden Beitrag wird ein Überblick über die Epidemiologie genetischer Faktoren gegeben. Beginnend mit einer Beschreibung genetisch mitbedingter Erkrankungen und einer Abschätzung ihres Anteils am gesamten Morbiditätsgeschehen, behandelt die Arbeit anschließend einzelne Bausteine der Krankheitsentstehung. Es werden Gen-Gen- und Gen-Umwelt-Interaktionen sowie ihre Einbindung in Wirkkreise (pathways) diskutiert. Weiterhin werden epidemiologische Studien und verschiedene Studiendesigns zur Aufdeckung genetischer Faktoren überblicksartig beleuchtet. Dabei wird auch auf Confounding durch genetische Heterogenität eingegangen. Abschließend wird die Problematik mangelnder Replizierbarkeit von Studienergebnissen diskutiert und mögliche Zukunftsperspektiven, wie z.B. Mendelsche Randomisierung, aufgezeigt.

## **Ü1: Komplexe Erkrankungen**

Erkrankungen mit genetischem Anteil lassen sich grob in zwei Kategorien einteilen, die *monogenen Erkrankungen*, die durch eine einzige Genvariante quasi-deterministisch verursacht werden und die so genannten *komplexen Erkrankungen*, die multifaktoriell bedingt sind und bei denen neben genetischen in aller Regel auch nicht-genetische Faktoren Mitverursacher sind. Bei Vorliegen der beteiligten Genvarianten tritt also die entsprechende Erkrankung nicht unbedingt auf, sondern es besteht lediglich eine genetische Prädisposition, die das Entstehen dieser Erkrankung begünstigt bzw. wahrscheinlicher macht. Beispiele komplexer Erkrankungen sind Krebs- und Herz-Kreislauf-Erkrankungen, Asthma, Multiple Sklerose und Diabetes mellitus.

Während die verursachenden Genvarianten monogener Erkrankungen vielfach schon lange bekannt sind, sind genetische Ursachen komplexer Erkrankungen erst in jüngerer Zeit Gegenstand der Forschung. Die komplexen Erkrankungen sind jedoch aus gesundheitspolitischer Sicht von besonderem Interesse, weil sie in der Bevölkerung häufig auftreten und, auf Grund der beteiligten nicht-genetischen Ursachen, besonderes Potenzial für Interventionen bieten. Zudem bestehen für nahezu alle der gegenwärtig zehn häufigsten Todesursachen bei Männern und Frauen in Deutschland (Ausnahmen sind Pneumonie und alkoholische Leberkrankheit) genetische Risikofaktoren. Sie sind zu den komplexen Erkrankungen zu zählen.

Komplexe Erkrankungen können, bei gleichem Phänotyp, Subformen aufweisen, die unterschiedliche genetische (und nicht-genetische) Ursachen haben. Einzelne Formen können dabei nahezu den Charakter einer monogenen Erkrankung besitzen. Bekanntestes diesbezügliches Beispiel ist der familiäre Brustkrebs, der hauptsächlich durch Mutationen in den Genen BRCA1 und BRCA2 verursacht wird. So besitzen BRCA1-Mutationsträgerinnen eine 65%-Wahrscheinlichkeit, im Laufe ihres Lebens an Brustkrebs zu erkranken, bei BRCA2 wird diese Wahrscheinlichkeit mit 45% angegeben ([3], [4]). Andererseits sind über 90% aller Brustkrebsfälle nicht durch eine dieser Genvarianten verursacht. Hier spielen klassische Risikofaktoren eine Rolle, aber auch weitere genetische Faktoren, zum Beispiel bei der Regelung des Zellzyklus. Man spricht hier von einer ätiologischen (genetischen) Heterogenität einer Erkrankung. Diese erschwert die epidemiologische Forschung zusätzlich.

>>> Tabelle 1 ungefähr hier einfügen

In Tabelle 1 sind die geschätzten Anteile genetischer Ursachen an der Gesamtheit aller Ursachen für einzelne Erkrankungen angegeben. Es ist zu bedenken, dass sich dieser Anteil zusammensetzt aus dem Anteil der familiären Subtypen einer Erkrankung und dem Anteil der genetischen Faktoren der sporadischen Subtypen (d.h. ohne erkennbare familiäre Häufung) dieser Erkrankung. Des Weiteren wirken genetische Faktoren unter Umständen auf indirektem Weg über einen klassischen Risikofaktor wie bspw. über den Blutdruck ([5]) oder das Suchtverhalten (zur Nikotinabhängigkeit siehe z.B. [6]). Diese Faktoren werden üblicherweise bei der Erfassung des genetischen Anteils einer Erkrankung nicht mit berücksichtigt, was zu seiner Unterschätzung führt. Es zeigt sich, dass ohnehin die Angabe einer konkreten Zahl, zumindest derzeit, in vielen Fällen schwierig oder unmöglich ist. Darüber hinaus ist zu bedenken, dass eine solche Angabe wie jedes Attributivrisiko von der Prävalenz der einzelnen Risikofaktoren abhängig ist. Dies zeigt sich deutlich am Beispiel Lungenkrebs: In der gegenwärtigen Situation mit einem geschätzten Anteil an Rauchern von 36% der männlichen bzw. an Raucherinnen von 21,5% der weiblichen Bevölkerung lassen sich mindestens 85% aller Lungenkrebsfälle auf das Rauchen zurückführen ([7]) und der Anteil der Lungenkrebsfälle mit vermutlich vorwiegend genetischer Komponente ist gemessen an der Gesamtheit der Lungenkrebsfälle sehr gering. Gäbe es hingegen keine Raucher in der Bevölkerung, würde die *absolute Zahl* der Lungenkrebsfälle drastisch sinken, aber der *relative Anteil* der Fälle mit vermutlich genetischen Ursachen stark ansteigen. Insofern können die in

Tabelle 1 genannten Zahlen nur eine Momentaufnahme darstellen. Es soll jedoch nicht unerwähnt bleiben, dass eine Subgruppe junger Lungenkrebspatienten existiert, bei denen eine familiäre Häufung nachgewiesen ist ([8]).

## Ü1: Ursachen komplexer Erkrankungen

Obwohl für die meisten komplexen Erkrankungen eine Vielzahl nicht-genetischer Risikofaktoren etabliert ist, wie zum Beispiel Rauchen als Hauptursache für Lungenkrebs, hat die Einbeziehung genetischer Faktoren eine Reihe neuer Impulse gegeben. Die bisherige epidemiologische Forschung beschränkte sich weitgehend auf die Quantifizierung von Einzeleffekten und die Etablierung von Dosis-Wirkungsbeziehungen für einzelne Risikofaktoren. Die Berücksichtigung genetischer Faktoren und von Stoffwechselprodukten eröffnet hingegen viel stärker als früher die Möglichkeit zu Grunde liegende Kausalzusammenhänge zu erkennen und zu modellieren. Hierfür bieten sich besonders Gene mit bekannter Funktion an, die sich naturgemäß wesentlich einfacher in eine biologische Kausalkette stellen lassen als bspw. ein klassischer Risikofaktor wie das Zigarettenrauchen, bei dem über 4000 einzelne Stoffe ([9]) involviert sind. Ein weiterer Trend ist die Betrachtung von genetisch beeinflussten vorgelagerten Endpunkten und anderen so genannten *intermediären Phänotypen*. Diese sind nicht nur für die Erforschung von Ursachen von Bedeutung, sondern bieten theoretisch Ansatzpunkte für eine effiziente Prävention (siehe auch [10]). In der Praxis kann die Erhebung von intermediären Phänotypen, vor allem an gesunden Probanden, jedoch schwierig sein. Können Laborparameter beispielsweise im Blut oder Urin noch vergleichsweise einfach in großen Studien ermittelt werden, so dürften Parameter, die durch aufwändigere Verfahren, wie z.B. Biopsien, bestimmt werden müssen, eher selten zu erheben sein.

Trotz der Bemühungen der letzten Jahre, ist die Forschung jedoch von einer ganzheitlichen Erfassung und Betrachtung des Ursachengeflechts komplexer Erkrankungen weit entfernt. Üblicherweise werden einzelne Genvarianten bzw. einzelne Ursachenstränge betrachtet. Zudem werden selten Zwischenschritte der Metabolisierung in die Analysen einbezogen. Ein weiterer Punkt ist, dass Phänotypen zumeist auf einen binären Status reduziert werden, der häufig künstlich gesetzt ist (zum Beispiel das Überschreiten eines Laborparameters, wie dem Cholesterinspiegel). Auch bezüglich der Anzahl der beteiligten Gene besteht grundsätzlich Unsicherheit. So stehen zwei Hypothesen nebeneinander:

**Hypothese 1: common-disease/common-variant.** Die genetische Prädisposition komplexer Erkrankungen wird durch *mehrere häufig auftretende Genvarianten* bestimmt, die jeweils nur einen *geringen Effekt* haben.

**Hypothese 2: heterogeneity hypothesis.** Die genetische Prädisposition komplexer Erkrankungen wird durch viele unterschiedliche *selten auftretende Genvarianten* bestimmt, die jeweils einen *großen Effekt* haben.

Diese beiden Hypothesen schließen sich gegenseitig nicht grundsätzlich aus. Auf bestimmte Erkrankungen könnte Hypothese 1 und auf andere Hypothese 2 zutreffen, auch Kombinationen beider Hypothesen wären denkbar. Es ist zurzeit nicht bekannt, welche dieser drei Möglichkeiten am häufigsten auftritt. Es lässt

sich jedoch feststellen, dass Hypothese 2 epidemiologischen Methoden nahezu nicht zugänglich ist, da hier trotz großer postulierter Effekte unrealisierbare Fallzahlen benötigt werden. Gleichzeitig ist hier die Anzahl der einzubeziehenden Gene, bei sehr niedriger Prävalenz der verursachenden Genotypen, sehr hoch ([11]). Häufig auftretende Genvarianten lassen sich hingegen in epidemiologischen Studien leichter untersuchen. Sie haben zudem eine größere Bedeutung für die öffentliche Gesundheit, da auf Bevölkerungsebene weit verbreitete Risikofaktoren mit kleinem Effekt bedeutsamer sind als seltene Risikofaktoren mit hohem Effekt (siehe [12]).

## **Ü1: Zusammenwirken verschiedener Faktoren bei der Krankheitsentstehung**

Zur Identifizierung genetischer Risikofaktoren einzelner Erkrankungen ist es wünschenswert, deren komplexes Wechselspiel untereinander und mit externen Faktoren einzubeziehen. Dies gilt umso mehr, als sich beispielsweise Effekte einzelner Gene gegenseitig maskieren (oder verstärken) können oder die genetische Ausstattung eines Individuums die Suszeptibilität (bzw. Resistenz) gegenüber Noxen beeinflussen kann. Phänomene dieser Art werden *Gen-Gen-Interaktionen* bzw. *Gen-Umwelt-Interaktionen* genannt.

**Beispiel einer Gen-Gen-Interaktion.** Ein Beispiel für das Vorliegen einer Gen-Gen-Interaktion ist die Hirschprung'sche Krankheit, eine erbliche Erkrankung des Dickdarms, zu deren Entstehung Genvarianten an drei unterschiedlichen Genorten beitragen ([13]). Nur die Anwesenheit aller drei Genvarianten führt in diesem Fall zum Auftreten der Erkrankung. Jedoch gilt auch die Hirschprung'sche Krankheit als heterogene Erkrankung, d.h. diese Gen-Gen-Interaktion betrifft nur eine Subform der Erkrankung.

**Beispiel einer Gen-Umwelt-Interaktion.** Ein Beispiel für das Vorliegen einer Gen-Umwelt-Interaktion ist das frühkindliche Asthma. Für dessen Entstehung scheinen mehrere unterschiedliche Gen-Umwelt-Interaktionen relevant zu sein. So wurden beispielsweise Genvarianten, die die Suszeptibilität von Expositionen gegenüber Passivrauchen ([14]) erhöhen, nachgewiesen. Weiterhin wurde bei Asthma ein Zusammenhang mit frühkindlichem Krippenbesuch (und damit einer frühen Exposition gegenüber unterschiedlichen Erregern) beobachtet ([15]). Diese Interaktionen beeinflussen die frühe Ausbildung und nachfolgende Entwicklung des Immunsystems und verändern so das Risiko für die Ausprägung einer allergischen Erkrankung.

Die Einbindung von Gen-Gen- und Gen-Umwelt-Interaktionen in komplexere Modelle (causal pathways) erlaubt die Abbildung von Mechanismen der Krankheitsentstehung. Problematisch hierbei ist zum einen die Datenerhebung zum anderen fehlen geeignete statistische Verfahren, um vorgeschlagene biologische Modelle anhand erhobener Daten auf ihre Korrektheit zu testen. So gibt es derzeit kein etabliertes Verfahren, um biologische Interaktionen im Rahmen einer statistischen Analyse ohne Vorwissen aufzudecken. Die wünschenswerte Einbeziehung von Variablen auf verschiedenen Stufen (wie Expositionen, Gene, Genexpressions- und Proteinlevel, Metaboliten und vorgelagerten Endpunkten) stellt erhebliche Anforderungen auch an die

Datenerhebung. Hier stoßen epidemiologische Studien häufig an ihre Grenzen, da zu deren Erhebung vielfach invasive Verfahren notwendig sind, die oft aus ethischen oder auch aus logistischen Gründen nicht eingesetzt werden. Es ist also mehr als fraglich, ob epidemiologische Daten allein genügen, um diese komplexen Kausalbeziehungen aufzudecken. Eine mögliche Lösung wären Studienkonzepte, die Daten aus unterschiedlichen Quellen, wie epidemiologischen Studien, tierexperimentellen Studien, klinischen Studien, in vitro-Experimenten etc., heranziehen ([16], [17]). Auch an dieser Stelle besteht erheblicher Forschungsbedarf bezüglich möglicher epidemiologischer Designs und einzusetzender statistischer Verfahren.

## **Ü1: Epidemiologische Studiendesigns zur Identifizierung genetischer Risikofaktoren**

Grundsätzlich teilen sich die Vorgehensweisen zur Identifizierung genetischer Risikofaktoren in zwei methodische Ansätze. Eine Möglichkeit ist es, bereits bekannte Gene, die vermutlich mit der Ausprägung einer Krankheit in Zusammenhang stehen, so genannte *Kandidatengene*, hinsichtlich ihres gemeinsamen Auftretens (*Assoziation*) in Bezug auf die Erkrankung zu untersuchen. Ein solches Verfahren wird häufig confirmatorisch, d.h. zur Absicherung bestehender Hypothesen, eingesetzt. Bei einem derartigen Vorgehen werden im Rahmen einer Studie die zu untersuchenden Genvarianten für jedes Individuum bestimmt und in die statistischen Analysen mit einbezogen.

Anders ist es, wenn derartige Kandidatengene nicht bekannt sind oder neue potenzielle Kandidatengene gesucht werden sollen. Eine Suche mittels so genannter *genetischer Marker* macht sich die Tatsache zunutze, dass auf einem Chromosom räumlich eng zusammen liegende Gene überzufällig häufig gemeinsam vererbt werden. Man spricht hier auch von *Kopplung*. Wichtigstes Merkmal dieser genetischen Marker ist, dass ihre Lage auf dem Genom und häufig auch ihre Funktion bekannt sind. Findet man in einer Studie einen Zusammenhang zwischen einem genetischen Marker und dem Auftreten einer Erkrankung, wird so die Lage eines potenziellen Kandidatengens eingegrenzt. Diese explorative Suche erfolgt oftmals genomweit und wird deshalb auch als *whole genome scan* bezeichnet. In genomweiten Assoziationsstudien können auf diese Weise eine große Anzahl an Hypothesen gleichzeitig getestet werden. Allerdings liegt hier ein multiples Testproblem vor, für das geeignet adjustiert werden muss. Vor wenigen Jahren galten diese genotypisierungsaufwendigen Studien noch als nicht realisierbar, aktuell werden sie jedoch immer öfter eingesetzt.

Um den Einfluss genetischer Faktoren auf eine Erkrankung zu erfassen, können prinzipiell klassisch epidemiologische Studiendesigns wie Fall-Kontroll-Studien, Kohortenstudien oder Querschnittsstudien verwendet werden. Zusätzlich wurden aus diesen klassischen epidemiologischen Studiendesigns mehrere neue Designs abgeleitet, die die Familienstruktur genetischer Information adäquat berücksichtigen.

## **Ü2: Klassische epidemiologische Studiendesigns**

Klassische Studiendesigns sind besonders dann geeignet, falls ein oder mehrere Kandidatengene untersucht werden sollen oder wenn eine begrenzte Anzahl zu prüfender Hypothesen über genetische Faktoren besteht. Diese genetischen Faktoren und auch zusätzliche Familieninformationen können wie jede andere Expositionsvariable in die Studie einbezogen werden. Somit können mit denselben statistischen Methoden, wie sie in der traditionellen Epidemiologie verwendet werden, Haupteffekte analysiert, aber auch in einem allerdings begrenzten Umfang Gen-Gen- und Gen-Umwelt-Interaktionen etabliert werden.

Das vermutlich intuitivste Studiendesign der Epidemiologie ist die *Kohortenstudie*, bei der eine definierte Bevölkerungsgruppe unter Exposition oder unter Risiko einer Exposition prospektiv über einen langen Zeitraum beobachtet wird. Für unterschiedliche Expositionsgruppen können dann Morbidität oder Mortalität direkt berechnet und verglichen werden. Für seltene Erkrankungen und vergleichsweise häufig auftretende Risikofaktoren stellen *Fall-Kontroll-Studien* eine wirtschaftliche Alternative zu den sehr zeit- und kostenaufwendigen Kohortenstudien dar. In Fall-Kontroll-Studien werden erkrankte Personen (Fälle) mit nicht erkrankten Personen (Kontrollen) hinsichtlich ihrer retrospektiv erhobenen Exposition miteinander verglichen. Andere Studiendesigns, die in der traditionellen Epidemiologie verwendet werden, sind z.B. eingestete Fall-Kontroll-Studien und Fall-Kohorten-Studien, in denen Fälle und Kontrollen aus einer bestehenden Kohorte gezogen werden. Jedoch besteht sowohl für Kohortenstudien, für Fall-Kontroll-Studien als auch für die meisten anderen klassischen Ansätze, die potentielle Gefahr von Confounding durch unberücksichtigte genetische Heterogenität (*Populationsstratifikation*) innerhalb der Studienpopulation.

### Ü3: Populationsstratifikation

Ein klassisches Beispiel für das Vorliegen von Populationsstratifikation stellt eine von Knowler und Kollegen veröffentlichte Studie zu genetischen Ursachen von Diabetes mellitus in einer Population von Pima- und Papago-Indianern dar ([18]). In der Gesamtgruppe der Studie zeigte sich ein Zusammenhang zwischen einer bestimmten genetischen Variante und dem Auftreten von Diabetes mellitus. Wurden die Daten jedoch getrennt für die einzelnen ethnischen Subgruppen analysiert, war dieser Zusammenhang nicht mehr vorhanden. Eine Nichtberücksichtigung des genetischen Hintergrundes hätte somit fälschlicherweise eine Assoziation zwischen der genetischen Variante und der Erkrankung nachgewiesen.

>>> Abbildung 1 ungefähr hier einfügen

In Abbildung 1 ist das Problem der Populationsstratifikation schematisch dargestellt. In Population 1 liegt innerhalb der Gruppe der Fälle und innerhalb der Gruppe der Kontrollen jeweils die gleiche Verteilung der drei Genotypen vor. Dies gilt auch für Population 2. Die Verteilung ist in den beiden Populationen unterschiedlich (in Population 1 überwiegt der Genotyp *aa*, in Population 2 ist es

der Genotyp  $Aa$ ), jedoch besteht innerhalb der beiden Populationen offensichtlich kein Zusammenhang zwischen Fall-Kontroll-Status und Genotypverteilung. In der Gesamtpopulation unterscheiden sich die Genotypverteilungen der Fälle und der Kontrollen hingegen deutlich voneinander. Hier wäre eine Assoziation zwischen Genotyp und Erkrankung nachweisbar. Diese resultiert jedoch nicht aus einem echten Zusammenhang zwischen Genotyp und Erkrankung, sondern einzig daraus, dass die Zusammensetzung der beiden Subpopulationen bei Fällen und Kontrollen unterschiedlich ist.

In einer aus genetischer Sicht heterogenen Gesellschaft, die Migrationsprozessen unterliegt, ist es selten möglich, genetisch homogene Subpopulationen zu identifizieren, um diese dann stratifiziert analysieren zu können. Daher wird für die Analyse epidemiologischer Studien ein Verfahren zum Umgang mit einer Populationsstratifikation benötigt. Derzeit werden hauptsächlich zwei unterschiedliche Verfahren eingesetzt, für die allerdings zusätzliche genetische Marker benötigt werden. Dabei handelt es sich zum einen um die Methode der so genannten *Genomic Controls* ([19]). Bei dieser Methode wird anhand von krankheitsunabhängigen genetischen Markern der Einfluss der Stratifikation geschätzt und kann so in nachfolgenden Analysen berücksichtigt werden. Zum anderen wird das *Structured-Association-Verfahren* ([20], [21]) eingesetzt. Bei diesem Verfahren wird für jede Person der individuelle genetische Anteil an den beteiligten Subpopulationen geschätzt. Auch diese Schätzung geht in die nachfolgenden Analysen mit ein. Der Ansatz reicht jedoch über die *Genomic Controls* hinaus, da hier explizit die Natur der Populationsstruktur erfasst wird.

## Ü2: Erweiterte Studiendesigns

Die meisten der im Folgenden dargestellten erweiterten Ansätze versuchen, das Problem der Populationsstratifikation zu umgehen, indem sie Familienmitglieder von Studienteilnehmern in das Studiendesign einbeziehen. Dahinter steht die Idee, dass Familienmitglieder jeweils aus der gleichen Population stammen und somit Einheiten mit einem genetisch homogenen Hintergrund bilden. Mögliche Varianten der klassischen epidemiologischen Fall-Kontroll-Studie sind familienbasierte Fall-Kontroll-Designs, in denen (reale oder hypothetische) Kontrollen aus der Familie der Fälle abgeleitet werden.

In *Fall-Geschwister-Studien* (case-sibling studies) werden jedem Fall Geschwister zugeordnet, die in dem Alter nicht erkrankt waren, in dem der zugehörige Fall erkrankte (Indexalter). Mittels Standardmethoden gematchter Fall-Kontroll-Studien, wie der bedingten logistischen Regression, können in einer solchen Studie genetische und nicht-genetische Faktoren sowie Gen-Umwelt-Interaktionen analysiert werden. In *Fall-Eltern-Studien* (case-parent studies) werden hingegen aus den nicht an den Fall übertragenen Allelen der Eltern jeweils drei hypothetische Pseudo-Geschwister (pseudo-siblings) konstruiert, die dann jeweils drei Kontrollen pro Fall bilden. Für die statistische Analyse finden Methoden 1:3-gematchter Fall-Kontroll-Studien Anwendung. Jedoch können in einem solchen Design keine Haupteffekte nicht-genetischer Faktoren geschätzt werden, da für die Pseudo-Geschwister keine Beobachtungen vorliegen. Während beide Designs sicher für Populationsstratifikation kontrollieren, unterliegen sie anderen potenziellen Verzerrungsquellen. Im Fall-Geschwister-Design können



Fälle ohne nicht erkrankte Geschwister nicht berücksichtigt werden, d.h. diese müssen aus der Analyse ausgeschlossen werden und beeinträchtigen so die Repräsentativität der Stichprobe. Das Fall-Eltern-Design erfordert die Probenentnahme und Genotypisierung beider Eltern und ist daher für Erkrankungen, die erst im höheren Alter auftreten nur sehr eingeschränkt verwendbar. Auch in diesem Fall müssen unvollständige Familien-Trios aus der Analyse ausgeschlossen werden. Ein Überblick über retrospektive Studiendesigns findet sich in ([22]).

## Ü2: Adoptions- und Zwillingsstudien

Aus historischer Sicht gibt es zwei weitere wichtige Studientypen zur Untersuchung genetischer Faktoren, die Adoptions- und die Zwillingsstudien. In *Adoptionsstudien* werden unterschiedliche Familienkonstellationen von Familien mit mindestens einem nicht biologisch verwandten Mitglied in Bezug auf die interessierende Krankheit miteinander verglichen, um den Einfluss genetischer Faktoren unabhängig von nicht-genetischen Faktoren erfassen zu können. Diese Vergleiche erstrecken sich z.B. auf

- biologische Eltern mit ihren biologischen Kindern vs. adoptierte Kinder,
- adoptierte vs. biologische Kinder,
- biologische Kinder, die zusammen aufwuchsen vs. biologische Kinder, die getrennt voneinander aufwuchsen.

In *Zwillingsstudien* werden die gemeinsamen Phänotypverteilungen monozygoter Zwillinge mit denen dizygoter Zwillinge verglichen. Sowohl Adoptionsstudien als auch Zwillingsstudien können erheblicher Verzerrung (sampling bias) unterliegen, insbesondere wenn die Studienpopulation aus Freiwilligen besteht. Während Adoptionsstudien nicht nur wegen ihrer methodologischen Probleme, sondern auch wegen erheblicher rechtlicher und ethischer Bedenken eher der Vergangenheit angehören, sind Zwillingsstudien nach wie vor von Interesse. Da monozygote Zwillinge genetisch identisch sind, sind sie ideal geeignet, um den Einfluss nicht-genetischer Faktoren zu untersuchen. Einige Länder haben daher Zwillingsregister etabliert, die die Durchführung populationsbasierter Zwillingsstudien erheblich erleichtern. Neuere Entwicklungen zu Zwillingsstudien werden in [23] beschrieben.

## Ü1: Mangelnde Replizierbarkeit von Studienergebnissen

Ein großes Problem bei der Erforschung genetischer Ursachen komplexer Erkrankungen stellt die mangelnde Replizierbarkeit von Studienergebnissen dar. Dieses Problem wurde im Mai 1999 im Editorial von *Nature Genetics* anlässlich einer dort 1998 veröffentlichten Studie über Alzheimer aufgegriffen. Der Veröffentlichung folgten mehrere Leserbriefe, da die Ergebnisse mit Daten anderer Studien nicht replizierbar waren. Als Konsequenz kündigten die Editoren der Zeitschrift an, dass zukünftig an die Publikation von Ergebnissen aus Assoziationsstudien in *Nature Genetics* höhere Anforderungen gestellt würden ([24]). Diese Anforderungen bezogen sich zum einen auf die Stärke und die statistische Signifikanz (Höhe des p-Werts) der Effekte. Zum anderen sollten die

Ergebnisse vor Publikation anhand einer unabhängigen Stichprobe repliziert worden sein oder durch physiologische Daten die funktionale Rolle des gefundenen Gens belegt sein. Eine Reihe anderer Journals haben mittlerweile ähnliche Maßnahmen getroffen. Dieser Schritt verdeutlicht auch die Hilflosigkeit angesichts einer Flut von unterschiedlichen Ergebnissen aus den verschiedenartigsten Studien, deren Autoren häufig anscheinend nicht einmal Notiz voneinander nehmen.

Nichtsdestotrotz greift ein solcher Schritt zu kurz, da es für eine mangelnde Replizierbarkeit neben einer Fülle von potenziellen Fehlern oder Nachlässigkeiten auch Ursachen gibt, die einer näheren Betrachtung wert sind, da sie zum Verständnis der Krankheit beitragen. Sillanpää und Auranen haben diesbezüglich eine sehr lesenswerte Übersicht erstellt ([25]). Sie unterscheiden Ursachen, die eher der Ätiologie zugeordnet werden, wie unterschiedliche Phänotypdefinitionen, ätiologische Heterogenität, unterschiedliche genetische Hintergründe oder Vorhandensein von Umweltfaktoren, die bei Vorliegen von Interaktion entsprechend unterschiedliche Resultate erzeugen können, sowie unterschiedliche Allelhäufigkeiten und unberücksichtigte Populationsstratifikation. Insbesondere können ungenau oder zu breit definierte Phänotypen, die keinem kausalen Mechanismus auf der Molekularebene zuzuordnen sind, eine Nicht-Replizierbarkeit bedingen. Eine zweite Kategorie bilden solche Ursachen, die eher technischer Natur sind. Zu diesen zählen u. a. unterschiedliche Studiendesigns, unterschiedliche Einschlusskriterien, unterschiedliche Stichprobenvarianz, Fehler in den Daten, Annahme unterschiedlicher Vererbungsmodelle und unterschiedliche statistische Verfahren und Modelle, die benutzt werden.

Genotypisierungsfehler stellen eine weitere Quelle für falsche oder verfälschte Resultate dar ([26]). Den größten Einfluss auf die Datenqualität hat die Qualität des biologischen Materials. Zu geringe Mengen und Verschmutzungen schaffen methodische Probleme und erschweren zum Beispiel die Amplifizierung und korrekte Typisierung der Proben. Im Allgemeinen werden keine Schätzungen über Fehlerraten veröffentlicht, auch wenn diese einen wesentlichen Hinweis auf die Qualität der Resultate geben. Pompanon et al. schlagen einen festen Ablauf für die Qualitätssicherung des Genotypisierungsprozesses vor, der, z.B. durch wiederholtes Genotypisieren, Fehlerquellen entdecken und minimieren hilft.

Eine weitere Hilfe, um Studienergebnisse in einen Zusammenhang mit anderen wissenschaftlichen Arbeiten zu stellen sind systematische Reviews zur Assoziation einzelner Genvarianten und Erkrankungen wie sie vom Human Genome Epidemiology Network (HuGENet) angefertigt werden.

## **Ü1: Ein viel versprechendes Konzept: Mendelsche Randomisierung**

Seit einigen Jahren, und insbesondere seit Erscheinen der Sonderausgabe des *International Journal of Epidemiology* 2004 zu diesem Thema, erfährt das Konzept der *Mendelschen Randomisierung* (Mendelian randomization) Beachtung. Die Idee, die hinter diesem Prinzip steht, ist es, durch Stratifizierung nach genetischer Ausstattung mögliche Artefakte bezüglich des Zusammenhangs zwischen Krankheit und Expositionen ausschließen zu können ([27]).

Der Grundgedanke zum Konzept der Mendelschen Randomisierung stammt ursprünglich von Martijn Katan (siehe Nachdruck in [28]). Dieser skizzierte eine Analysestrategie zur Prüfung der Hypothese, dass ein niedriger Cholesterinspiegel das Tumorwachstum begünstigt. Zwar zeigte sich in Studien ein Zusammenhang zwischen Krebs und Cholesterinspiegel, es war aber umstritten, ob der Cholesterinspiegel für das Auftreten von Krebs verantwortlich war oder ob umgekehrt die Krebserkrankung den Cholesterinspiegel beeinflusst. Die Idee war, stattdessen den Zusammenhang zwischen dem Vorliegen bestimmter Varianten des Gens für Apolipoprotein E (ApoE) und dem Auftreten einer Krebserkrankung zu untersuchen. Ein niedriger Cholesterinspiegel ist mit bestimmten Varianten des ApoE-Gens assoziiert. Würde ein niedriger Cholesterinspiegel die Entstehung von Krebs begünstigen, sollte sich daher ein Zusammenhang zwischen diesen ApoE-Varianten und dem Auftreten einer Krebserkrankung zeigen. Umgekehrt würde ein fehlender Zusammenhang zwischen dem ApoE-Gen und Krebs nahe legen, dass es sich bei dem beobachteten Zusammenhang zwischen Cholesterinspiegel und Krebs um ein Artefakt handelt. Der entscheidende Vorteil, den das ApoE-Gen gegenüber dem Cholesterinwert für eine solche Analyse besitzt ist, dass es von Geburt an invariant und die Wirkrichtung somit eindeutig ist. Das Konzept der Mendelschen Randomisierung bietet sich an, viele derartige bekannte Probleme der Epidemiologie zu lösen. Weitere praktische Beispiele finden sich in [27]. Eine weitergehende Diskussion der dem Prinzip zu Grunde liegenden Annahmen geben Nitsch et al. in [29].

### **Ü1: Fazit**

Die Fortschritte auf dem Gebiet der Genetik und der genetischen Epidemiologie haben zunehmend auch in der klassischen Epidemiologie die Suche nach genetischen Faktoren ausgelöst, die komplexen Erkrankungen zugrunde liegen (siehe auch [30]). Der Anteil genetischer Ursachen am Krankheitsgeschehen ist häufig nur klein, verglichen mit anderen Risikofaktoren, wie beispielsweise dem Tabakrauchen und der Ernährung. Das heißt, bei den meisten Erkrankungen leisten nicht-genetische Faktoren, wie der Lebensstil, die Arbeitsbedingungen, soziale und psychische Faktoren sowie die Umwelt den Hauptbeitrag. Dennoch gibt die Erforschung genetischer Faktoren wichtige Einblicke in die kausalen molekularen Mechanismen der Krankheitsentstehung. Neben familiär bedingten Subformen einzelner Erkrankungen interessiert auch die genetisch bedingte individuelle Suszeptibilität gegenüber einzelnen Expositionen. Auf methodischer Ebene besteht jedoch noch eine Reihe bislang ungelöster Probleme. So werden derzeit bestenfalls nur sehr vereinfachte Modelle bzw. Teilaspekte ätiologischer Zusammenhänge geprüft und das komplexe Zusammenspiel zwischen Genen, Expositionen, Stoffwechselkreisläufen und Erkrankung nur ungenügend abgebildet. Es ist zu erwarten, dass sich durch die Verbesserung der technischen Möglichkeiten, wie z.B. der High-throughput-Technologien oder der Etablierung von Biobanken, immer weitere Spielräume für das Forschungsgebiet ergeben. So werden, wie oben bereits erwähnt, mittlerweile die ersten genomweiten Assoziationsstudien durchgeführt. Nach wie vor werden jedoch statistische und epidemiologische Methoden benötigt, die die komplexe Realität adäquat modellieren helfen.

### **Ü1: Weiterführende Internet-Adressen**

### *Allgemeine Informationen und Nachrichten*

DNA from the beginning  
<http://www.dnafb.org/dnafb/>

U.S. Department of Energy Office of Science  
<http://doegenomes.org/>

Genomics and Health Weekly Update  
<http://www.cdc.gov/genomics/update/current.htm>

### *Datenbanken*

Genetic Association Database  
<http://geneticassociationdb.nih.gov/>

OMIM – Mendelian Inheritance in Men  
<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>

### *Organisationen und Netzwerke*

Nationales Genomforschungsnetz  
<http://www.ngfn.de/>

Human Genome Epidemiology Network (HuGENet)  
<http://www.cdc.gov/genomics/hugenet/>

## Literatur

- [1] Thomas DC (2000) Genetic epidemiology with a capital E. *Genet Epidemiol* 19: 289-300
- [2] Khoury MJ, Millikan R, Little J, Gwinn M (2004) The emergence of epidemiology in the genomics age. *Int J Epidemiol* 33: 936-944
- [3] Lux MP, Fasching PA, Beckmann MW (2006) Hereditary breast and ovarian cancer: review and future perspectives. *J Mol Med* 84: 16-28
- [4] Smith KL, Robson ME (2006) Update on Hereditary Breast Cancer. *Curr Oncol Rep* 8: 14-21
- [5] Staessen JA, Wang J, Bianchi G, Birkenhager WH (2003) Essential hypertension. *Lancet* 361: 1629-1641
- [6] Munafò MR, Clark TG, Johnstone EC, Murphy MFG, Walton RT (2004) The genetic basis for smoking behavior: A systematic review and meta-analysis. *Nicotine Tob Res* 6: 583-598
- [7] Statistisches Bundesamt (Hrsg.) (1998) Gesundheitsbericht für Deutschland: Gesundheitsberichterstattung des Bundes. Stuttgart: Metzler-Poeschel
- [8] Broman K, Pohlabein H, Jahn I, Ahrens W, Jöckel, KH (2000) Aggregation of lung cancer in families: results from a population-based case-control study in Germany. *Am J Epidemiol* 152: 497-505
- [9] Deutsches Krebsforschungszentrum (2002) Produktregulierung und Verbraucherinformation. Auszug aus den Handlungsempfehlungen für eine wirksame Tabakkontrollpolitik. Heidelberg
- [10] Khoury MJ, Davis R, Gwinn M, Lindgren ML, Yoon P (2005) Do we need genomic research for the prevention of common diseases with environmental causes? *Am J Epidemiol* 161: 799-805
- [11] Yang Q, Khoury MJ, Friedman JM, Little J, Flanders WD (2005) How many genes underlie the occurrence of common complex diseases in the population? *Int J Epidemiol* 34:1129-1137
- [12] Heller R (2005) Evidence for Population Health. Oxford: Oxford University Press
- [13] Gabriel SB, Salomon R, Pelet A, Angrist M, Amiel J, Fornage M, Attié-Bitach T, Olson JM, Hofstra R, Buys C, Steffann J, Munnich A, Lyonnet S, Chakravarti A (2002) Segregation at three loci explains familial and population risk in Hirschsprung disease. *Nature Genetics* 31: 89-93
- [14] Choudhry S, Avila PC, Nazario S, Ung N, Kho J, Rodriguez-Santana JR, Casal J, Tsai HJ, Torres A, Ziv E, Toscano M, Sylvia JS, Alioto M, Salazar

M, Gomez I, Fagan JK, Salas J, Lilly C, Matallana H, Castro RA, Selman M, Weiss ST, Ford JG, Drazen JM, Rodriguez-Cintron W, Chapela R, Silverman EK, Burchard EG (2005) CD14 tobacco gene-environment interaction modifies asthma severity and immunoglobulin E levels in Latinos with asthma. *Am J Respir Crit Care Med* 172: 173-182

- [15] Hoffjan S, Nicolae D, Ostrovnaya I, Roberg K, Evans M, Mirel DB, Steiner L, Walker K, Shult P, Gangnon RE, Gern JE, Martinez FD, Lemanske RF, Ober C (2005) Gene-Environment Interaction Effects on the Development of Immune Responses in the 1st Year of Life. *Am J Hum Genet* 76: 696–704
- [16] Marks GB (2006) Environmental factors and gene–environment interactions in the aetiology of asthma. *Clin Exp Pharmacol Physiol* 33: 285–289
- [17] Thomas DC (2005) The need for a systematic approach to complex pathways in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev* 14: 557-559
- [18] Knowler WC, Williams RC, Pettitt DJ, Steinberg AG (1988)  $GM^{3;5, 13, 14}$  and type 2 diabetes mellitus: An association in American Indians with genetic admixture. *Am J Hum Gen* 43: 520-526
- [19] Devlin B, Roeder K (1999) Genomic control for association studies. *Biometrics* 55: 997-1004
- [20] Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959
- [21] Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping in structured populations. *Am J Hum Genet* 67: 170-181
- [22] Weinberg CR, Umbach DM (2000) Choosing a retrospective design to assess joint genetic and environmental contributions to risk. *Am J Epidemiol* 152: 197-203
- [23] Boomsma D, Busjahn A, Peltonen L (2002) Classical twin studies and beyond. *Nat Rev Genet* 3: 872-882
- [24] Anonymous (1999) Freely associating. *Nat Genet* 22: 1–2
- [25] Sillanpää MJ, Auranen K (2004) Replication in genetic studies of complex traits. *Ann Hum Genet* 68: 646-657
- [26] Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. *Nat Rev Genet* 6: 847-859
- [27] Smith GD, Ebrahim S (2003) ‘Mendelian randomization’: can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 32:1-22

- [28] Katan MB (2004) Apolipoprotein E isoforms, serum cholesterol, and cancer. 1986. *Int J Epidemiol* 33:9
- [29] Nitsch D, Molokhia M, Smeeth L, DeStavola BL, Whittaker JC, Leon DA (2005) Limits to causal inference based on mendelian randomization: A comparison with randomized controlled trials. *Am J Epidemiol* 163: 397-403
- [30] Hoffmann W, Latza U, Greiser KH, Kroke A, Nieters A, Schulze M, Steiner M, Terschüren C, Wjst M (2002) Biologische Marker in der Epidemiologie: Begriffe, Anwendungen, Perspektiven (Teil II). *Gesundheitswesen* 64:145-152

**Tabelle 1: Anteil genetischer Ursachen ausgewählter Erkrankungen**  
(Quelle, sofern nicht anders angegeben: Gesundheitsberichterstattung des Bundes)

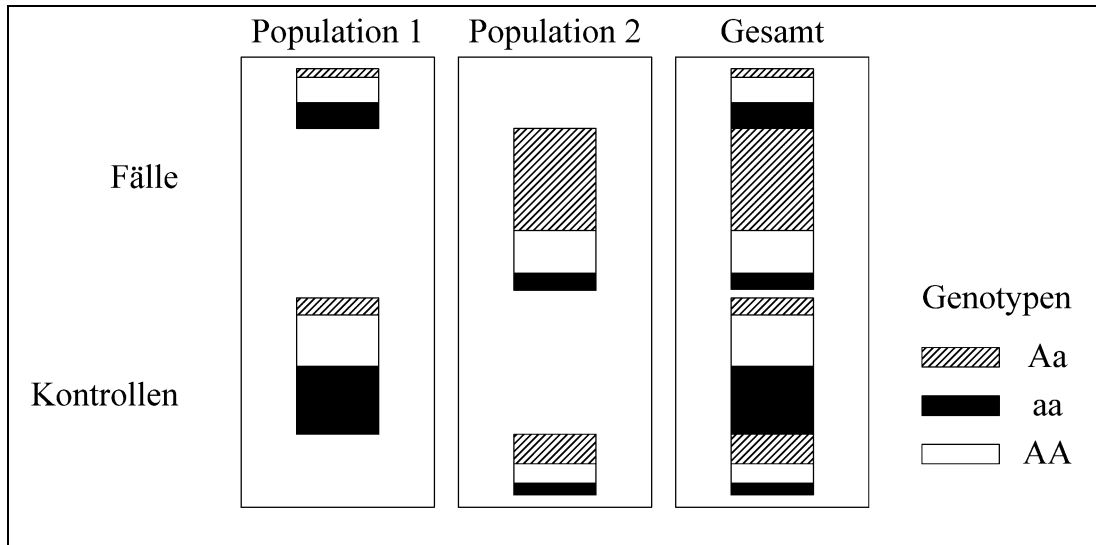
<b>Erkrankung</b>	<b>Genetische Ursachen (geschätzter Anteil in %)</b>	<b>Bemerkungen</b>
Bösartige Neubildungen insgesamt	etwa 5%-10%	-
Bronchialkarzinom	~ 0% (siehe Text)	Jedoch wird der Anteil genetischer Ursachen der Nikotinabhängigkeit auf mindestens 50% geschätzt ([a]).
Mammakarzinom	etwa 5%-10%	Familiäre Variante identifiziert (siehe auch Text).
Kolonkarzinom	nicht angegeben	Familiäre Varianten identifiziert (>6% aller Fälle).
Diabetes mellitus	nicht angegeben	Anzahl konkordanter eineiiger Zwillingspaare: Typ I: 30-50%; Typ II: bis zu 90%.
Adipositas	nicht angegeben	Ergebnis verschiedener Zwillings- und Adoptionsstudien: 60-80% der BMI-Varianz sind genetisch bedingt. Geschätzte Heritabilität ca. 80% ([b]).
Morbus Alzheimer	nicht angegeben	Familiäre Varianten identifiziert (ca. 5-10% aller Erkrankungen); bei den sporadischen Formen, die ca. 90% der Erkrankungen ausmachen gibt es ebenfalls Hinweise auf eine genetische Prädisposition; hier wurden bislang ca. 30 möglicherweise relevante Gene lokalisiert.
Akuter Myokardinfarkt	nicht angegeben	Der Einfluss genetischer Ursachen beschränkt sich vermutlich vor allem auf intermediäre Phänotypen, wie Hypertonie etc.; insgesamt wird der Anteil von Wang et al. in [c] mit 31% angegeben.
Hypertonie	20-50% ([d])	-
Schlaganfall	nicht angegeben	Der Einfluss genetischer Ursachen beschränkt sich vermutlich vor allem auf intermediäre Phänotypen, wie Diabetes, Hypertonie ([e]).
Asthma	etwa 50%	Anzahl konkordanter eineiiger Zwillingspaare aus verschiedenen Studien: 60-80%.
Chronische obstruktive Lungenerkrankung (COPD)	nicht angegeben	Es wurden mehrere potenzielle Kandidatengene identifiziert, der Anteil genetischer Ursachen an der Entwicklung von COPD ist jedoch noch unklar ([f]).

[a] Li MD (2006) The genetics of nicotine dependence. *Curr Psychiatry Rep* 8: 158-164

[b] Bell CG, Walley AJ, Froguel P (2005) The genetics of human obesity. *Nat Rev Genet* 6: 221-34



- [c] Wang Q, Rao S, Shen GQ, Li L, Moliterno DJ, Newby LK, Rogers WJ, Cannata R, Zirzow E, Elston RC, Topol EJ (2004) Premature myocardial infarction novel susceptibility locus on chromosome 1P34-36 identified by genomewide linkage analysis. *Am J Hum Genet* 74:262-271
- [d] Gong, M, Hübner N (2006) Molecular genetics of human hypertension. *Clin Sci* 110: 315-326
- [e] Flossmann E, Rothwell PM (2006) Family history of stroke does not predict risk of stroke after transient ischemic attack. *Stroke* 37: 544-546
- [f] Silverman EK (2002) Genetic Epidemiology of COPD. *Chest* 121 (Suppl.): 1S-6S



**Abbildung 1: Schematische Darstellung von Confounding durch unberücksichtigte genetische Heterogenität (Populationsstratifikation)**