

## **4. DISKUSSION**

In der vorliegenden Studie konnte gezeigt werden, daß in den Massenmedien selten aber im Wesentlichen korrekt über die Methoden der Pränataldiagnostik berichtet wird und daß ein Presstext in der Lage ist, das Wissen und die Einstellung der Lesenden zu beeinflussen. Im folgenden sollen diese Ergebnisse zusammengefaßt und mit Voruntersuchungen verglichen werden. Anschließend werden die Ergebnisse kritisch hinterfragt.

### **4.1 ZUSAMMENFASSUNG DER DATEN UND VERGLEICH ZU VORUNTERSUCHUNGEN**

#### **4.1.1 Inhaltsanalyse**

Die Medien können nur bedingt dazu beitragen, über die Methoden der Pränataldiagnostik aufzuklären, da nur in einem Drittel der Texte über die Pränataldiagnostik auf eine einzelne Untersuchungsmethode eingegangen wird. Das sind in den 136 Texten untersuchten Texten aus drei Publikationen im Zeitraum von 7 Jahren nur 43 Texte. Darüber hinaus ist die Beschreibung der Methoden der Pränataldiagnostik nicht immer ausgewogen sondern folgt bestimmten Mustern, wie im Folgenden gezeigt werden soll.

Die Bewertung der Methoden der Pränataldiagnostik in den untersuchten Presstexten spiegelt zunächst einmal die Bewertung der Methoden der Pränataldiagnostik in der Fachliteratur wieder. Zu einem ähnlichen Ergebnis kamen auch Bubela und Caulfield (2004). Sie verglichen wissenschaftliche Publikationen aus dem Bereich der Genetik mit Artikeln aus Tageszeitungen und kamen zu dem Schluß, daß die Mehrheit der Zeitungsartikel die Einschätzung der wissenschaftlichen Artikel korrekt wieder gibt. Die korrekte Wiedergabe der wissenschaftlichen Artikel bedeutet aber nicht immer, daß die Medien ausgewogen über ein Thema berichten. Dazu sollen zwei Beispiele aus der vorliegenden Studie genannt werden: Der Triple Test und die Fetalzellanalyse.

In der recht ausgewogenen Berichterstattung über den Triple Test entsprechen die Schwerpunkte in etwa denen, die auch in der Fachliteratur eine wichtige Rolle spielen. Dort werden zwei Vorteile genannt. Zum einen ermögliche der Triple Test eine Pränataldiagnostik auch für jüngere Frauen. Zum anderen könnten ältere Schwangere auf eine Amniozentese verzichten, wenn nach einem Triple Test nur noch eine geringe Wahrscheinlichkeit besteht, ein Kind mit einem Down-Syndrom zu bekommen (Harper, 1998; Haddow und Palomaki, 1993; Fortuny, 2000). Bei der Auswertung der Texte wurde dieses Thema als „Wahl der Methode“ kodiert. In den Texten über den Triple Test, kam es besonders häufig vor (36% der Texte). Beispielsweise in der Süddeutschen Zeitung. Dort heißt es: „Bluttests, die schwangerschaftsspezifische Stoffwechselprodukte im Serum untersuchen, helfen – auch bei jüngeren Frauen – gezielt nach Chromosomenstörungen des Feten zu fahnden, und sie können bei älteren manche Amniozentese überflüssig machen“ (Brüser, 1993).

Weniger ausgewogen wird über die Amniozentese berichtet. Sie wird häufig mit der Fetalzellanalyse verglichen, die als risikofreie Alternative beschrieben wird. In Beiträgen, in denen neben der Amniozentese auch die Fetalzellanalyse erwähnt wird, werden signifikant häufiger die Risiken der Amniozentese genannt. Auch bei diesem Thema folgen die Autoren der Medienberichte den Argumenten aus der wissenschaftlichen Diskussion. In einem Beitrag für The Lancet beispielsweise schreiben Poon und Mitarbeiter (2000): Die Pränataldiagnostik mittels Analyse fetaler DNA aus mütterlichem Plasma könnte unsere Abhängigkeit von invasiven Methoden reduzieren und so zu sichereren Untersuchungen für Mutter und Kind führen. In der Süddeutschen Zeitung findet sich eine ähnliche Einschätzung: „Die Techniken, die momentan zur Verfügung stehen ... sind alle mit einem Fehlgeburtsrisiko von ... 0,5 bis zwei Prozent verbunden. Einen erfolgversprechenden Ansatz, bei dem dieses Risiko umgangen wird, verfolgt eine Forschergruppe ... der Universität Münster. Ihr ist es ... gelungen, aus dem Blut der Schwangeren fetale Zellen zu isolieren und mit diesen erfolgreich Nachweise auf bestimmte Chromosomenstörungen zu führen“ heißt es gleich in der Einleitung des Beitrages (Jacobsen, 1993).

Anlaß der Berichterstattung über die Pränataldiagnostik waren vor allem politische Ereignisse (zum Beispiel die Spätabtreibungsdebatte) und wissenschaftliche Fortschritte (zum Beispiel die Entwicklung der Fetalzellanalyse). Aber auch lokale Themen waren Grund, über die Pränataldiagnostik zu berichten (zum Beispiel die Woche für das Leben). Diese Beiträge zu lokalen Themen werden von besonders vielen Menschen wahrgenommen, da sich 84% der Mediennutzer für lokale Nachrichten interessieren. Im Vergleich dazu interessieren sich 68% für Politik und 25% für Wissenschaft (Meyn, 1999).

#### **4.1.2 Experimentelle Studie – Wissen**

Nach der Intervention können beantworten die Probanden mehr Fragen richtig als vorher. Zu ähnlichen Ergebnissen kam auch eine Studie zur Wirkung eines Aufklärungsfaltblatts zum Mundbodenkarzinom (Humphris und Field, 2003). Besonders profitieren in der vorliegenden Studie die männlichen Probanden, Probanden unter 35 und Probanden ohne Kinder. Dabei ist zu bedenken, daß das Wissen zum Zeitpunkt der Abfrage ein Maximum erreicht hat und mit der Zeit wieder abnehmen wird. Besonders wenn in kurzer Zeit viel Wissen präsentiert wird, vergessen medizinische Laien zahlreiche Fakten wieder. Durch Verwendung einer einfachen Sprache, klarer Kategorien und einiger Wiederholungen läßt sich der Lernzuwachs allerdings verbessern (Ley, 1979).

Auch die Einschätzung der Wahrscheinlichkeit wird durch die Intervention beeinflusst. Probanden in der Prozentzahlen-Gruppe schätzen Wahrscheinlichkeiten niedriger ein als Probanden in der natürliche-Häufigkeiten-Gruppe. Das gilt sowohl für die Wahrscheinlichkeit, ein Kind mit Down-Syndrom zu bekommen als auch für die Einschätzung der Risiken der invasiven Diagnostik. Ein anderes Ergebnis ergibt sich bei der Abschätzung der Zuverlässigkeit des Triple Test: Diese schätzen die Probanden in der Prozentzahlen-Gruppe etwas höher ein als die in der natürliche-Häufigkeiten-Gruppe.

Dieser Einschätzung liegt zugrunde, daß in der Intervention keine direkte Aussage zur Zuverlässigkeit gemacht wird sondern lediglich Daten zur Häufigkeit von positiven Ergebnissen bei einem Triple Test mit der Häufigkeit eines Down-Syndroms verglichen werden. Vermutlich schätzen die Probanden aus der Prozentzahlen-Gruppe die Häufigkeit von positiven Ergebnissen ähnlich niedrig ein wie die Häufigkeit des Down-Syndroms und werten daher die Zuverlässigkeit des Triple Test auf verbalen Skalen etwas höher. Somit liegt also auch dieser Einschätzung eine als niedrig wahrgenommene Wahrscheinlichkeit zugrunde, so daß zusammenfassend festgestellt werden kann, daß die Probanden in der Prozentzahlen-Gruppe die Wahrscheinlichkeiten niedriger einschätzen als die in der natürliche-Häufigkeiten-Gruppe.

Aus Voruntersuchungen ist bereits bekannt, daß Probanden Wahrscheinlichkeiten exakter einschätzen, wenn Sie ihre Informationen in natürlichen Häufigkeiten und nicht in Prozentangaben bekommen (Hoffrage et al, 2000, Gigerenzer und Edwards, 2003). Bei einer Untersuchung zum positiv prädiktiven Wert verschiedener diagnostischer Tests schätzten 46% der Probanden in der natürliche-Häufigkeiten-Gruppe die Wahrscheinlichkeit einer Erkrankung bei positivem Testergebnis richtig ein, in der Prozentzahlengruppe sank der Anteil auf 10% (Hoffrage und Gigerenzer, 1998). Dieses Ergebnis kann in der vorliegenden Studie nicht reproduziert werden. In beiden Gruppen schätzt etwa die Hälfte der Probanden den positiv prädiktiven Wert des Triple Tests richtig ein.

Der Grund dafür liegt vermutlich in der Gestaltung des Textes. Dort erfahren Probanden aus beiden Gruppen neben den Zahlen, aus denen sie den positiv prädiktiven Wert abschätzen sollen, daß bei einem positiven Testergebnis im Triple Test weitere Untersuchungen notwendig sind, um zu einem zuverlässigen Ergebnis zu kommen. Diese Überlegungen werden durch Vorarbeiten von Rothman und Kiviniemi (1999) unterstützt. Die Autoren kommen in einem Übersichtsartikel zu dem Ergebnis, daß bei der Abschätzung von Wahrscheinlichkeiten neben Zahlen auch andere Faktoren eine Rolle spielen. Dazu zählen Faktoren, die erklären, wie ein Risiko zustande kommt (antecedents) und die Folgen für den Betroffenen

(consequences). Rothman und Kiviniemi schlagen vor, daß diese antecedents und consequences in der Vermittlung von Gesundheitsrisiken stärker berücksichtigt werden sollten.

#### **4.1.3 Experimentelle Studie – Einstellung**

Das in der Intervention vermittelte Wissen beeinflusst die Präferenzen der Probanden in Bezug auf die pränatale Diagnostik. Vor dem Lesen des Textes würden zwei Drittel einem Triple Test (63%) und ein Drittel einer invasiven Diagnostik (Amniozentese: 37%, Chorionzottenbiopsie: 20%) zustimmen. Fast alle halten eine Sonographie für notwendig (93%). Diese Werte sind vergleichbar mit Daten, die Jorgensen (1995) an schwangeren Frauen erhoben hat. Von diesen gaben 79% an, alle Frauen sollten einen AFP-Test angeboten bekommen (dieser ist mit dem Triple Test vergleichbar); 70% meinten, allen Frauen sollte eine Sonographie angeboten werden; und 20% sagten, daß allen Frauen eine invasive Diagnostik angeboten werden sollte.

Nach der Intervention sinkt die Zustimmung zum Triple Test (47%), zur Amniozentese (13%) und zur Chorionzottenbiopsie (7%), möglicherweise weil der Eindruck entstanden ist, der Triple Test sei unzuverlässig und die invasive Diagnostik gefährlich. Die Zustimmung zur Sonographie bleibt unverändert hoch. Schwangere Frauen, die tatsächlich vor der Entscheidung stehen, ob sie eine Methode der Pränataldiagnostik wahrnehmen sollen, reagieren ähnlich auf zusätzliche Information: weniger von ihnen stimmen einer Amniozentese zu, nachdem sie in einer Art Unterrichtsklasse über die Methoden der Pränataldiagnostik informiert wurden. Hingegen nehmen etwas mehr Schwangere einen Triple Test in Anspruch. Die Anzahl der Ultraschalluntersuchungen wird durch die zusätzliche Information, ähnlich wie in der vorliegenden Studie, nicht verändert (Thornton et al, 1995). Das liegt möglicherweise daran, daß die Ultraschalluntersuchung von den schwangeren Frauen weniger genutzt wird, um medizinisch wichtige Entscheidungen zu treffen, sondern um beispielsweise das Geschlecht des Kindes zu erfahren (Stephens et al, 2000).

Der in der Intervention vorgestellte Umgang mit einem positiven Testergebnis beeinflusst die Einstellung der Probanden in Bezug auf Schwangerschaftsabbruch und das Leben mit einer bei der Pränataldiagnostik diagnostizierten Behinderung, dem Down-Syndrom. Probanden, die einen Text lasen, in dem sich eine Schwangere Frau zum Abbruch einer Schwangerschaft entscheidet, stimmen nach dem Lesen des Textes häufiger einem Abbruch bei schwerer Behinderung zu. Besonders stark durch den Text beeinflussen lassen sich junge Probanden ohne Kinder und solche, die vor der Intervention neutral eingestellt waren, das heißt einen Schwangerschaftsabbruch bei schwerer Behinderung weder stark befürworteten noch stark ablehnten.

Zu einem vergleichbaren Ergebnis kamen Lang und Lang (1961) bei einer Untersuchung des Einfluß von Fernsehdebatten auf die Wahlentscheidung der US-Amerikaner im Präsidentschaftswahlkampf. Sie fanden heraus, daß durch die Fernsehdebatten nicht diejenigen beeinflusst werden, die sich bereits entschieden haben, sondern die Unentschiedenen. Bei einer tatsächlichen Entscheidung für oder gegen einen Schwangerschaftsabbruch spielen selbstverständlich auch andere Faktoren, wie beispielsweise die Darstellung des Schweregrads der Behinderung durch die behandelnden Ärzte, eine Rolle (Marteau, 1995).

## **4.2 KRITISCHE BETRACHTUNG DER ERGEBNISSE**

### **4.2.1 Inhaltsanalyse**

In der wissenschaftlichen Literatur zur Inhaltsanalyse unterscheidet man zwischen quantitativen und qualitativen Verfahren. Mit Hilfe der quantitativen Inhaltsanalyse lassen sich vor allem die expliziten Inhalte von Texten erfassen (z.B. Nennung von bestimmten Personen und deren Bewertung). Die qualitative Inhaltsanalyse ist eine Erweiterung der quantitativen Inhaltsanalyse (Mayring, 1991). Diese ermöglicht es, auch implizite Inhalte und verborgene Sinnstrukturen zu erfassen. Dazu verwendet man beispielsweise die Kontextanalyse, in der das direkte Textumfeld und Zusatzmaterial, zum Beispiel über den Kommunikator, ausgewertet werden. Eine

weitere Methode der qualitativen Inhaltsanalyse ist die Analyse der Aspekte eines Themas, die nicht genannt werden (Clarke et al., 2005). Die vorliegende Untersuchung beschränkt sich auf eine quantitative Inhaltsanalyse.

Bei der Untersuchung von Texten zum Thema Wissenschaft und Technik werden Inhaltsanalysen meistens an Presstexten gemacht (Kepplinger et al, 1991). Auch zur inhaltsanalytischen Untersuchung von anderen Themen, zum Beispiel Politik, werden in der Regel Presstexte, also Artikel aus Zeitungen und Zeitschriften, herangezogen. Presstexte haben gegenüber Fernsehsendungen und Radiobeiträgen den Vorteil, daß die Botschaften bereits in lesbarer Form vorliegen. Sie können daher direkt ausgewertet werden (Merten, 1983). Die Analyse von Fernsehbildern und Radiobeiträgen ist anfälliger für Fehler, weil beispielsweise der Unterton in der Stimme der Sprecher von den Untersuchern sehr unterschiedlich kodiert werden kann. Für die Analyse von Zeitungen spricht auch, daß Menschen den Berichten aus Zeitungen eher glauben als Beiträgen aus Illustrierten oder dem privaten Fernsehen (Meyn, 1999). Das hat den Vorteil, daß man – wenn auch mit gewissen Einschränkungen – von dem Inhalt der Beiträge in den untersuchten Medien auf die Ansicht der Menschen schließen kann (Merten, 1983).

Untersuchungen zur Politikberichterstattung beruhen häufig auf Analysen der großen überregionalen Zeitungen (Die Welt, Frankfurter Allgemeine Zeitung, Süddeutsche Zeitung und Frankfurter Rundschau). Diese haben den Vorteil, daß sie das gesamte politische Spektrum der bundesdeutschen Tagespresse repräsentieren (Wilke, 1994). Auf die in der vorliegende Untersuchung vorgenommenen Analyse der regionalen Zeitungen verzichteten einige Untersucher, da die oben genannten großen Zeitungen eine Meinungsführerfunktion haben. Von den großen überregionalen Zeitungen stand in der dieser Untersuchung zugrunde liegenden Datenbank ([www.genios.de](http://www.genios.de)) für den Untersuchungszeitraum nur die Süddeutsche Zeitung zur Verfügung. Darüber hinaus haben die in der vorliegenden Untersuchung gewählten Medien den Vorteil, daß sie verschiedene Darstellungsformen abdecken: Das Politische Magazin (Spiegel), die überregionale Tageszeitung (Süddeutsche Zeitung) und die Regionalzeitung (Berliner Zeitung). Die vorliegende Untersuchung

beschränkt sich auf den Zeitraum vom 01.01.1993 bis zum 01.01.2000, da nur für diesen Zeitraum ein kostenloser Zugang zu der oben genannten Datenbank bestand.

Die Zielgruppe an Publikationen für diese Untersuchung waren also Texte, die im Zeitraum vom 01.01.1993 bis zum 01.01.2000 in Politischen Magazinen, überregionalen Tageszeitungen und regionalen Tageszeitungen erschienen sind. Aus dieser Zielgruppe wurden die Auswahleinheiten (also die untersuchten Publikationen) anhand der Auflage ausgewählt. Es läßt sich daher keine sichere Aussage über alle in Deutschland veröffentlichten Printmedien machen. Es ist allerdings unwahrscheinlich, daß in anderen Printmedien wesentlich häufiger über die Methoden der Pränataldiagnostik berichtet wird als in den untersuchten Printmedien. Das wichtigste Ergebnis der Untersuchung, daß Printmedien nur sehr begrenzt zur Aufklärung über die Methoden der Pränataldiagnostik beitragen können, läßt sich also vermutlich auf alle in Deutschland veröffentlichten Printmedien verallgemeinern.

Die Bestimmung der Analyseeinheiten (also der untersuchten Texte) erfolgte in der vorliegenden Untersuchung sehr viel gründlicher als in vergleichbaren Untersuchungen. Wells und Mitarbeiter (2001) beispielsweise verwenden in einer Studie zur Mammographie als Suchworte für die Gewinnung der zu untersuchenden Texte aus der Datenbank lediglich Fachworte wie Mammographie und Mammogramm. In der vorliegenden Studie werden als Suchworte neben den Fachbegriffen (z.B. pränatal) auch die alltagsüblichen Begriffe (z.B. vorgeburtlich) verwandt, was die Vollständigkeit der Suchergebnisse erhöht.

Bei der Analyse der Texte wurde keine Prüfung auf Interkoderreabilität vorgenommen. In anderen inhaltsanalytischen Studien werden häufig mehrere Kodierer unabhängig voneinander zumindest einen Teil der Texte aus. (zum Beispiel Geller et al. 2003). Als Maß für die Qualität derartiger Studien wird die Interkoderreabilität angegeben (Merten, 1983). Darauf wurde in der vorliegenden Studie verzichtet, da es sich nicht um eine journalistische sondern um eine medizinische Untersuchung handelt. In journalistischen Untersuchungen werden

neben objektiven Merkmalen wie Thema und Länge der Beiträge auch stärker subjektiv empfundene Merkmale wie politische Werturteile kodiert. Das macht eine Kodierung durch mehrere Untersucher notwendig. In der vorliegenden Arbeit wurden hingegen medizinische Fakten kodiert, was auch ein einzelner medizinisch vorgebildeter Untersucher leisten kann.

#### **4.2.2 Experimentelle Studie – Studiendesign**

Das Ziel der Studie war es, die Wirkung eines Textes auf Wissen und Einstellung der Lesenden zu messen. Die Wirkung auf das Wissen wurde gemessen als ein Vergleich der Antworten der Probanden vor dem Lesen des Textes und nach dem Lesen des Textes. Eine Plazebokontrolle erfolgte nicht, da es schwer zu vermitteln gewesen wäre, warum ähnlich klingende Fragen ohne jegliche Intervention zweimal beantwortet werden müssen. Außerdem wurden vor dem Lesen des Textes und nach dem Lesen des Textes jeweils unterschiedliche Fragen gestellt. Somit kann nicht sicher ausgeschlossen werden, daß die Veränderung nur auf die wiederholte Befragung oder nur auf die unterschiedlichen Fragen zurückzuführen ist. Darüber hinaus war nach dem Lesen des Textes ein Zurückblättern möglich so daß nicht mit Sicherheit gesagt werden kann, daß Wissenszuwachs im engeren Sinn erfolgt ist.

Ein möglicherweise besseres Studiendesign verwenden Humphris und Mitarbeiter (1999) bei ihrer Evaluierung eines Informationsfaltblattes über Mundbodenkarzinom. Zur Messung der Wirkung des Faltblatts auf das Wissen der Lesenden verwendeten sie folgendes Design: Gruppe A mußte den Fragebogen vor und nach dem Lesen des Faltblattes ausfüllen, Gruppe B bekam den Fragebogen nur nach dem Lesen des Faltblattes und Gruppe C beantwortete den Fragebogen, ohne das Faltblatt gelesen zu haben. Auf diese Weise konnte ausgeschlossen werden, daß das Beantworten des Fragebogens vor dem Lesen des Faltblattes zu einem aufmerksameren Lesen des Faltblattes und somit zu besseren Ergebnissen im Wissenstest führt.

Die Wirkung auf die Einschätzung der Wahrscheinlichkeiten und die Einstellung der Probanden wurde gemessen als randomisierte Studie, wobei die Antworten der Probanden zwischen den unterschiedlichen Studienarmen verglichen wurden. Dabei wurde ein einfaches Randomisierungsverfahren, das alternate assignment (Kramer, 1988), verwendet. Ein aufwändigeres Randomisierungsverfahren, beispielsweise durch Zufallszahlen, ist bei dem Studiendesign vermutlich nicht nötig. Der Erfolg der Randomisierung ist im Ergebnisteil belegt: in beiden Gruppen waren die Konfounder (Alter, Religiösität, medizinische Vorbildung) gleich verteilt. Die Unterschiede zwischen den Gruppen sind also am ehesten durch die Unterschiede in den Studienarmen zu erklären.

In der Medienwirkungsforschung wird zur Untersuchung der Wirkung von Presseberichten auf die Einstellung der Rezipienten jedoch ein sogenanntes Mehrmethodendesign verwendet. Dies besteht aus zwei inhaltlich aufeinander abgestimmten Elementen: einer Inhaltsanalyse und einer Rezipientenbefragung. So untersuchten Merten und Mitarbeiter (1992) beispielsweise die Wirkung von Nachrichtensendungen im Fernsehen, indem sie an einem Abend in einer bestimmten Zahl von Haushalten anriefen. Sie gingen davon aus, daß ein Großteil der Menschen eine der großen Nachrichtensendungen am Abend gesehen hatte. Die Angerufenen wurden zu bestimmten politischen Themen befragt. Außerdem wurde in Erfahrung gebracht, welche Nachrichtensendung der Befragte gesehen hatte. Dann zogen Merten und seine Mitarbeiter Schlüsse vom Inhalt der Nachrichtensendung auf die Wirkung bei den Zuschauern.

Ein ähnliches Design wurde angewendet in Studien zur Wirkung der Medienberichte über die Ankündigung der nahezu vollständigen Entschlüsselung des Genoms im Sommer 2000. Tambor und Mitarbeiter (2002) verglichen den Inhalt von 55 Medienberichten mit den Ergebnissen einer Telefonumfrage (random-digit-dialing). Geller und Mitarbeiter (2003) verglichen den Inhalt von 36 Fernseh- und Zeitungsberichten mit den Ergebnissen einer Befragung unter jungen Ärzten. Dieses Studiendesign hat gegenüber dem in der vorliegenden Studie verwendeten Design den Vorteil, daß nicht nur eine Momentaufnahme nach der Wirkung eines einzelnen

Presstextes erfolgt. Das Mehrmethodendesign ist für die Bearbeitung der Fragestellung der vorliegenden Arbeit allerdings ungeeignet, denn über Pränataldiagnostik wird nur selten berichtet und nur wenige Menschen werden daher in einer kurzen Zeitspanne mehrere Beiträge über Pränataldiagnostik lesen, so daß unmöglich genug Probanden befragt werden können um Rückschlüsse von der Analyse der Medienberichte auf die Ergebnisse der Umfrage zu ziehen.

#### **4.2.3 Experimentelle Studie – Studienpopulation**

Für die vorliegende Studie wurde keine Zufallsstichprobe aus der Zielpopulation gezogen. Vielmehr wurden die Verteilungsorte der Fragebogen so gewählt, daß erwartet werden konnte, daß der Fragebogen von Probanden ausgefüllt wird, die den Kriterien der Zielpopulation entsprechen. Dies kann zu sampling bias führen (Kramer, 1988). Dieser wird jedoch durch die Randomisierung minimiert, da die Konfunder, wie oben beschrieben, in den unterschiedlichen Studienarmen gleich verteilt waren. Dennoch ist das Ergebnis aufgrund der nicht ordnungsgemäßen Ziehung der Stichprobe nicht ohne weiteres auf die Zielpopulation oder auf die gesamte deutsche Bevölkerung verallgemeinerbar. Letzlich konnte nur gezeigt werden, daß die Probanden auf eine bestimmte Art und Weise auf den verteilten Text reagierten. Da jedoch eine sozial heterogene Gruppe (Probanden mit und ohne Kinder, Studierende an Fachhochschulen und an Universitäten, Studierende in sozialen und technischen Berufen) untersucht wurde, kann angenommen werden, daß sich Menschen aus der Zielpopulation ähnlich verhalten werden wie die Probanden in der vorliegenden Studie.

Eine weitere Einschränkung der Generalisierbarkeit der Studienergebnisse besteht aufgrund der niedrigen Rücklaufquote von 57%. In den mit dieser Studie vergleichbaren Untersuchungen zur Wirkung von Medienberichten über die Entschlüsselung des Humangenoms wurde allerdings nur eine ähnliche Rücklaufquote erreicht. Von den 190 von Geller und Mitarbeitern (2003) verteilten Fragebögen wurden 123 ausgewertet (Rücklaufquote 65%). Tambor und Mitarbeiter (2002) kontaktierten 965 Personen, von denen 407 an der Befragung teilnahmen

(Rücklaufquote 42%). Idealerweise sollte eine Rücklaufquote von mindestens 80% erreicht werden. Andernfalls besteht die Gefahr des self-selection-bias, das heißt, daß nur diejenigen den Fragebogen ausfüllen, die sich besonders für das Thema interessieren (Gallus und Lühe, 1998). Für einen Unterpunkt der Studie, die Änderung der Einstellung zum Down-Syndrom wurde daher untersucht, ob sich die Wirkung des Textes auch in der Gruppe der Probanden findet, bei der eine Rücklaufquote von 100% erreicht wurde, weil ihnen Unterrichtszeit zum Ausfüllen der Fragebögen zur Verfügung gestellt wurde. Die Analyse in dieser Untergruppe kam zu einem ähnlichen Ergebnis wie die Untersuchung der Gesamtgruppe<sup>18</sup>. Daher ist davon auszugehen, daß die gezeigte Wirkung des Textes auf Wissen und Einstellung der Probanden nicht wesentlich durch self-selection-bias beeinflusst ist.

#### **4.2.4 Experimentelle Studie – Messung der Zielvariablen**

Das Ziel der vorliegenden Studie war es, die Wirkung eines Presstextes auf Wissen und Einstellung der Lesenden zu messen. Zu diesem Zweck wurde ein Fragebogen entwickelt, der in einem pre-Test auf Verständlichkeit und Vollständigkeit der Fragen getestet wurde. Harte Testkriterien wie Reliabilität und Validität wurden allerdings nur unzureichend berücksichtigt, wie im Folgenden gezeigt wird.

Die Fragen zur Messung des Wissens der Probanden basieren auf einem bereits getesteten Fragebogen von Marteau und Mitarbeitern (1988). Dieser Fragebogen wurde jedoch speziell für schwangere Frauen entwickelt, so daß die Fragen zum Teil geändert werden mußten. Der geänderte Fragebogen wurde vor der Anwendung nicht auf Reliabilität und Validität geprüft. Die Prüfung der Reliabilität der Wissensfragen sollte in zukünftigen Studien anhand der Kuder-Richardson-Formel (zum Beispiel KR-21) erfolgen. Diese bestimmt die split-half-Reliabilität für alle

---

<sup>18</sup> Der mittlere Unterschied zwischen beiden Studienarmen war auf der Skala der Frage nach einem erfüllten Leben mit Down-Syndrom in der Gruppe mit 100% Rücklauf 0,62 (95% Konfidenzintervall: 0,13—1,11), in der Gesamtgruppe 0,71 (95% Konfidenzintervall: 0,39—1,02). Bei der Frage nach einem selbstständigen Leben war der mittlere Unterschied auf der Skala zwischen beiden Studienarmen 0,64 in der Gruppe mit 100% Rücklauf (95% Konfidenzintervall: 0,12—1,12) und 0,74 in der Gesamtgruppe (95% Konfidenzintervall: 0,42—1,07).

denkbaren Teilungen eines Tests. Sie ist der Test-Retest-Reliabilität überlegen, da Lerneffekte zwischen den beiden, für die Bestimmung der Test-Retest-Reliabilität nötigen, Testungen ausgeschlossen sind (zum Beispiel Humphris et al. 1999). Der KR-21-Wert für die vor dem Lesen des Textes in der vorliegenden Studie verwendeten Fragen beträgt 0,15, für die nach dem Lesen des Textes verwendeten Fragen beträgt der KR-21-Wert 0,40. Diese Werte sind niedriger als für reliable Tests geforderte Werte, so daß nicht sicher gesagt werden kann, ob die einzelnen Items des Tests das Wissen in etwa gleich gut messen.

Die Prüfung der Validität der Prüfungsfragen sollte in zukünftigen Studien anhand der „criterion validity“ untersucht werden. So wäre beispielsweise zu untersuchen, ob bei Fragen zum Thema Pränataldiagnostik Gynäkologen besser abschneiden als Allgemeinmediziner und diese wiederum besser abschneiden als medizinische Laien. In der vorliegenden Studie beantworteten vor dem Lesen des Textes die Medizinstudenten unter den Probanden statistisch signifikant mehr Fragen richtig als die Nicht-Medizinstudenten (Mittelwert der Medizinstudenten: 1,90, 95%-Konfidenzintervall: 1,73—2,07; Mittelwert der Nicht-Medizinstudenten: 1,50, 95%-Konfidenzintervall: 1,43—1,59).

Die Fragen zur Einstellung sollten beantwortet werden auf einer monopolen Skala mit Zahlenvergabe und verbaler Extrempunktbeschreibung. Diese Skala wurde ebenfalls nicht auf Reliabilität und Validität geprüft. In zukünftigen Studien sollte die Reliabilität der Einstellungsskala geprüft werden durch Test-Retest-Reliabilitätsprüfung (zum Beispiel Humphris et al., 2000). Eine Prüfung der Validität ist sehr schwierig, da eine tatsächliche Entscheidung (als Ausdruck der Einstellung) nur im Längsschnitt zu erfassen wäre. Eine solche Untersuchung liegt beispielsweise vor zur Wirkung einer Werbekampagne der Firma Novartis, die über Pilzerkrankungen der Fußnägel (Onychomykose) aufklärte. Laut einer Kohortenstudie von t´Jong und Mitarbeitern (2004) stieg in der Zeit dieser Kampagne die Verschreibungshäufigkeit für ein Antikmykotikum der Firma Novartis an.

Zusammenfassend läßt sich sagen, daß nur der Wissensfragebogen validiert ist. Die Reliabilität des Wissensfragebogen, gemessen mit der Kuder-Richardson-Formel, ist ungenügend. Die Validität und Reliabilität der Einstellungsfragen wurde nicht geprüft. Das bedeutet für die Auswertung, daß die Ergebnisse dieser Studie erst nach einer Wiederholung der Untersuchung mit besser evaluierten Instrumenten als sicher gelten können. Die vorliegende Studie gibt also Hinweise auf eine Verbesserung des Wissens durch die Intervention jedoch keine sichere Messung einer Einstellungsänderung. Es konnte lediglich gezeigt werden, daß die Probanden in Abhängigkeit vom Studienarm die Fragen zur Einstellung unterschiedlich beantworteten. Dies kann man werten als einen Hinweis auf die Beeinflussung der Probanden durch die Intervention.

#### **4.3 FAZIT**

In den untersuchten Printmedien wird selten über die Methoden der Pränataldiagnostik berichtet und dabei werden bestimmte Schwerpunkte gesetzt, die zwar inhaltlich richtig sind aber zu einem verzerrtem Bild führen können. Es müssen also andere Wege der Aufklärung über die Methoden der Pränataldiagnostik gefunden werden. Eine Möglichkeit sind Informationsfaltblätter, die zum Beispiel über Arztpraxen oder Hochschulen und Berufsschulen in der Öffentlichkeit verteilt werden könnten. Die Ergebnisse der vorliegenden experimentellen Studie weisen darauf hin, daß das Lesen eines Textes dazu beitragen kann, daß die Lesenden etwas über die Methoden der Pränataldiagnostik lernen und sich daher gezielter für bestimmte Untersuchungen entscheiden. Aufgrund der oben beschriebenen methodischen Mängel kann die vorliegende Untersuchung aber nur als Pilotstudie gelten. Ausgehend davon erscheint es sinnvoll, ein Informationsfaltblatt über die Methoden der Pränataldiagnostik genauer zu evaluieren und dabei aus den Erfahrungen der vorliegenden Studie zu lernen.

Dazu sollte in folgenden Schritten vorgegangen werden. Zunächst sollte ein Faltblatt entwickelt werden, welches einen möglichst hohen Wert auf der Medical Information Design Assessment Scale (MIDAS) erhält. Mit dieser Skala werden beispielsweise

die Formulierung und das Layout von medizinischen Informationen bewertet (Krass et al., 2002). Die Wahrscheinlichkeiten im Text sollten als natürliche Häufigkeiten ausgedrückt werden. Dies erhöht, wie oben beschrieben, die Verständlichkeit, besonders wenn auch die Bedingungen für bestimmte Risiken (antecedents) und ihre Folgen (consequences) beschrieben werden. Ferner sollte in dem Faltblatt auf die exemplarische Beschreibung der Entscheidung einer Betroffenen verzichtet werden, da dies vermutlich die Lesenden beeinflussen könnte und somit dem Ziel der genetischen Beratung entgegensteht, nicht-direktiv Optionen aufzuzeigen.

In einem nächsten Schritt sollte das Faltblatt in einer Plazebo-kontrollierten randomisierten Studie mit drei Studienarmen evaluiert werden. In Studienarm A sollte, wie oben beschrieben, ein Fragebogen vor und nach dem Lesen des Faltblatts ausgefüllt werden. Die Probanden in Studienarm B sollten den Fragebogen nur nach dem Lesen des Faltblatts beantworten. Im Studienarm C sollten die Probanden nur den Fragebogen ohne ein Faltblatt bekommen. Die Randomisierung sollte man als Gruppenrandomisierung durchführen, um Kontaminationseffekte zu vermeiden. So könnte man beispielsweise an einem Befragungstermin jeweils nur einen Studienarm untersuchen. Als Setting könnte man, wie in der vorliegenden Studie, Fachhochschulen oder, wie Humphris und Mitarbeiter (1999), Arztpraxen wählen. Als Einschlußkriterien wären alle Personen zu wählen, die jünger sind als 40 Jahre.

Als Zielvariablen zu wählen wären Wissen und Einstellung zu den Methoden der Pränataldiagnostik. Der Wissensfragebogen sollte Fragen enthalten zu Prinzip und Risiken der Testmethoden sowie zur Zuverlässigkeit und Konsequenz der jeweiligen Testergebnisse. Der Wissensfragebogen sollte auf Reliabilität und Validität geprüft werden. Zur Reliabilitätsprüfung könnte man die oben beschriebene Kuder-Richardson-Formel benutzen. Zur Validitätsprüfung ist zu untersuchen, ob sich die Testergebnisse zwischen Gynäkologen, Allgemeinmedizinern und medizinischen Laien unterscheiden. Der Einstellungsfragebogen sollte fragen, welche Methoden der Pränataldiagnostik der Proband selber wahrnehmen beziehungsweise seiner Partnerin empfehlen würde. Zur Reliabilitätsprüfung könnte man die Test-Retest-

Reliabilität heranziehen. Eine Validitätsprüfung ist wie oben beschrieben nur schwer möglich.

In der statistischen Auswertung sollten folgende Punkte zwischen den Studienarmen verglichen werden: Median und 95%-Konfidenzintervall der Anzahl der richtig beantworteten Fragen; Anteil (inklusive des 95%-Konfidenzintervalls) der Probanden, die eine bestimmte Untersuchung selber wahrnehmen beziehungsweise ihrem Partner empfehlen würden. Ferner wäre zu untersuchen, ob die Bereitschaft, bestimmte Untersuchungen wahrzunehmen, mit dem Wissen über die Methoden der Pränataldiagnostik korreliert. Dies wäre zu untersuchen anhand einer nach Wissenspunkten stratifizierten Auswertung des Anteils (inklusive des 95%-Konfidenzintervalls) der Probanden, die eine bestimmte Untersuchung für richtig halten.

Die These wäre, daß die Probanden, analog zu der vorliegenden Studie, nach dem Lesen des Faltblatts mehr Wissensfragen richtig beantworten und durch die Aufklärung gezielter bestimmte Untersuchungen wählen, daß heißt zwischen Screening-Untersuchungen (Triple Test und Ultraschall) und Untersuchungen zur Bestätigung der Verdachtsdiagnose bei auffälligen Screeningtests (Amniozentese und Chorionzottenbiopsie) unterscheiden können.