

INSTITUT FÜR STATISTIK  
DER LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN



Bachelorarbeit

Exit Polls:  
Geschichte und statistische Aspekte

Manuela Hölzlwimmer

Aufgabensteller und Betreuer: Prof. Dr. Helmut Küchenhoff

Abgabetermin: 18. Mai 2009

---

## **Zusammenfassung**

Die Arbeit befasst sich mit dem Thema „Exit Polls“ vornehmlich in den Vereinigten Staaten, Großbritannien und Deutschland. Dabei wird auf die historische Entwicklung der Wahltagsbefragung in diesen drei Ländern eingegangen. Es werden sowohl die Vorläufer behandelt, wie die Prognose auf Grundlage von ausgezählten Stimmen und die Analyse von charakteristischen Bezirken in den USA, als auch neuere Entwicklungen wie der Übergang zum Paneldesign in Großbritannien. Aus statistischer Sicht liegt ein besonderer Augenmerk auf der Stichprobenziehung für die Exit Polls, den verschiedenen Schätzverfahren, wobei besonders die Quotientenschätzung in den USA und die Regressionsmodelle in Großbritannien Beachtung finden, und den Gewichtungen, die für Antwortverweigerung und Auswahlwahrscheinlichkeiten vorgenommen werden.

Außerdem wird die Genauigkeit von Exit Polls thematisiert und welche Diskussionen sich dabei ergeben. Dabei soll unter anderem aufgezeigt werden, welche Gründe es für die falschen Prognosen in Florida bei der Präsidentschaftswahl 2000 gab und wie sich die Abweichung der Exit-Poll-Ergebnisse bei der Wahl 2004 in Amerika von den tatsächlichen Resultaten erklären lässt. Schließlich werden mögliche Neuerungen bei der Wahltagsbefragung behandelt, die durch die Veränderung der politischen und sozialen Realität nötig geworden sind.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>5</b>
<b>2</b>	<b>Allgemeine Fakten zum Exit Poll</b>	<b>6</b>
<b>3</b>	<b>Geschichte der Exit Polls ...</b>	<b>8</b>
3.1	... in den USA . . . . .	8
3.2	... in Deutschland . . . . .	11
3.3	... in Großbritannien . . . . .	12
<b>4</b>	<b>Methodik für die USA, Großbritannien und Deutschland</b>	<b>14</b>
4.1	Stichprobenziehung . . . . .	14
4.1.1	Stichprobengröße . . . . .	14
4.1.2	Stichprobenziehung . . . . .	15
4.2	Fragebogen und Durchführung der Befragung . . . . .	17
4.3	Schätzung und Prognose . . . . .	19
4.3.1	Gewichtung . . . . .	20
4.3.2	Quotientenschätzung in den USA . . . . .	23
4.3.3	Prognose am Wahlabend . . . . .	25
4.4	Evaluation . . . . .	25
4.5	Wahltagsbefragung in Deutschland . . . . .	26
4.5.1	Stichprobenziehung . . . . .	27
4.5.2	Gewichtung . . . . .	27
4.5.3	Analyse . . . . .	28
4.6	Exit Polls in Großbritannien . . . . .	29
4.6.1	Auswahl der Stichprobenbezirke . . . . .	29
4.6.2	Durchführung der Wahltagsbefragung . . . . .	31
4.6.3	Prognose . . . . .	31
<b>5</b>	<b>Verschiedene Fehlertypen und Fehlerquellen</b>	<b>34</b>
<b>6</b>	<b>Die Exit Polls in den Jahren 2000 und 2004 in den USA</b>	<b>36</b>
6.1	Die Präsidentschaftswahl 2000 . . . . .	36
6.2	Die Präsidentschaftswahl 2004 . . . . .	38
<b>7</b>	<b>Genauigkeit von Prognosen auf Grundlage von Exit Polls</b>	<b>41</b>
7.1	Analyse der Exit-Poll-Daten zur amerikanischen Präsidentschaftswahl 2004 . . .	41

7.1.1	Beschreibung des Datensatzes . . . . .	41
7.1.2	Ergebnisse für den Exit Poll 2004 . . . . .	43
7.2	Genauigkeit der 18-Uhr-Prognose in Deutschland . . . . .	46
<b>8</b>	<b>Aktuelle Forschungsfragen</b>	<b>48</b>
8.1	Die Befragung von Latino-Wählern . . . . .	48
8.2	Vorzeitige Wähler und Briefwähler . . . . .	49
<b>9</b>	<b>Eine kritische Auseinandersetzung mit Exit Polls</b>	<b>51</b>

# 1 Einleitung

Obwohl Wahlberichterstattungen zu den Fernsehübertragungen mit den höchsten Einschaltquoten gehören, machen sich die wenigsten Menschen Gedanken darüber, wie es möglich ist, dass die Fernsehsender bereits direkt nach der Schließung der Wahllokale erste Ergebnisse verkünden können und diese in vielen Fällen den tatsächlichen Wahlausgang sehr gut treffen. Möglich gemacht werden diese Prognosen durch Exit Polls, im Deutschen auch Wahltagsbefragungen genannt. Dabei handelt es sich um die Befragung von Wählern direkt nach Verlassen des Wahllokals, um herauszufinden, für wen die Wähler gestimmt und warum sie sich so entschieden haben.

Obwohl Exit Polls zeitlich unbegrenzt von großem Nutzen sind, war der Grund für ihre Entwicklung vor allem der Wunsch nach umfassender Berichterstattung am Wahlabend und die Tatsache, dass man möglichst bald verlässliche Ergebnisse benötigte, um Politiker zu einer Stellungnahme zu bewegen. So braucht es nicht zu verwundern, dass Exit Polls eine „Erfindung“ der Medien sind. Da Vorwählerhebungen mit zahlreichen Schwierigkeiten umgehen müssen, sind sie oftmals keine sehr verlässliche Grundlage für Analysen am Wahlabend. Zahlreiche Beispiele aus den vergangenen Jahren wie die britischen Unterhauswahlen 1992, als die Vorwahlbefragungen einen deutlichen Sieg der Labourpartei voraussagten, tatsächlich aber die Tories gewannen, belegen diese Schwäche (Moon 1999). Solide durchgeführte Exit Polls dagegen haben in den allermeisten Fällen sehr verlässliche Prognosen geliefert. Doch auch die Wahltagsbefragungen wurden in letzter Zeit bei mehreren Wahlen mit ihren Grenzen konfrontiert, beispielsweise bei der Präsidentschaftswahl 2000 in den USA. Die Fälle, in denen es Probleme mit Prognosen auf der Basis von Wahltagsbefragungen gab, werden im Folgenden genauso thematisiert wie die Faktoren für einen gelungenen Exit Poll.

Während das grundsätzliche Vorgehen in den meisten Ländern ähnlich ist, gibt es auch Unterschiede, da verschiedene geographische und politische Voraussetzungen auch unterschiedliches Vorgehen erfordern. So sehen sich beispielsweise die deutschen und die amerikanischen Statistiker mit unterschiedlichen Wahlsystemen konfrontiert. Deutschland hat das Verhältniswahlrecht und der Stimmenanteil, den eine Partei bei der Wahl erhält, wird auch in etwa in dieselbe Anzahl an Sitzen umgerechnet. Beim britischen Mehrheitswahlrecht dagegen gilt das Prinzip „the winner takes it all“ und so liegt die Hauptaufgabe darin, den siegreichen Kandidaten zu prognostizieren. In den Vereinigten Staaten gilt wie in Großbritannien das Mehrheitswahlrecht, aber die beiden Länder verfolgen aufgrund unterschiedlicher Voraussetzungen verschiedene Strategien beim Exit Poll. Deshalb wird ein Teil dieser Arbeit die Unterschiede und Gemeinsamkeiten bei den Wahltagsbefragungen in den USA, in Großbritannien und in Deutschland thematisieren.

## 2 Allgemeine Fakten zum Exit Poll

Wahltagsbefragungen haben zwei Ziele: Prognose und Analyse. So ermöglichen es Exit Polls manchmal, bereits bevor die erste Stimme ausgezählt ist, den Gewinner einer Wahl zu verkünden oder wenn die Entscheidung nicht allzu knapp ist, zumindest im Laufe der Auszählung den Sieger zu prognostizieren. Dies ist besonders für die Fernsehsender, die am Wahlabend um Zuschaueranteile kämpfen, von großer Bedeutung. Hinzu kommt noch die Möglichkeit, am Wahlabend erklären zu können, wer wen warum gewählt hat. Allerdings werden die Exit-Poll-Daten auch nach dem Wahltag verwendet, um Wählerverhalten zu untersuchen und Wahlforschung zu betreiben. Nicht zuletzt, da die Wahltagsbefragungen vor der Veröffentlichung der Ergebnisse durchgeführt werden, sind sie von solch zentraler Bedeutung. Auf diese Weise kann nämlich der sogenannte „Bandwagon-Effekt“ vermieden werden. Das bedeutet, dass Menschen nach der Wahl bei Befragungen oftmals angeben, den Sieger gewählt zu haben, obwohl dies nicht der Fall war.

Manchmal haben Exit Polls auch noch eine dritte Aufgabe: Sie dienen der Kontrolle, ob eine Wahl korrekt abgelaufen ist. Oftmals führen ausländische Gruppen wie die „Organisation für Zusammenarbeit und Sicherheit in Europa“ (OSZE) in Staaten, in denen Wahlbetrug zu befürchten ist, Exit Polls durch, um eine zweite Quelle für die Wahlabsicht der Wähler zu haben. Da Wahltagsbefragungen wie andere Umfragen auch gewissen Fehlern unterliegen, können Differenzen zwischen den Exit Polls und den Stimmauszählungen allerdings lediglich ein Hinweis auf einen Wahlbetrug sein.

Exit Polls haben zahlreiche Vorteile gegenüber Umfragen vor der Wahl. So kann der Interviewer dadurch, dass er direkt vor dem Wahllokal steht, leicht mit der Zielperson Kontakt aufnehmen. Im Gegensatz zu Telefonbefragungen, wo ein Festnetzanschluss Grundvoraussetzung für eine Teilnahme ist, haben alle Wähler, die ihre Stimme persönlich abgeben, die Möglichkeit in die Stichprobe zu gelangen. Außerdem wird bei Exit Polls im Gegensatz zu anderen Wahlumfragen das große Problem umgangen, zwischen tatsächlichen und vermeintlichen Wählern unterscheiden zu müssen. Da Wählen eine sozial erwünschte Handlung ist, geben viele Befragte nicht zu, dass sie nicht zur Wahl gehen werden. Daher versucht man in den meisten Wahlbefragungen, mit Hilfe einiger Fragen herauszufinden, ob der Befragte tatsächlich zur Wahl gehen wird. Diese Frage stellt sich bei den Exit Polls nicht, da die Personen, die das Wahllokal verlassen, auch tatsächlich ihre Stimme abgegeben haben. Ein weiterer Vorteil der Wahltagsbefragung ist, dass keine Absicht, sondern ein tatsächliches Verhalten abgefragt wird. Insbesondere in Zeiten schwindender Parteibindung und immer kurzfristiger getroffener Wahlentscheidungen kann die zeitliche Nähe zum Wahlakt ein entscheidender Faktor sein. Dies trifft auch auf die Erinnerung der Wähler an ihre Wahlentscheidung zu. Es hat sich nämlich gezeigt, dass sogenannte

Recall-Fragen, die herausfinden sollen, wen der Befragte bei der letzten Wahl gewählt hat, oftmals recht unzuverlässig sind. Das gilt auch für taktische Wähler, die sich eigentlich einer anderen Partei verbunden fühlen, oder für Menschen, die sich wenig für Politik interessieren und daher eine kurzfristige Bauchentscheidung getroffen haben. Indem man die Wähler direkt nach der Abgabe ihrer Stimme zu ihrem Wahlverhalten befragt, wird die Wahrscheinlichkeit einer falschen Erinnerung auf ein Minimum reduziert (Roth 2008). In manchen Fällen erhält man aber trotzdem falsche Angaben, zum Beispiel, weil die Personen glauben, dass sie einem bestimmten Kandidaten ihre Stimme gegeben haben, aber in Wirklichkeit jemand anderen gewählt haben. Diesen Fall hatte man in den USA bei der Präsidentschaftswahl 2000 in Florida, wo einige Wähler aufgrund eines ungünstigen Fragebogendesigns statt Al Gore unbeabsichtigt Patrick Buchanan gewählt haben (Scheuren und Alvey 2008).

Schließlich ermöglichen Wahltagsbefragungen sehr hohe Fallzahlen zu vertretbaren Kosten. In Deutschland werden bei einer Bundestagswahl von jedem Meinungsforschungsinstitut in etwa 20.000 Menschen befragt, wohingegen die Fallzahl bei Meinungsumfragen normalerweise im Bereich zwischen 1.000 und 2.000 Interviews liegt (Hilmer und Hofrichter 2001). Dadurch können mit Hilfe von Wahltagsbefragungen auch gut die Anteile kleiner und radikaler Parteien ermittelt werden.

Allerdings haben auch Exit Polls Schwachpunkte. Obwohl immer mehr Menschen per Brief wählen, sind bei der klassischen Wahltagsbefragung nur Aussagen über Urnenwähler möglich. Außerdem ist es in den USA teilweise möglich, seine Stimme schon vor dem eigentlichen Wahltag abzugeben. Dies ist solange kein Problem, wie sich die „normalen“ Urnenwähler nicht von den Briefwählern und vorzeitigen Wählern unterscheiden. Das ist aber nicht immer der Fall. Bei der amerikanischen Präsidentschaftswahl 2000 in Florida zum Beispiel war der Anteil der Wähler, die für den republikanischen Kandidaten George W. Bush gestimmt haben, unter den Briefwählern wesentlich höher als unter den Urnenwählern. Eine weitere Schwäche ist, dass der Fragebogen relativ kurz und einfach sein muss, da er als Selbstausfüller konzipiert ist. Versuche mit unterschiedlichen Fragebogenlängen haben gezeigt, dass mit zunehmender Länge auch die Anzahl der Antwortverweigerer steigt. Antwortverweigerung ist ein Punkt, der zunehmend an Bedeutung gewinnt. Während man in Deutschland ganz annehmbare Ausschöpfungsquoten von 70% und mehr hat, liegt der Durchschnitt in den USA bei etwa 50% und teilweise sogar darunter (Hilmer und Hofrichter 2001, National Election Pool 2004).

## 3 Geschichte der Exit Polls ...

### 3.1 ... in den USA

Der erste Exit Poll wurde Frankovic (1992) zufolge 1940 in Denver durchgeführt, wo Wähler in Wahlkabinen, die außerhalb des Wahllokals aufgestellt worden sind, befragt wurden. Allerdings wurde diesem ersten Versuch wenig Aufmerksamkeit geschenkt. Ihr Durchbruch gelang den Exit Polls erst mit dem Aufkommen der Fernsehanstalten, die vom Wunsch nach schneller und umfassender Berichterstattung am Wahlabend getrieben die Entwicklung der Wahltagsbefragung forcierten und finanzierten.

Bis 1948 war das gängige Vorgehen, dass die Nachrichtenagenturen die Wahlergebnisse in der Wahlnacht und an den folgenden Tagen sammelten und sie den Zeitungen und Radiosendern übermittelten. Diese unterrichteten anschließend in den Tagen danach ihre Leser beziehungsweise Zuhörer über den Ausgang der Wahl. In den 1950er Jahren setzte man erstmals Computer am Wahlabend ein, um schon früher eine Tendenz angeben zu können. Dabei verarbeitete man die Stimmen aus den einzelnen Countys, die bis zu diesem Zeitpunkt ausgezählt waren,<sup>1</sup> in einem Regressionsmodell, um die Siegchancen der Kandidaten zu prognostizieren. Damit konnte zum Beispiel ausgesagt werden, dass die Chance John Kennedys, Richard Nixon zu schlagen, bei 3 zu 2 liegt.

Da es recht lange dauert, bis Ergebnisse auf County-Ebene vorliegen, begann der Fernsehsender CBS 1962 auf Grundlage von Ergebnissen in Wahlbezirken, die bereits ausgezählt sind, den Wahlsieger zu prognostizieren. Die Wahlbezirke, deren Wahlergebnisse verwendet wurden, wählte der Sender über eine Quotenstichprobe aus. Zumindest in den ersten Jahren konnten damit auch gute Resultate erzielt werden. Auf diese Weise konnte CBS unter anderem 1962 die Gewinner in 13 Gouverneurs- und Senatswahlen korrekt vorhersagen. Allerdings kam es 1966 zu Problemen. Dies lag jedoch nicht an der Stichprobe, sondern am implementierten Qualitätsprüfungsverfahren. Dabei wurden Bezirke, bei denen sich das aktuelle Wahlergebnis zu stark vom Wahlergebnis der letzten vergleichbaren Wahl unterschied, aus der Untersuchung ausgeschlossen. In Georgia und Maryland wurden bei den Gouverneurswahlen jedoch einige Ergebnisse nicht berücksichtigt, die eigentlich repräsentativ für die Wahl gewesen wären. Der Grund für diese Resultate war, dass in Wahlbezirken mit einem hohen Anteil afroamerikanischer

---

<sup>1</sup>In den USA bestehen die einzelnen Staaten aus verschiedenen Countys. Ihre Bedeutung für die Präsidentschaftswahl entspricht in etwa den deutschen Wahlkreisen. Die Countys werden in „precincts“ unterteilt, das sind die Wahlbezirke. Zu jedem Wahlbezirk gehört normalerweise ein Wahllokal, doch teilweise stimmen auch mehrere Bezirke am selben Ort ab. Die Stimmen werden entweder in den einzelnen Wahlbezirken ausgezählt und dann an das jeweilige County übermittelt oder es werden alle Stimmen eines Countys zentral ausgezählt.



Bewohner traditionell eher demokratisch gewählt wird. Bei diesen Wahlen vertraten jedoch die demokratischen Kandidaten konservativere Ansichten als ihre republikanischen Gegner. Dies führte bei den Wählern zu einem kollektiven Wechsel hin zu den moderaten republikanischen Kandidaten. Über das Qualitätsprüfungsverfahren wurde entschieden, diese Resultate zu vernachlässigen, was schließlich zu Fehlprognosen führte. In der Folge kam es dann zu zahlreichen Veränderungen, unter anderem der Einführung von wahrscheinlichkeitsbasierten Stichprobenverfahren.

Die erste „richtige“ Wahltagsbefragung wurde vom Fernsehsender CBS anlässlich der Gouverneurswahl 1967 in Kentucky durchgeführt. Es wurden keine Quotenstichproben mehr verwendet, sondern eine Zufallsauswahl aus den Wahlbezirken getroffen, wobei die Auswahlwahrscheinlichkeit proportional zur Anzahl der abgegebenen Stimmen bei einer vorausgangenen Wahl war. 1968 wurden erstmals anlässlich einer Präsidentschaftswahl Exit Polls durchgeführt. Dazu wurden Wähler in 20 Staaten befragt. Außerdem begann man 1968 zusätzlich Daten zu Geschlecht und Rasse zu erheben. In den 1970er Jahren führten dann auch NBC und ABC Exit Polls durch (Mitofsky und Edelman 2002).

Bis dahin gab es zwei Quellen für Wahlanalysen: Vorwahlbefragungen und die Analyse von sogenannten „tag precincts“. „Tag precincts“ sind Wahlbezirke, die mehrheitlich von einer bestimmten Bevölkerungsgruppe bewohnt werden. Beispielsweise schloss man aus dem Wahlverhalten in Bezirken mit mehrheitlich schwarzen Bewohnern auf das Wahlverhalten der schwarzen Bevölkerung. Es stellte sich jedoch heraus, dass Personen, die in einer homogenen Nachbarschaft leben, anders wählen als solche, die in Gegenden mit gemischter Bevölkerung wohnen. Dies ist besonders problematisch, da die Mehrheit der Bevölkerung in einer eher heterogenen Nachbarschaft lebt. Wie wenig aussagekräftig die Untersuchung von „tag precincts“ ist, hat eine Analyse der Exit-Poll-Ergebnisse für die Präsidentschaftswahl 1972 gezeigt. So schwankte der Anteil an Nixon-Wählern in der schwarzen Bevölkerung zwischen 6% in ärmeren innerstädtischen Vierteln und 34% in den wohlhabenderen Vororten. Empirische Untersuchungen haben gezeigt, dass die Untersuchung von „tag precincts“ nur für geographische Merkmale, wie die Unterscheidung zwischen Stadt und Land, brauchbare Ergebnisse liefert. Außerdem gibt es für manche Unterscheidungsmerkmale, zum Beispiel das Alter der Wähler, solche Gebiete gar nicht. Auch Vorwahlbefragungen sind für detaillierte Analysen nur begrenzt hilfreich. Denn in der Zeit zwischen der Vorwahlbefragung und der Wahl kann sich noch viel verändern, da der Wahlkampf weitergeht. Ein weiteres Problem stellen die geringen Fallzahlen dar, weil normalerweise nur zwischen 1.000 und 2.000 Personen befragt werden. Wenn man nun eine bestimmte Teilpopulation untersuchen möchte, beispielsweise Erstwähler, so wird der Anteil dieser Gruppe an den Befragten schnell sehr klein und es können keine gesicherten Aussagen getroffen werden (Mitofsky 1991). Mit den Daten aus den Exit Polls hatte man nun eine neue Quelle für die Analysen am Wahlabend.

Das war bis zu Beginn der 1980er Jahre auch der Hauptzweck der Wahltagsbefragungen. Um ausführlichere Analysen durchführen zu können, wurden längere Fragebögen entwickelt. Mit

diesen wurden neben Rasse und Geschlecht noch weitere Informationen erhoben, zum Beispiel Gründe für die Wahlentscheidung. Die Informationen über den Wahlausgang dagegen basierten ausschließlich auf der Hochrechnung von Wahlergebnissen in ausgewählten Bezirken. 1980 wagte dann der Sender NBC erstmals eine Prognose auf Basis einer Wahltagsbefragung und in den nächsten Jahren zogen auch die Sender ABC und CBS nach. Allerdings gab es starke Kontroversen, da NBC 1980 bereits vor Schließung der letzten Wahllokale den Sieg Ronald Reagans über Jimmy Carter verkündete. Anschließend wurden in mehreren Bundesstaaten Anti-Exit-Poll-Gesetze verabschiedet. Diese schrieben den Interviewern oftmals einen derart großen Abstand vom Wahllokal vor, dass sie nicht mehr erkennen konnten, wer gerade seine Stimme abgegeben hatte. Obwohl die Fernsehsender teilweise Erfolg mit ihren Klagen gegen diese Gesetze hatten, haben sie auch heute noch mit Regelungen zu kämpfen, die einen großen Abstand zwischen Interviewer und Wahllokal vorschreiben und damit die Arbeit der Interviewer enorm erschweren, wenn nicht sogar unmöglich machen. Nach dem Bericht von Edison Media Research und Mitofsky International (2005) ist dies einer der Gründe für die schlechten Resultate bei den Wahltagsbefragungen zur Präsidentschaftswahl 2004.

Da die Exit Polls aufgrund des Konkurrenzdrucks unter den Fernsehsendern immer kostspieliger wurden (um schnellere und bessere Ergebnisse zu liefern, mussten Befragungen in immer mehr Bezirken durchgeführt werden), arbeiten die wichtigsten Fernsehsender, Zeitungen und die Nachrichtenagentur „Associated Press“ (AP) seit 1993 unter dem Namen „Voter News Service“ (VNS) zusammen. Beim VNS handelt es sich um einen Zusammenschluss des „Voter Research and Surveys“ (VRS) und des 1964 gegründeten „News Election Services“ (NES). VRS war von den Fernsehsendern Ende der 1980er Jahre gegründet worden, um sich die Erhebung der Exit-Poll-Daten zu teilen und der NES war seit seiner Gründung für das Zusammentragen der ausgezählten Stimmen zuständig. Die Kräftebündelung durch die Zusammenarbeit der Medien ermöglichte es, sehr große Umfragen durchzuführen. Allerdings hat es den Nachteil, dass die Ergebnisse weniger verlässlich sind, da sie nicht überprüft werden können. Lediglich die LA Times führt noch unabhängig Wahltagsbefragungen durch, jedoch im wesentlich kleineren Stil. Problematisch war es anfangs auch, sich auf ein gemeinsames Vorgehen zu einigen, da sich die Arbeitsweisen der verschiedenen Fernsehanstalten voneinander unterschieden. Dies betrifft auf der einen Seite die Auswahl der Stichprobenbezirke und auf der anderen Seite auch die Konzeption des Fragebogens. Es war vorgesehen, dass der VNS die Daten sammelt, diese aufbereitet und bei einer eindeutigen Datenlage einen Kandidaten zum Sieger erklärt. Doch 1994 lies ABC die Ergebnisse aus der Wahltagsbefragung von einem eigenen Team analysieren und erklärte einige Rennen schon vor dem VNS für entschieden. Die anderen Fernsehsender, die 1994 noch auf den VNS warten mussten, wenn sie jemanden zum Sieger erklären wollten, legten sich in den folgenden Jahren auch eigene „Entscheidungsteams“ zu (Konner 2003, Mitofsky und Edelman 2002).

Allerdings wurde der VNS nach dem schlechten Abschneiden in den Jahren 2000 und 2002 aufgelöst und seitdem werden die Exit Polls vom „National Election Poll“ (NEP) durchgeführt

und die ausgezählten Stimmen auf County-Ebene von der Associated Press gesammelt.

### 3.2 ... in Deutschland

Die Wahlberichterstattung der deutschen Fernsehsender beginnt mit den ersten computergetstützten Hochrechnungen bei der Bundestagswahl 1965. Dazu wurden Wahlkreisergebnisse verwendet, weshalb die ersten Hochrechnungen erst um 21 Uhr vorlagen. In den folgenden Jahren gingen die beiden Fernsehanstalten dann dazu über, die Daten auf Wahlbezirksebene hochzurechnen. Dies führte zu einer Beschleunigung der Wahlberichterstattung. Die Einführung von Wahltagsbefragungen erfolgte schließlich durch das Meinungsforschungsinstitut Infas, mit dem die ARD bis 1996 kooperierte. Infas arbeitete mit Forschern aus den USA zusammen und adaptierte die dort verwendeten Methoden für Deutschland. Das ZDF arbeitete zwischen 1964 und 1974 mit einer Gruppe Politikwissenschaftler der Uni Mannheim zusammen. Daraus entstand schließlich die Forschungsgruppe Wahlen (FGW), die bis heute die Wahltagsbefragungen für das ZDF durchführt. Die ARD dagegen arbeitet seit 1997 mit Infratest dimap zusammen. 1976 wurden erstmals Wähler am Wahltag befragt, die Ergebnisse wurden aber ausschließlich zu Analysezwecken verwendet. Volle Anwendung fanden die Exit Polls 1978 bei der Berichterstattung zu den Landtagswahlen in Hamburg und Niedersachsen. Nach Schließung der Wahllokale konnte auf Basis der Wahltagsbefragungen das Ergebnis recht exakt prognostiziert werden. Daraufhin kam es zu einer öffentlichen Debatte über die Exit Polls und inwiefern diese Daten während des Wahltags missbraucht werden könnten. Schließlich wurde sogar ein Verbot der Wahltagsbefragung diskutiert. Die beiden Nachrichtensender ARD und ZDF entschieden sich letztendlich für eine Selbstbeschränkung und stellten die Wahltagsbefragungen für über ein Jahrzehnt ein.

Seit Anfang 1990 sind Exit Polls jedoch wieder fester Bestandteil der Wahlsendungen, nachdem ein Veröffentlichungsverbot vor 18 Uhr in die Wahlgesetze der Länder und des Bundes aufgenommen worden war (Roth 2008). Zum Durchbruch haben den Wahltagsbefragungen auch die guten Prognosen für die Bundestagswahl 1990 verholfen. Dies war ein besonders kompliziertes Unterfangen, da für Ostdeutschland keine alten Wahlergebnisse vorlagen, die man als Grundlage für gebundene Hochrechnungen hätte verwenden können. Zusätzlich hatten die Vorwahlumfragen einen Sieg der SPD vorhergesehen, während die Exit Polls den korrekten Sieg der Unionsparteien prognostizierten. Seither werden die Daten aus der Wahltagsbefragung zur Prognose und zur detaillierten Beschreibung des Wahlergebnisses benutzt (Hilmer und Hofrichter 2001).

### 3.3 ... in Großbritannien

In Großbritannien wurden Exit Polls schon früher als in den USA zur Prognose verwendet. Der Grund dafür ist, dass die britischen Fernsehsender nicht die Möglichkeit haben, die Ergebnisse von bereits ausgezählten Wahlbezirken hochzurechnen, da die Stimmen nur auf Wahlkreisebene ausgezählt werden.<sup>2</sup> Die Geschichte der Exit Polls in Großbritannien beginnt mit dem Nachrichtenanbieter ITN, der die Ergebnisse dem Fernsehsender ITV zur Verfügung stellte, und dem Meinungsforschungsinstitut Harris. Diese führten 1974 erstmals Wahltagsbefragungen durch. ITN verfolgt den Ansatz, in erster Linie in umkämpften Wahlkreisen Interviews durchzuführen und zwar bevorzugt in solchen, in denen diejenige Partei regierte, die in den Umfragen zurücklag. Viele Wahlkreise in Großbritannien werden von einer Partei dominiert und Wechsel dort sind sehr unwahrscheinlich. Da die Hauptaufgabe darin besteht, für jede Partei die Anzahl der Sitze zu prognostizieren, konzentriert man sich auf Wahlkreise, in denen es bei der letzten Wahl ein knappes Ergebnis gab. Dabei kann „knapp“ aber durchaus 16 Prozentpunkte Differenz zwischen den beiden führenden Parteien bedeuten. ITN prognostizierte auf Grundlage des ersten Exit Polls zwar eine Labourmehrheit mit 15 Sitzen, während es tatsächlich nur drei Sitze Vorsprung wurden, allerdings sagte BBC auf Basis von Vorwahlbefragungen einen Labourvorsprung von 137 Sitzen voraus. BBC behielt die Wahlprognose auf Grundlage von Vorwahlbefragungen bis 1992 bei, wobei der Sender zwar in manchen Jahren recht exakte Ergebnisse erzielte (auf zwei Sitze genau bei der Wahl 1983), in vielen Jahren jedoch beträchtlich daneben lag. Nachdem die Prognosen für die Wahl 1987 wieder einmal sehr ungenau waren, entschloss sich auch BBC, bei der Wahl 1992 Exit Polls zu verwenden. Während ITN noch immer mit Harris zusammenarbeitete, begann BBC die Zusammenarbeit mit dem „National Opinion Poll“ (NOP). In den Jahren zwischen 1987 und 1992 führten NOP und BBC bei kleineren Wahlen Experimente mit verschiedenen Fragebogenlängen und auch mit anonymen oder nicht anonymen Befragungen durch. Schließlich konnte der durchschnittliche Fehler auf weniger als 0.5% pro Partei reduziert werden. Allerdings erzielte man bei den landesweiten Wahlen 1992 keine guten Ergebnisse. Obwohl lediglich ein knapper Sieg der Konservativen prognostiziert worden war, siegten diese mit 65 Sitzen Vorsprung vor Labour. Allerdings wich auch die Sitzprognose von ITN, mit einem Vorsprung der Konservativen von elf Sitzen, bei dieser Wahl deutlich vom tatsächlichen Ergebnis ab (Matthias und Cowling 1995). Es ist jedoch eine unzulässige Aussage, die Exit Polls hätten versagt, da nur die Prognose der Sitze betrachtet wurde. Tatsächlich war ein großer Teil der Abweichung auf die fehlerbehafteten Modelle zurück zu führen, die zur Vorhersage verwendet wurden. Außerdem wurden für einen Teil der Sitze Expertenmeinungen eingeholt und auf diese Weise entschieden, welche Partei den jeweiligen Sitz erhalten wird. Diese Vorgehensweise

---

<sup>2</sup>Großbritannien ist ebenfalls in Wahlkreise unterteilt, die sogenannten „constituencies“. In jedem Wahlkreis wird über einem Sitz im Unterhaus abgestimmt. Die Partei mit der Mehrheit der Sitze im Unterhaus stellt die Regierung. Die „constituencies“ bestehen aus mehreren „wards“. Das ist die Unterteilung für lokale Wahlen. Die „wards“ bestehen wiederum aus „polling districts“, die am ehesten den deutschen Wahlbezirken entsprechen.

ist für solche Sitze üblich, die aus irgendeinem Grund als „speziell“ angesehen werden, zum Beispiel, weil dort ein Kandidat erneut antritt, der in der Zwischenzeit die Partei gewechselt hat. Man geht davon aus, dass sich solche Sitze anders verhalten und daher das Ergebnis nicht über das Standardverfahren prognostiziert werden kann. Deshalb verlässt man sich in solchen Fällen auf Experten. 1992 orientierte sich die Entscheidung der Experten allerdings stark an den Vorwahlbefragungen, welche eine deutliche Führung der Labour-Partei sahen. So kam es dann, dass viele dieser „special seats“ bei der Wahlprognose der Labour-Partei zugeschlagen wurden, schlussendlich aber an die Konservativen gingen.

Bei den Wahlen 1997 gelangen dann beiden Gruppen (BBC mit NOP und ITN arbeitete inzwischen mit Market and Opinion Research International (MORI) zusammen) wieder bessere Prognosen. Während BBC den Labour-Vorsprung um rund 20 Sitze überschätzte, lag die ITN-Schätzung circa 20 Sitze unter dem tatsächlichen Resultat. Allerdings war dies nicht unbedingt der Verdienst exakter Exit Polls oder guter Modelle. So hätten die Modelle eigentlich einen zu geringen Anteil für Labour vorausgesagt. Da die Wahltagsbefragungen aber einen größeren Labouranteil ergeben hatten als dies tatsächlich der Fall war, hoben sich die beiden Fehler auf und man erhielt eine vertretbare Prognose.

2004 beendeten die BBC und ITN/ITV, vornehmlich aus Kostengründen, ihre Konkurrenzsituation und arbeiten seitdem recht erfolgreich zusammen (Moon 1999, Curtice und Firth 2008).

## 4 Methodik für die USA, Großbritannien und Deutschland

Wahltagsbefragungen stellen eine logistische Meisterleistung dar, denn es müssen an einem Tag tausende Interviews durchgeführt und ausgewertet werden. Dies ist nur durch gute Planung im Vorfeld und automatisierte Abläufe am Wahltag möglich. Obwohl bei einer Wahltagsbefragung eine sehr große Anzahl an Wählern befragt wird, handelt es sich nach wie vor um eine Stichprobe. Daher ist es eine der größten Herausforderungen vor dem Wahltag, eine repräsentative Stichprobe zu ziehen, die am Wahlabend korrekte Schlüsse auf die Grundgesamtheit ermöglicht.

### 4.1 Stichprobenziehung

#### 4.1.1 Stichprobengröße

Allgemein gilt, dass in Ländern mit Mehrheitswahlrecht eher größere Stichproben in Bezug auf die Anzahl der Wahlbezirke notwendig sind. Dies kommt daher, dass man Daten auf der jeweiligen Ebene benötigt, auf der die Sitze vergeben werden. Beim Verhältniswahlrecht dagegen kann man auch mit Daten auf nationaler Ebene arbeiten.

Abhängig von der Größe eines Landes, dem dort geltenden Wahlrecht und der Komplexität des Wahlvorganges variiert die Stichprobengröße. Bei den Präsidentschaftswahlen 1992 in den USA beispielsweise wurden Interviews in 1.310 Wahlbezirken durchgeführt und dabei ungefähr 177.000 Wähler befragt (Mitofsky und Edelman 1995). In Deutschland hat man bei Landtagswahlen zwischen 150 und 200 Bezirke und es werden insgesamt etwa 5.000 bis 10.000 Wähler befragt, wobei die Fallzahlen in letzter Zeit vor allem bei Infratest dimap deutlich erhöht wurden (Hilmer und Hofrichter 2001, Infratest dimap 2009). Bei Bundestagswahlen werden von Infratest dimap rund 400 Wahlbezirke mit Interviewern besetzt und von der Forschungsgruppe Wahlen ungefähr 250. In den ausgewählten Wahlbezirken werden von jedem Institut Daten von über 20.000 Wählern erhoben (z.B. Forschungsgruppe Wahlen 2005, Infratest dimap 2005). Diese Stichprobengröße (und das Vorgehen bei der Auswahl derselben) ermöglicht getrennte Schätzungen für die alten und die neuen Bundesländer. Dies ist nötig, da sich diese beim Wahlverhalten noch immer systematisch unterscheiden. Wenn man Schätzungen für bestimmte kleinere Einheiten möchte, zum Beispiel für Städte, so muss die Stichprobengröße entsprechend erhöht werden, da für eine fundierte Prognose mindestens zwischen 25 und 30 Wahlbezirke Voraussetzung sind (Hilmer 2008).

Ferner kann die Auswahl der Bezirke auch noch von weiteren Aspekten des Wahlrechts beeinflusst werden. So gibt es in Deutschland für eine Partei neben der Überwindung der 5%-Hürde auch noch die Möglichkeit, über den Gewinn von drei Direktmandaten in den Bundestag einzuziehen. 1994 hatte die PDS nur über ihre vier gewonnenen Direktmandate den Einzug geschafft,

obwohl sie lediglich 4,4% der Stimmen erhalten hatte. Um bei der Bundestagswahl 1998 eine Prognose über einen erneuten Einzug der PDS in den Bundestag abgeben zu können, wurden Wahlkreise, in denen die Möglichkeit bestand, dass die PDS ein Direktmandat bekommen könnte, mit zusätzlichen Interviewern ausgestattet (Hilmer und Hofrichter 2001).

#### 4.1.2 Stichprobenziehung

Der erste Schritt bei der Durchführung von Exit Polls ist die Stichprobenziehung. Aufgrund ökonomischer und statistischer Überlegungen hat man sich für ein zweistufiges Auswahlverfahren entschieden. Dazu werden die Befragungen in einer repräsentativen Auswahl an Stimmbezirken durchgeführt, in welchen wiederum eine Wählerstichprobe gezogen wird. Dieses Vorgehen hat mehrere Vorteile. So können auf diese Weise mit einem verhältnismäßig geringem Aufwand eine hohe Anzahl an Interviews durchgeführt werden. Außerdem können die Ergebnisse aus den Wahlbezirken relativ zügig mit den ausgezählten Stimmen abgeglichen werden, wodurch Verzerrungen und Fehler entdeckt werden können. Allerdings hat dieses Vorgehen auch einen entscheidenden Nachteil, nämlich die hohe Clusterung. Clusterstichproben sind dann besonders gut, wenn sich die Cluster möglichst ähnlich sind und ein verkleinertes Abbild der Grundgesamtheit sind. Dazu müssen die Cluster selbst möglichst heterogen sein und idealerweise alle Fälle abdecken. Man kann jedoch nicht davon ausgehen, dass die einzelnen Wahlbezirke ein repräsentatives Abbild der Grundgesamtheit darstellen. Allerdings ermöglicht es der Effizienzgewinn durch die Clusterung, eine wesentlich größere Stichprobe zu befragen, als dies bei einer einfachen Zufallsstichprobe der Fall wäre. Indem man verschiedenste Stimmbezirke in die Stichprobe zieht, erhält man auch die unterschiedlichen Aspekte der Grundgesamtheit.

Im Folgenden wird die Auswahl der Stimmbezirke exemplarisch für die USA dargestellt.

In den USA finden am Tag der Präsidentschaftswahl noch zahlreiche weitere Wahlen statt, für die in der Wahltagsbefragung ebenfalls Daten erhoben werden sollen. Dies führt zu einem recht komplexen Stichprobendesign. Dabei werden erst für jeden Staat einzeln Stichprobenbezirke gezogen. Aus diesen Bezirken wird wiederum ein Teil ausgewählt, die nationale Stichprobe, wo spezielle Fragebögen für die Präsidentschaftswahl an die Wähler ausgeteilt werden. Es wird jedoch auch bei den „state samples“ abgefragt, für welchen Präsidentschaftskandidaten der Wähler gestimmt hat. Zusätzlich werden aber noch Wahlentscheidungen bei Gouverneurs- und Senatswahlen abgefragt. Es ist sinnvoll, dass die Ziehung der Stichprobe staatenweise erfolgt, da der nächste amerikanische Präsident derjenige Kandidat wird, der die meisten Wahlmänner gewonnen hat. Diese Wahlmänner werden jedoch auf Staatenebene vergeben (Levy 1983, Mitofsky und Edelman 1995).

Bevor die Stichprobenbezirke gezogen werden, wird eine Schichtung der Wahlbezirke mit Hilfe verschiedener Merkmale, die für den Wahlausgang von Bedeutung sind, durchgeführt. Durch Schichtung versucht man, eine sehr einseitige Stichprobe zu vermeiden. Typische Schichtungsmerkmale sind geographische Eigenheiten und die Stimmanteile der Parteien bei einer vorausgegangenen Wahl, die mit der aktuellen Wahl vergleichbar ist. Ziel ist es, dass die Auswahl der

Wahlbezirke die verschiedenen geographischen Teile des Staates und die Stimmanteile der Demokraten und Republikaner bei der Vergleichswahl widerspiegelt. Aus den einzelnen Schichten wird dann eine bestimmte Anzahl von Stichprobenbezirken gezogen. In den USA werden die Wahlbezirke dazu mit einer Auswahlwahrscheinlichkeit versehen, die entweder proportional zur Anzahl der abgegebenen Stimmen in einem bestimmten Jahr oder zur Anzahl der registrierten Wähler ist. Dadurch werden Bezirke mit vielen Wählern mit einer größeren Wahrscheinlichkeit gezogen als solche mit einer kleinen Wählerschaft. Diese ungleichen Auswahlwahrscheinlichkeiten werden dadurch ausgeglichen, dass die Wähler in den ausgewählten Bezirken mit vielen Wählern eine geringere Wahrscheinlichkeit haben, befragt zu werden. Auf diese Weise wird eine möglichst gleich große Befragungswahrscheinlichkeit für alle Wähler gewährleistet. Wenn Stimmbezirke annähernd dieselbe Größe haben, kann auch eine Zufallsauswahl aus den Schichten erfolgen.

Außerdem bekommen manche Bezirke eine höhere Auswahlwahrscheinlichkeit zugewiesen. Dies ist zum Beispiel bei Bezirken der Fall, die über einen großen Anteil an schwarzen Wählern verfügen. Folglich sind solche Bezirke in der Stichprobe überrepräsentiert. Um dafür einen Ausgleich zu schaffen, werden sie bei der Auswertung heruntergewichtet. Die Stichprobenziehung berücksichtigt zudem verschiedene weitere Faktoren wie die Zeitzone oder das Vorhandensein von Wahlmaschinen. Bei der Stichprobenziehung muss auch in Betracht gezogen werden, dass sich die Bedingungen seit der letzten Wahl verändert haben könnten, zum Beispiel durch Umzüge oder Veränderung von Wahlbezirksgrenzen. Dies kann dazu führen, dass eine bestimmte soziodemographische Gruppe oder geographische Region überproportional stark vertreten ist. Daher müssen auch immer umfangreiche Recherchen betrieben werden, um eventuell vorhandene Veränderungen zu entdecken (Levy 1983).

Die Anzahl der ausgewählten Bezirke richtet sich danach, ob eine knappe Entscheidung erwartet wird und wie groß in Amerika die Staaten beziehungsweise in Deutschland die Bundesländer sind, in denen die Befragung durchgeführt wird. In den USA hat man in Staaten, in denen ein eindeutiges Ergebnis erwartet wird oder die sehr klein sind, etwa 20 Stichprobenbezirke ausgewählt. In größeren oder sehr umkämpften Staaten sind es meist mehr, zum Beispiel 50 Bezirke in Ohio bei der Präsidentschaftswahl 2004 (Scheuren und Alvey 2008).

Bis Ende der 1980er Jahre verwendeten die amerikanischen Fernsehanstalten und Zeitungen unterschiedliche Methoden zur Stichprobenziehung. So war bei CBS die Exit-Poll-Stichprobe eine Zufallsauswahl aus den Stichprobenbezirken, die dazu dienten, ausgezählte Stimmen bereit zu stellen, welche ebenso wie die Exit-Poll-Daten in die Modelle eingespeist wurden. ABC dagegen verwendete ein mehrstufiges wahrscheinlichkeitsbasiertes Design, das an Telefoninterviews gekoppelt wurde. Auch die Anzahl der Stichprobenbezirke unterschied sich zwischen den Fernsehsendern. Im Jahr 1980 umfasste die Stichprobe von CBS 150, von ABC 375 und von NBC 400 Wahllokale. Da die Exit Polls inzwischen gemeinschaftlich durchgeführt werden, musste man sich auch auf ein einheitliches Vorgehen bei der Stichprobenziehung verständigen (Levy 1983).



Bei der Auswahl der Wähler handelt es sich um eine systematische Auswahl. Ziel ist es, dass jeder Wähler, der in diesem Bezirk persönlich zur Wahl geht, die gleiche Wahrscheinlichkeit hat, in die Stichprobe zu gelangen. Dazu wird für jeden Stimmbezirk in der Stichprobe ein Befragungsintervall  $k$  festgelegt. Ausgehend von einem bestimmten Wähler, der das Wahllokal verlässt, wird dann jeder  $k$ -te Wähler befragt. Das Befragungsintervall kann unterschiedlich groß sein. So variierten die Intervalle bei der Präsidentschaftswahl 1980 in den USA zwischen jedem Wähler in extrem kleinen Stimmbezirken und jedem 70. Wähler in sehr großen Bezirken (Levy 1983). 2004 hatte man mit einem  $k$  zwischen 1 und 10 eine wesentlich geringere Varianz in der Schrittweite (Edison Media Research und Mitofsky International 2005). In den USA versucht man  $k$  so zu bestimmen, dass in etwa 100 Wähler pro Stichprobenbezirk befragt werden. Allerdings beruhen diese Berechnungen auf einer Schätzung der Wahlbeteiligung und so kann es passieren, dass die Anzahl der durchgeführten Interviews auch einmal deutlich unter oder über 100 Personen liegt. So wurden zum Beispiel bei der amerikanischen Präsidentschaftswahl 1992 im Durchschnitt 134 Wähler pro Wahllokal interviewt (Mitofsky und Edelman 1995). Wenn man im Laufe des Wahltages feststellt, dass ein schlecht praktikierbares Befragungsintervall festgelegt wurde, wird es manchmal geändert. Allerdings hat Klorman (1976) gezeigt, dass Wähler mit unterschiedlichen soziodemographischen Merkmalen zu verschiedenen Tageszeiten wählen. Daher wird man höchstwahrscheinlich eine verzerrte Stichprobe erhalten, wenn man das Intervall während der Befragung anpasst. Allerdings kann auch das Beibehalten des Befragungsrhythmus zu Problemen führen und zwar dann, wenn der Interviewer sehr viele Personen verpasst, die er eigentlich befragen müsste, weil zu viele Personen das Wahllokal verlassen. Alternativ zu einer PPS-Stichprobe wäre es auch denkbar, alle Wahlbezirke mit gleicher Wahrscheinlichkeit zu ziehen. Wenn bei dieser Variante für jeden Wähler die Wahrscheinlichkeit, in die Stichprobe zu gelangen, gleich groß sein soll, dann muss in jedem Wahlbezirk dasselbe Befragungsintervall verwendet werden. Es wird also beispielsweise in jedem Bezirk, der in die Stichprobe gelangt ist, jeder zehnte Wähler befragt. Dies ist solange kein Problem, wie sich die Wahlbezirke nicht allzu sehr in ihrer Größe unterscheiden. Wenn sich die Wahlbezirke jedoch deutlich in der Anzahl ihrer Wähler unterscheiden, hat dies zur Folge, dass bei einem recht klein gewählten Intervall die Interviewer in Bezirken mit vielen Wählern wahrscheinlich Personen verpassen, die befragt werden sollten. Oder wenn das Intervall relativ groß gewählt wird, bezahlt man in kleinen Bezirken einen Interviewer für den ganzen Tag, obwohl dieser nur ein paar Interviews durchführt (Moon und McGregor 1992).

## 4.2 Fragebogen und Durchführung der Befragung

Bei der konkreten Durchführung tritt der Interviewer an jeden  $k$ -ten Wähler heran und bittet ihn, einen Fragebogen auszufüllen. Falls die angesprochene Person eine Teilnahme ablehnt, wird in Deutschland das ungefähre Alter und das Geschlecht notiert und in den USA zusätzlich

die Rasse. Diese Angaben werden benötigt, um bei der Auswertung eine Ausfallgewichtung durchzuführen (Hilmer und Hofrichter 2001). Anschließend befolgt der Interviewer wieder sein Befragungsintervall. Es gibt jedoch auch andere Möglichkeiten, um mit Antwortverweigerung umzugehen. So kann zum Beispiel die Person befragt werden, die auf den Antwortverweigerer folgt oder der nächste Wähler mit denselben Merkmalen in Bezug auf Alter und Geschlecht wie die Person, die eine Teilnahme abgelehnt hat. Diese Vorgehensweise war bei MORI in Großbritannien vor der Zusammenarbeit mit NOP üblich (Curtice und Firth 2008).

Der verwendete Fragebogen ist als Selbstausfüller konzipiert, so dass der Befragte das Ausfüllen selbstständig erledigen können sollte. Dies und die Abgabe des Fragebogens in eine Urne sorgen dafür, dass die Angaben der Befragten anonym bleiben. In Großbritannien wurden bei den BBC-Wahltagsbefragungen anfangs persönliche Interviews durchgeführt. Nachdem man jedoch festgestellt hatte, dass man durch eine anonyme Befragung den Anteil der Antwortverweigerer verringern konnte, wechselte man zu selbstständig ausfüllbaren Fragebögen. Damit gelang es auch, die Überrepräsentanz von Labour-Wählern in der Stichprobe etwas einzudämmen (Moon und McGregor 1992). Obwohl sich die meisten Experten darüber einig sind, dass ein anonymen Fragebogen die verlässlichsten Resultate liefert (z. B.: Bishop und Fisher 1995), gibt es doch Ausnahmen von dieser Regel. Denn in Ländern mit hoher Analphabetenrate ist zu erwarten, dass man mit einem Fragebogen, der selbstständig ausgefüllt werden muss, verzerrte Ergebnisse bekommt. Daher hat man sich in Mexiko für ein gemischtes Erhebungsverfahren entschieden. Dabei stellt der Interviewer dem Befragten zuerst einige Fragen zu demographischen Fakten, sozialem Status, höchster Schulbildung und Fragen zu politischen Themen. Anschließend füllt der befragte Wähler anonym einem simulierten Stimmzettel aus, der exakt dem offiziellen entspricht, und wirft ihn in eine Box. Schließlich werden im persönlichen Interview noch weitere demographische Daten erhoben (Bautista et al. 2008).

Der Exit Poll-Fragebogen kann unterschiedliche Fragen enthalten. Immer abgefragt wird die aktuelle Wahlentscheidung und Variablen, die als wichtig für die Wahlentscheidung angenommen werden, wie Alter und Geschlecht. Darüber hinaus können auch Dinge wie Gewerkschafts- und Konfessionszugehörigkeit oder entscheidende Wahlkampfthemen erfragt werden. In Deutschland enthält der Fragebogen meist weniger Fragen als in den USA, da man in Deutschland versucht, manche Fragen schon in Vorwahlbefragungen zu klären, um so die Wahltagsbefragung zu entlasten und auf diese Weise die Ausschöpfungsquote zu erhöhen. Daher müssen die Wähler in den USA in der Regel zwischen 20 und 50 Fragen beantworten, in Deutschland dagegen nur zwischen 10 und 20. Diese Fragen werden auf die Vorder- und Rückseite eines DinA4-großen Zettels gedruckt (Hofrichter 1999). In den USA gibt es außerdem unterschiedliche Fragebögen, unter anderem wegen der Kooperation der Fernsehsender. Da jeder auch individuelle Fragen stellen möchte, ergibt sich ein äußerst umfangreicher Fragenkatalog. Wenn dieser von jedem Befragten beantwortet werden müsste, würde die Antwortbereitschaft noch weiter sinken. Die Lösung ist, dass an jedem ausgewählten Wahllokal in der nationalen Stichprobe unterschiedliche Fragebögen in einer fest vorgegebenen Reihenfolge verteilt werden. In den Vereinigten Staaten müssen für

die Wahltagsbefragungen in den Staaten sehr viele verschiedene Fragebogenversionen gedruckt werden, da auch Fragen zu Wahlentscheidungen bei Gouverneurs- und Senatswahlen gestellt werden. Man hat nun versucht, die Anzahl der Fragebogenvarianten zu reduzieren und ist dazu übergegangen, bei den Kandidaten, die für das Repräsentantenhaus kandidieren, nur den Namen der Partei anzugeben, da sonst für jeden Wahlbezirk ein eigener Fragebogen erstellt werden hätte müssen. Dies hat allerdings dazu geführt, dass besonders beliebte Kandidaten bei der Wahltagsbefragung oftmals wesentlich schlechter abschnitten als es tatsächlich der Fall war. Dies liegt daran, dass die Menschen sich oftmals für solche Kandidaten nicht aufgrund ihrer Parteizugehörigkeit entscheiden, sondern wegen deren Persönlichkeit und daher auch nur die Namen der Kandidaten kennen, aber nicht wissen, zu welcher Partei sie gehören. Allerdings nimmt man dieses Problem in Kauf, da man sonst einen immensen Mehraufwand hätte. Bei der Präsidentschaftswahl 1992 konnte auf diese Weise die Anzahl der unterschiedlichen Fragebögen von ungefähr 1.200 auf 204 reduziert werden (Mitofsky und Edelman 1995).

Die Ergebnisse der Exit Polls werden in regelmäßigen Abständen an die Zentrale des durchführenden Meinungsforschungsinstituts weitergeleitet. Nach der Schließung des Wahllokals ist es auch noch die Aufgabe des Interviewers, der Stimmauszählung beizuwohnen und die Ergebnisse des jeweiligen Wahlbezirkes an die Zentrale weiterzugeben (Hilmer und Hofrichter 2001).

### 4.3 Schätzung und Prognose

Nachdem die Daten aus der Wahltagsbefragung an die Zentrale übermittelt worden sind, beginnt man dort, diese zu analysieren. Aufgrund des großen Zeitdrucks werden die Daten jedoch normalerweise nicht individuell beurteilt, sondern mit automatischen Prozeduren überprüft. Dasselbe gilt für die Gewichtung und Auswertung. Auch hier werden vorgefertigte Algorithmen verwendet, um die Datenmengen in einem akzeptablen Zeitraum bewältigen zu können.

Direkt nach Schließung der Wahllokale können in manchen Fällen bereits die Sieger einer Wahl bekannt gegeben werden. Dies ist der Fall, wenn der Ausgang so eindeutig ist, dass man nur auf Grundlage der Exit-Poll-Daten bereits eine Entscheidung treffen kann. Meist müssen jedoch noch die Ergebnisse der Stimmauszählung aus einigen Stichprobenbezirken abgewartet werden. Dabei handelt es sich um die Wahlbezirke, in denen Wahltagsbefragungen durchgeführt worden sind, sowie um einige zusätzlich ausgewählte Bezirke. Die Ergebnisse der Stimmauszählung werden dann entweder von der Person weitergeleitet, die die Wahltagsbefragung durchgeführt hat, oder von einem Reporter, der speziell dafür in den Wahlbezirk geschickt worden ist, falls in einem ausgewählten Bezirk kein Exit Poll stattgefunden hat. Diese Daten können einerseits zur Kontrolle der Ergebnisse aus den Wahltagsbefragungen verwendet werden. Andererseits werden sie auch in den Prognosemodellen und zur Gewichtung verwendet. Im Laufe des Abends werden dann zunehmend Exit-Poll-Daten durch ausgezählte Stimmen ersetzt. In den meisten Fällen kann auf diese Weise der Wahlsieger bestimmt werden. Wenn der Wahlausgang besonders knapp ist, muss noch abgewartet werden, bis die Wahlkreise (fast) vollständig ausgezählt

sind (Mitofsky und Edelman 2002).

#### 4.3.1 Gewichtung

Ein wichtiger Aspekt bei der Arbeit mit den Daten aus Exit Polls ist die Gewichtung. 2004 beispielsweise wurden Daten im Internet veröffentlicht, die eine mehr oder weniger große Führung John Kerrys zeigten. Als am Wahlabend dann der Amtsinhaber George Bush zum Sieger erklärt wurde, gab es Zweifel an der Rechtmäßigkeit dieser Wahl. Der National Election Pool, der für die Durchführung und Analyse der Exit Polls zuständig war, erklärte allerdings, dass es sich bei den Zahlen im Internet um ungewichtete Daten gehandelt hat. Da die Demokraten bei Wahltagsbefragungen traditionell überrepräsentiert sind, musste das Ergebnis von Kerry heruntergewichtet werden, um dem tatsächlichen Anteil zu entsprechen. Dieses Beispiel zeigt anschaulich, dass die Ergebnisse von Exit Polls ohne Gewichtung oftmals wenig brauchbar sind. Dabei gibt es mehrere Faktoren, die durch eine Gewichtung ausgeglichen werden sollen. Dazu gehören unter anderem ungleiche Auswahlwahrscheinlichkeiten und Antwortverweigerung. Levy (1983) beschreibt eine mögliche Vorgehensweise bei der Gewichtung am Beispiel von CBS bei der Präsidentschaftswahl 1980.

#### Vorgehen bei Gewichtung am Beispiel von CBS

CBS verwendete drei verschiedene Gewichte: „within-precinct weights“, „precinct weights“ und ein Gewicht, um den Stichprobenfehler zu verringern. Dabei tragen die ersten beiden Gewichte den unterschiedlichen Auswahlwahrscheinlichkeiten bei der Stichprobenziehung Rechnung. Durch die „within-precinct weights“ wurden die unterschiedlichen Auswahlwahrscheinlichkeiten der einzelnen Wähler und die verschiedenen Non-Response-Raten innerhalb der Wahlbezirke berücksichtigt. Denn die Wähler in den ausgesuchten Stimmbezirken hatten je nach Größe des Wahlbezirks eine andere Wahrscheinlichkeit, in die Stichprobe zu gelangen, da in Wahlbezirken mit einer höheren Anzahl an Wählern größere Befragungsintervalle verwendet wurden. Beim Non-Response im Zusammenhang mit amerikanischen Exit Polls handelt es sich sowohl um den klassischen Non-Response, wenn ein angesprochener Wähler die Teilnahme am Exit Poll verweigert, als auch um den Umstand, dass der Interviewer einen Wähler verpasst, den er eigentlich befragen hätte sollen. Zur Berechnung des „within-precinct“-Gewichts wurde die Anzahl der abgegebenen Stimmen in einem Wahlbezirk durch die Anzahl der durchgeführten Interviews geteilt. Ersteres wird nach Schließung der Wahllokale über die während des Tages abgegebenen Stimmen bestimmt. Wenn die Wahl noch läuft, dann wird der Zähler mit Hilfe historischer Daten und bisheriger Beobachtungen zur Wahlbeteiligung geschätzt.

Die „precinct weights“ wurden auf die bereits gewichteten Bezirksdaten angewandt. Diese Gewichtung diente dazu, die unterschiedlichen Auswahlwahrscheinlichkeiten der einzelnen Bezirke zu berücksichtigen. Bei CBS wurden bei der Präsidentschaftswahl 1980 die Auswahlwahrscheinlichkeiten für die Wahlbezirke auf Basis der abgegebenen Stimmen bei Wahlen in den Jahren

1976 und 1978 berechnet. Diese Daten wurden dann auch für die Berechnung der „precinct weights“ verwendet.

Das dritte Gewicht diente dazu, den Stichprobenfehler zu verringern. Dabei handelt es sich um „a ratio estimate in which precinct-level returns aggregated across the entire sample are manipulated in accordance with expected vote outcome“ (Levy 1983: 62).

Die Formel für die Quotientenschätzung lautet:

$$E = \left( \frac{\sum x_i}{\sum y_i} \right) Y$$

- E = estimated final vote for a candidate „tonight“ (e.g., Reagan vote, 1980)
- $x_i$  = weighted, precinct-level vote for a candidate „today“ (e.g., Reagan vote, 1980, Baltimore Ward 3, Precinct 11)
- $y_i$  = weighted, known precinct-level vote for „comparable“ candidate in previous election (e.g., Ford vote, 1976, Baltimore Ward 3, Precinct 11)
- Y = actual statewide vote for „comparable“ candidate in previous election (e.g., Ford, 1976 general election, Maryland) (Levy 1983: 63)

Die Gewichtung über den Quotientenschätzer liefert nur das erwartete Wahlergebnis für einen Kandidaten. Daher wird die Berechnung für alle zur Wahl stehenden Kandidaten einzeln durchgeführt. Wenn der Gewichtungsprozess abgeschlossen ist, beginnt die Analyse der Umfrage mit den gewichteten Daten (Levy 1983).

### Gewichtung bei Non-Response

Ein großes Problem in Amerika ist die hohe Anzahl der Antwortverweigerer. So lag die Non-Response-Rate bei der Präsidentschaftswahl 2004 im Durchschnitt bei 47%. Dabei reichte die Spanne von 61,4% in Indiana bis 33,3% in Tennessee (National Election Pool 2004). In Deutschland liegt die Partizipationsrate mit ungefähr 70% wesentlich höher. Auch der Anteil der Personen, die die Rückseite des in der Regel beidseitig bedruckten Fragebogens nicht ausfüllen, ist mit rund 10% in den USA deutlich größer als die ungefähr 2% in Deutschland (Hofrichter 1999). Dies könnte unter anderem eine Folge der geringeren Anzahl an Fragen in Deutschland sein und der Tatsache, dass in Amerika Umfragen noch wesentlich exzessiver genutzt werden als in Deutschland. Warren Mitofsky hält es auch für möglich, dass die Logos der Fernsehsender auf den Fragebögen die Antwortbereitschaft reduzieren. „The conservatives see CNN, the liberals see FOX, and the next thing you know, we have got a problem“ (Mitofsky in Scheuren und Alvey 2008: 65). Hinzu kommt, dass die Wahllokale in Deutschland mit mehreren Interviewern besetzt werden, während in Amerika jedem Stichprobenbezirk meist nur ein Interviewer zugeteilt wird. Lediglich in Bezirken, in denen man eine hohe Wahlbeteiligung erwartet oder

die keinen guten Ruf genießen, werden zum Teil auch zwei Interviewer eingesetzt. Dies hat zur Folge, dass man neben den Antwortverweigerern auch noch den Fall hat, dass der Interviewer Personen verpasst, die er eigentlich befragen müsste. Dies kann beispielsweise passieren, wenn gleichzeitig viele Leute das Wahllokal verlassen. So war ein Viertel des Non-Response in den USA durch Interviewer verschuldet, die Wähler verpasst haben. Außerdem hatten die Interviewer zum Beispiel bei den VRS Exit Polls 1992 jede Stunde 10 Minuten Pause. Während ihrer Pause „they did a hand tally of the number of respondents who said they voted for each candidate. Interviewers telephoned results of the hand tally and the individual responses to each question to a central site three times during the day“ (Mitofsky und Edelman 1995: 83). Auch in dieser Zeit können keine Interviews durchgeführt werden.

Eine hohe Zahl Antwortverweigerer muss allerdings nicht unbedingt zu schlechten Ergebnissen führen. So stellten Mitofsky und Edelman (1995) fest, dass der Anteil der Antwortverweigerer nicht mit den Abweichung der Exit Polls von den tatsächlichen Ergebnissen korreliert ist. Außerdem konnte in einer Studie nachgewiesen werden, dass die Resultate bei geringerem Non-Response teilweise sogar eine stärkere Verzerrung aufwiesen (Merkle und Edelman 2002).

Es ist jedoch wichtig zu beachten, dass sich die Antwortverweigerer nicht gleichmäßig auf alle Bevölkerungsgruppen verteilen. Man weiß zum Beispiel, dass es ältere Personen eher ablehnen, an einem Exit Poll teilzunehmen. Noch weitaus gravierender (für die Prognose am Wahlabend) ist das Problem, dass konservative Wähler seltener an Wahltagsbefragungen teilnehmen als Wähler anderer Parteien. So hat man in Großbritannien meist einen höheren Anteil an Labour-Wählern in den Exit-Poll-Daten als in den ausgezählten Stimmen und in Amerika trifft dies auf die Wähler der demokratischen Partei zu.

In Deutschland und den USA versucht man, das Non-Response-Problem durch Gewichtung zu beseitigen. Hierzu werden Korrekturen anhand der erfassten Merkmale Alter, Geschlecht und Rasse vorgenommen. Allerdings müssen diese Anpassungen nicht unbedingt zu großen Veränderungen führen. So hatte bei der amerikanischen Präsidentschaftswahl 1992 nur die Altersadjustierung einen wirklichen Effekt. Außerdem unterschied sich zwar auch die Antwortbereitschaft für Weiße und Schwarze, dies führte bei den Ergebnissen allerdings zu keiner nennenswerten Veränderung (Mitofsky und Edelman 1995). Solche soziodemographischen Gewichtungen sind jedoch nicht ganz unproblematisch, da man davon ausgeht, dass Personen mit denselben Merkmalen in Bezug auf Alter, Rasse und Geschlecht auch dieselbe Partei wählen. Ob man allerdings wirklich davon ausgehen kann, dass eine schwarze junge Frau, die an der Wahltagsbefragung teilnimmt, dasselbe Wahlverhalten zeigt wie eine schwarze junge Frau, die nicht teilnimmt, ist fraglich.

### **Weitere Gewichtungsmöglichkeiten**

Für die Gewichte können weitere Faktoren berücksichtigt werden: das Stichprobendesign, Informationen aus offiziellen Quellen (z.B. Wahlstatistiken) und Informationen, die über die Wahltagsbefragung gesammelt wurden (z.B. Recall-Fragen). Bei der Gewichtung aufgrund des Stich-

probendesigns wird ungleichen Auswahlwahrscheinlichkeiten Rechnung getragen. Obwohl man versucht, einigermaßen gleiche Auswahlwahrscheinlichkeiten für alle Urnenwähler zu bekommen, gibt es einige Fälle, wo dies nicht der Fall ist. So werden in den USA teilweise Bezirke mit einer großen Minderheitenpopulation bei der Stichprobenziehung mit einer größeren Auswahlwahrscheinlichkeit versehen, als Bezirke dieser Größe normalerweise erhalten würden, und tauchen dadurch verstärkt in der Stichprobe auf. Um dieses Oversampling auszugleichen, werden die Ergebnisse dieser Bezirke bei der Analyse wieder heruntergewichtet. Eine ähnliche Situation hatte man in New Hampshire, wo bei der Präsidentschaftswahl 2004 einige sehr große Wahlbezirke mit der Auswahlwahrscheinlichkeit Eins versehen wurden und somit mit Sicherheit in der Stichprobe auftauchten (National Election Pool 2004).

Eine weitere Möglichkeit zur Gewichtung stellt die sogenannte Recall-Frage dar. In vielen Exit-Poll-Fragebögen wird abgefragt, welche Partei der Befragte bei der letzten Wahl gewählt hat. Diese Frage ermöglicht es anhand des Unterschiedes zwischen den Stimmen, die eine Partei bei der letzten Wahl tatsächlich erhalten hat und dem Anteil der Personen, die bei den Exit Polls angaben, diese Partei gewählt zu haben, zu gewichten. Allerdings ist diese Form der Gewichtung ziemlich umstritten. Das Problem ist, dass man davon ausgeht, dass sich die Befragten korrekt an ihr Verhalten bei der letzten Wahl erinnern, was nicht unbedingt gegeben sein muss. Die Gewichtung der Daten auf der Grundlage von fehlerhaften Erinnerungen kann dazu führen, dass die Stichprobe durch die Gewichtung noch stärker verzerrt wird. Allerdings gibt es auch Befürworter der Recall-Gewichtung, die argumentieren, dass sich die fehlerhaften Angaben wieder ausgleichen (Hofrichter 1999).

#### 4.3.2 Quotientenschätzung in den USA

Prinzipiell reicht es, die Daten mit der inversen Auswahlwahrscheinlichkeit zu gewichten, um einen unverzerrten Schätzer zu erhalten. Allerdings ist bereits am Beispiel der Gewichtung bei CBS angeklungen, dass durch den geschickten Einsatz zusätzlicher Informationen Schätzungen deutlich verbessert werden können. Bei der „zusätzlichen Information“ sollte es sich um ein Hilfsmerkmal handeln, das mit der untersuchten Variable hoch korreliert ist. Bei Wahlprognosen bieten sich dafür meist Wahlergebnisse aus der Vergangenheit an.

Es gibt verschiedene Möglichkeiten, wie die Vergleichswahl zur Schätzung der aktuellen Wahlergebnisse verwendet werden kann. Es wären sowohl Quotienten- als auch Differenzen- und Regressionschätzer möglich. Für die USA hat sich gezeigt, dass Quotientenschätzer die stabilsten Ergebnisse liefern. Der Regressionsschätzer würde wahrscheinlich zu besseren Ergebnissen führen, allerdings nur wenn man die Koeffizienten einigermaßen korrekt schätzen kann. Dazu benötigt man jedoch eine gewisse Menge an Daten, die zu Beginn des Wahlabends oftmals noch nicht vorliegen. Das Problem beim Quotientenschätzer dagegen ist, dass er nicht linear ist. Dies kann zu einer Verzerrung führen, wenn nur wenige Daten vorliegen.

In Amerika wurden bei der Präsidentschaftswahl 2000 laut Warren Mitofsky und Murray Edel-

man (2002) am Wahlabend vier verschiedene Schätzer berechnet, die die Daten der Wahlbezirke verwendeten. Dabei handelt es sich um zwei Quotientenschätzer und zwei einfache unverzerrte Schätzer, für die keine Ergebnisse von vorausgegangenen Wahlen verwendet wurden.

Im Nenner des Verhältnisschätzers stehen die Stimmen für eine Partei, die sie bei einer geeigneten Vergleichswahl in denselben Bezirken erhalten hat, in welchen für die aktuelle Wahl bereits Ergebnisse in Form von Exit-Poll-Daten oder ausgezählten Stimmen vorliegen. Die am besten geeignete Wahl wird über die Höhe der Korrelation mit der aktuellen Wahl bestimmt. Bis einschließlich 2000 wurden alle Berechnungen nur auf Grundlage einer Vergleichswahl durchgeführt. Da man in diesem Jahr jedoch feststellen musste, dass die ausgewählte Wahl sehr schlechte Resultate lieferte, hat man beschlossen, dass die Quotientenschätzer zukünftig für mehrere Wahlen berechnet werden. Außerdem wurden Überlegungen angestellt, multivariate Schätzungen mit drei Wahlen als Zielgröße zu verwenden.

Der Verhältnisschätzer wird nicht für die einzelnen Wahlbezirke berechnet, sondern entweder für einen ganzen Staat oder einen größeren Teil des Staates. Der Quotient aus den Stimmen bei der aktuellen Wahl und den Stimmen bei einer vorausgegangenen Wahl wird dann mit dem offiziellen Endergebnis der Vergleichswahl multipliziert, wobei es sich dabei um das Ergebnis für das jeweilige Gebiet handelt, für das der Quotientenschätzer berechnet wurde.

Bevor der Quotientenschätzer berechnet wird, findet eine nachträgliche Schichtung statt. Dabei gibt es zwei verschiedene Varianten: geographische Schichtung und Schichtung nach Stimmenteilen bei einer vorausgegangenen Wahl. Bei der geographischen Schichtung wird jeder Bundesstaat meist entlang von County-Grenzen in ungefähr fünf Schichten unterteilt. Bei der zweiten Art der Schichtung versucht man, die Wahlbezirke in jedem Staat in fünf möglichst homogene Gruppen - in Bezug auf die Parteienteile bei einer älteren Wahl - aufzuteilen. Falls in einem Staat ein größerer Anteil schwarzer Wähler lebt, gibt es eine zusätzliche sechste Schicht, die aus Wahlbezirken mit hauptsächlich schwarzen Wählern besteht. Der Grund dafür liegt unter anderem in den Erfahrungen aus dem Jahr 1966. Die schwarze Bevölkerung Amerikas wählt in der Regel demokratisch. Doch in diesem Jahr gab es einen fast vollständigen Umschwung zum republikanischen Kandidaten. Wenn eine solche Veränderung nicht beachtet werden würde, hätte das schwerwiegende Konsequenzen für die Qualität der Wahlprognose. Bei den beiden Schätzern, die keine vorausgegangene Wahl als Grundlage verwenden, gibt es ebenfalls diese beiden Arten der nachträglichen Schichtung.

Außerdem wird zu jeder Schätzung immer noch der Stichprobenfehler berechnet. Dieser wird benötigt, um eine Entscheidung über den besten Schätzer zu treffen. Bis 2000 wurde der Stichprobenfehler über eine Taylorentwicklung berechnet. Da man sich im Zuge der Diskussionen über das schlechte Abschneiden der Prognosemodelle bei der Wahl 2000 dafür entschieden hat, in Zukunft komplexere Schätzungen durchzuführen, erschien es sinnvoll, auch über eine neue Art der Varianzberechnung nachzudenken, da für das neue Vorgehen eine Berechnung über die Taylorentwicklung als zu kompliziert erschien. Daher wollte man die Taylorentwicklung durch eine „replication variance calculation“ (Mitofsky 2003: 56) ersetzen.



### 4.3.3 Prognose am Wahlabend

Es gibt zwei verschiedene Ausgangslagen bei der Prognose am Wahlabend: Entweder sollen Stimmanteile für eine Partei oder einen Kandidaten prognostiziert werden. Dies ist zum Beispiel in Deutschland der Fall. In Großbritannien dagegen müssen die Stimmenanteile auf Sitze umgerechnet werden. Dabei handelt es sich um eine zusätzliche Hürde, da die Stimmanteile nicht Eins zu Eins in Sitze umgewandelt werden können, sondern die Anzahl der gewonnenen Sitze stark davon abhängt, wie die Stimmen verteilt sind. Während es für kleine Parteien günstiger ist, wenn sich die Stimmen in wenigen Wahlkreisen konzentrieren und sie in diesen dann die Mehrheit der Stimmen auf sich vereinen können, ist es für große Parteien besser, wenn sich die Stimmen stärker verteilen und sie immer nur mit knappem Vorsprung einen Sitz gewinnen. Am Wahlabend gibt es unterschiedliche Prognosemodelle, die auf verschiedenen Daten basieren. So gibt es Modelle, die mit Daten auf Wahlbezirksebene arbeiten. Dafür wird eine Stichprobe von Wahlbezirken gezogen, von der die Exit-Poll-Stichprobe normalerweise eine Teilmenge ist. Da die Stimmen in den einzelnen Wahlbezirken meist recht schnell ausgezählt sind, erhält man auch früh ein Ergebnis. Solche Modelle sind jedoch fehleranfälliger als Modelle, die die ausgezählten Stimmen der Wahlkreise verwenden. Hinzu kommt, dass Ergebnisse oftmals nur auf Wahlkreis-Ebene zu haben sind. Wahlkreis-Ebenen-Modelle wurden in erster Linie für knappe Wahlausgänge entwickelt, da es bis zum nächsten Morgen dauern kann, bis die Auszählungen beendet sind. Meist zählen die Wahlbezirke ihre Stimmen selbst aus und melden das Ergebnis an den entsprechenden Wahlkreis, wo die Ergebnisse des ganzen Wahlkreises gesammelt werden. Zusätzlich gibt es noch Modelle, die sowohl Wahlbezirks- als auch Wahlkreisergebnisse verwenden (Gawiser 1994).

In den USA werden die Analysen zwar in erster Linie vom NEP gemacht, allerdings ist jeder Fernsehsender für die Auswertung seiner eigenen Fragen zuständig. Außerdem können die Fernsehsender auf Grundlage ihrer eigenen Analysen selbstständig jemanden zum Sieger erklären, unabhängig davon, ob dieser Kandidat auch beim NEP einen „Call-Status“ hat, also alle Bedingungen erfüllt, dass man ihn zum Sieger erklären kann. Für diese Auswertungen bekommen die Medien die aufbereiteten Daten in Kreuztabellen zur Verfügung gestellt.

In Amerika ist die Minimal-Anforderung, um eine Entscheidung am Wahlabend zu verkünden, dass der kritische Wert das 99,5%-Quantil der t-Verteilung übersteigt. Der kritische Wert wird berechnet, indem von der geschätzten Differenz zwischen den Stimmanteilen der beiden führenden Kandidaten der geschätzte Effekt der Briefwähler abgezogen wird. Anschließend wird noch durch den Standardfehler geteilt (Mitofsky und Edelman 2002).

## 4.4 Evaluation

Exit Polls sind eine der wenigen Umfragearten, bei denen es möglich ist, die Ergebnisse der Befragung mit den tatsächlichen Werten in der Grundgesamtheit zu vergleichen. Sobald alle Stimmen ausgezählt sind, kann man theoretisch eindeutig den Erfolg oder Misserfolg einer

Wahltagsbefragung beurteilen. Praktisch gestaltet sich der Versuch, die Qualität der Exit-Poll-Daten zu beurteilen, jedoch etwas schwieriger. So werden in Großbritannien die Wahlergebnisse nur auf Wahlkreisebene ausgezählt und können somit nicht mit den Daten, die für einzelne Wahlbezirke erhoben wurden, verglichen werden. Dasselbe Problem hat man teilweise auch in den USA, wenn in Countys beispielsweise Wahlmaschinen verwendet werden, bei denen das Ergebnis nicht nach Wahlbezirken aufgeschlüsselt werden kann.

Für eine Beurteilung von außen ergibt sich oftmals noch eine zusätzliche Schwierigkeit. Levy (1983) unternahm für die amerikanische Präsidentschaftswahl von 1980 den Versuch, die Exit-Poll-Daten zu evaluieren. Dafür konnte er jedoch nur die Ergebnisse von ABC benutzen, da deren Studiendesign so konzipiert war, dass auch ohne Gewichtung sinnvolle Ergebnisse erzielt wurden. Bei den anderen Fernsehsendern waren dagegen keine Daten zugänglich, die nur mit den Auswahlwahrscheinlichkeiten gewichtet wurden. Verfügbar waren nur diejenigen Exit-Poll-Daten, die bereits an die tatsächlichen Wahlergebnisse angepasst worden sind. Mit solchen Werten ist ein Vergleich jedoch nicht mehr sinnvoll.

Die Stichprobenabweichung definiert Levy als Differenz zwischen der aus der Stichprobe gewonnenen Prozentzahl der siegreichen Kandidaten und den offiziellen Ergebnissen. Da es sich bei einem Exit-Poll-Sample um eine stark geclusterte Stichprobe handelt, berechnet Levy zusätzlich den Effekt der Clustergröße auf den Stichprobenfehler durch  $1 + \rho(\bar{n} - 1)$ , wobei  $\bar{n}$  die durchschnittliche Stichprobengröße jedes Clusters und  $\rho$  ein Maß für die Clusterhomogenität ist und empirisch bestimmt wurde. Die Wahlergebnisse wurden als binomiale Variable behandelt, wobei der Stimmanteil des siegreichen Kandidaten ein Wert war und die Summe aller anderen Stimmanteile als zweiter Wert angesehen wurde. „By taking into account the size of the election day samples, the frequency distribution of the binomial voting-percentage variable, and an adjustment for the cluster sampling design, it was possible to determine the degree to which sample deviation varied from theoretically expected sampling error“ (Levy 1983: 65). Bei der Präsidentschaftswahl 1980 lag der theoretische Fehler von ABC bei  $\pm 2\%$ , wohingegen der tatsächliche Fehler lediglich  $-0,2\%$  betrug.

## 4.5 Wahltagsbefragung in Deutschland

Die Wahltagsbefragung in Deutschland hat einige Gemeinsamkeiten mit den Exit Polls in den USA. Dies ist nicht weiter verwunderlich, da die Methoden ursprünglich aus den Vereinigten Staaten übernommen worden sind. Allerdings unterscheiden sich die beiden Länder sowohl geographisch als auch wahlrechtlich, weshalb es auch einige Unterschiede gibt. Das betrifft zum Beispiel die Anzahl der Interviewer (in Deutschland werden jedem Wahlbezirk in der Stichprobe zwei Interviewer zugewiesen) und den Abstand vom Wahllokal (die Interviewer in Deutschland stehen direkt am Ausgang). Außerdem werden den ganzen Tag über Befragungen durchgeführt (Blumenthal 2004). Auf viele Unterschiede kann jedoch nicht genauer eingegangen werden, da wenig konkrete Details zu den Wahltagsbefragungen in Deutschland publiziert werden. Dies

betrifft vor allem die Prognose auf Grundlage der Exit Polls. Der Grund für die fehlenden Informationen ist die Konkurrenz von Infratest dimap und der Forschungsgruppe Wahlen.

#### 4.5.1 Stichprobenziehung

Zur Stichprobenziehung greift man auch in Deutschland auf ein zweistufiges Verfahren zurück, wobei als erstes eine Schichtung nach Regionen und dem Wahlverhalten bei der letzten vergleichbaren Wahl vorgenommen wird. Anhand der Schichtungsmerkmale erfolgt die Erstellung einer Liste der Wahlbezirke. Aus dieser Liste werden mehrere systematische Stichproben gezogen und die beste ausgewählt. Das wichtigste Kriterium für die Qualität der Stichprobe ist, dass sie die Ergebnisse der vergangenen Wahl möglichst gut widerspiegelt, da man zur Wahlprognose auf gebundene Hochrechnungen zurückgreifen möchte. Dabei bildet die sogenannte „verbundene Stichprobe“ die Basis für die Vorhersage der Ergebnisse der aktuellen Wahl. Probleme können dabei durch Veränderungen in den Wahlbezirken auftreten. Dies kann einerseits die Wählerschaft und andererseits die Grenzen des Bezirks betreffen. Wenn sich die Grenzen von ausgewählten Stimmbezirken verändert haben, können diese entweder ausgeschlossen werden oder man versucht, die Ergebnisse der letzten Wahl auf die neuen Grenzen umzurechnen. In Deutschland ist die Wahrscheinlichkeit, mit der die Wahlbezirke gezogen werden, unabhängig von der Größe der Wahlbezirke. Ein solches Vorgehen ist gerechtfertigt, da die Variabilität der Stimmbezirksgrößen in Deutschland nicht sehr groß ist, weil es eine Obergrenze für die Größe von Wahlbezirken gibt. Diese liegt bei ungefähr 2500 wahlberechtigten Personen, während die durchschnittliche Stimmbezirksgröße in Deutschland etwas unter 1000 Wahlberechtigte beträgt. In Deutschland gibt es auch Anstaltsbezirke, zum Beispiel Krankenhäuser oder Gefängnisse. Diese eigenständigen Wahlbezirke haben jedoch spezielle Eigenschaften, wie eine starke Fluktuation bei den Bewohnern, und werden daher normalerweise ausgeschlossen. Auch bei sehr kleinen Bezirken kann man sich überlegen, diese aus Kostengründen bei der Stichprobenziehung nicht zu berücksichtigen (Hofrichter 1999).

#### 4.5.2 Gewichtung

Neben den Gewichtungsmöglichkeiten, die auch in den USA bestehen, gibt es in Deutschland noch eine zusätzliche Informationsquelle. Seit 1957 wird die repräsentative Wahlstatistik erstellt. Dabei werden in einigen zufällig ausgewählten Wahlbezirken die Stimmzettel so gekennzeichnet, dass getrennt nach Geschlecht und nach fünf Altersgruppen ausgezählt werden kann. Der repräsentativen Wahlstatistik liegt eine sehr große Datenmenge zu Grunde. So umfasste die Stichprobe für die Bundestagswahl 2005 1,7 Millionen der 48 Millionen Wähler in Deutschland (Roth 2008). Die Daten über das Wahlverhalten und die Wahlbeteiligung der zehn Gruppen bei zurückliegenden Wahlen können für eine soziodemographische Gewichtung der Wahltagsbefragungsdaten genutzt werden. Gegebenenfalls kann zusätzlich noch nach verschiedenen Regionen unterschieden werden. Lediglich für die Bundestagswahlen 1994 und 1998 gibt es keine solche

Statistik. In diesen Jahren versuchte man, über die Wahltagsbefragung das Fehlen der Wahlstatistik auszugleichen. Allerdings haben Wahltagsbefragungen den entscheidenden Nachteil, dass keine genaue Schätzung der Wahlbeteiligung in den einzelnen Gruppen möglich ist. Seit der Bundestagswahl 2002 sind auch Briefwähler in der repräsentativen Wahlstatistik enthalten. In Deutschland werden Recall-Fragen laut Hilmer (2008) eher selten zur Gewichtung verwendet, da sie mit zahlreichen Problemen behaftet sind. Sie werden vorrangig als Grundlage für Wählerwanderungsanalysen verwendet.

#### 4.5.3 Analyse

In Deutschland sind die Meinungsforschungsinstitute für die Analyse der Daten und ihre grafische Aufbereitung zuständig. Sie erstellen die 18-Uhr-Prognose und die Hochrechnungen am Wahlabend, die auf ausgezählten Stimmen basieren. Die Daten werden den ganzen Wahltag über analysiert, um Muster zu erkennen und einen ersten Eindruck von der Wahl zu bekommen. Die Wahlprognose, die um 18 Uhr veröffentlicht wird, beruht nur auf den Ergebnissen der Wahltagsbefragungen. Diese Prognosen sind meistens recht genau, lediglich für kleine Parteien ist es oft schwierig vorherzusagen, ob sie die 5%-Hürde erreichen werden.

Nach 18 Uhr stehen zusätzlich die ersten ausgezählten Stimmen zur Verfügung und werden in ein Computermodell eingespeist, um die Sitzverteilung zu schätzen. Später am Abend sind dann auch Ergebnisse für die Wahlkreise verfügbar und werden ebenfalls in den Modellen verwendet. Es muss beachtet werden, dass die ausgezählten Stimmen, die als erstes eintreffen, meist nicht repräsentativ für die ganze Wahl sind, da Bezirke mit weniger Wählern normalerweise schneller ausgezählt sind als solche mit vielen Wählern. Da sich diese Bezirke aber auch in ihrem Charakter unterscheiden können (beispielsweise ländliche im Gegensatz zu städtischen Bezirken), kann man anhand der ersten Ergebnisse einen falschen Eindruck vom Ausgang der Wahl erhalten. „Die H. [Hochrechnung, Anm. des Autors] von Parteiateilen erfolgt in der Bundesrepublik Deutschland stets auf der Basis von Stichproben aus Wahlbezirken mit Hilfe gebundener Schätzverfahren, wobei als zusätzliches Merkmal die Wahlergebnisse einer vorhergehenden Wahl dienen. H. am Wahlabend erfolgen aber nicht erst dann, wenn alle Wahlbezirke, die zur Stichprobe gehören, ihr Wahlergebnis gemeldet haben, sondern schon zu einer Zeit, wo nur ein Teil der Stichprobenergebnisse vorliegt, wodurch das Fehlerintervall der H. entsprechend größer wird. Bei noch unvollständigen Stichproben liefert die Differenzschätzung für große und die Verhältnisschätzung für kleine Parteien erfahrungsgemäß die besten Werte“ (Gibowski 1985: 343). Regressionsschätzer sind erst dann sinnvoll, wenn die Resultate von mehr als 30 Wahlbezirken vorliegen. Außerdem können die verschiedenen Prognoseverfahren auch mit Schichtung nach bestimmten Merkmalen kombiniert werden (Hofrichter 1999).

Auch in Deutschland hat man das Problem der Briefwähler. Diese werden in sogenannten Briefwahlbezirken organisiert und extra ausgezählt. Besonders in Stadtstaaten können Briefwähler zu einem Problem werden, da man teilweise einen Briefwähleranteil von bis zu 20% hat. Die Schätzungen für Briefwähler können über Vorwählerhebungen erfolgen und zusätzlich können

die Modelle im Nachhinein an die Ergebnisse angepasst werden. Besonders in Deutschland, wo man bei nahezu allen Wahlen per Brief wählen kann, ist zweiteres eine gute Möglichkeit um Hochrechnungsmodelle zu optimieren (Hilmer 2008).

## 4.6 Exit Polls in Großbritannien

Während die Exit Polls in Deutschland und den USA eine Menge Gemeinsamkeiten haben, wird in Großbritannien ein anderer Ansatz verfolgt. Zwei entscheidende Unterschiede sind, dass in Großbritannien die Befragungen jedes Mal in denselben Wahllokalen stattfinden und die Prognose mit Hilfe von Regressionsmodellen erfolgt.

Ein Grund für das abweichende Vorgehen in Großbritannien sind die besonderen Schwierigkeiten bei der Wahlprognose, mit denen die britischen Wahlforscher konfrontiert werden. Eine Herausforderung ist, dass in Großbritannien die Wahlergebnisse nicht in den einzelnen Wahlbezirken ausgezählt und verkündet werden, sondern in den Wahlkreisen. Daher können die Ergebnisse vergangener Wahlen nicht zu Schichtungszwecken verwendet werden, um so eine repräsentative Auswahl der Wahlbezirke zu gewährleisten. Außerdem hat Großbritannien das Mehrheitswahlrecht. Dabei wird in jedem der Wahlkreise (2005 gab es 646) über einen Sitz im Unterhaus abgestimmt. Diesen Sitz erhält diejenige Partei, die die meisten Stimmen in einem Wahlkreis auf sich vereinen kann. Daher gibt es nicht unbedingt einen klaren Zusammenhang zwischen dem landesweiten Stimmenanteil einer Partei und der Anzahl an gewonnenen Sitzen. Schließlich finden die Wahlen nicht an einem arbeitsfreien Tag, sondern am Donnerstag statt, was dazu führt, dass viele Wähler erst abends ihre Stimme abgeben. Deshalb fällt der Großteil der Daten relativ spät am Wahltag an und muss dann in einem äußerst kurzen Zeitraum analysiert werden (Curtice und Firth 2008).

### 4.6.1 Auswahl der Stichprobenbezirke

Während es bei ITV schon länger üblich war, bei jeder Wahl diesselben Wahllokale mit Interviewern zu besetzen, führte man ein solches Design bei BBC erst bei der Unterhauswahl 2001 ein. 1999 hatte man noch eine Stichprobe von 140 Wahlkreisen mit einer Auswahlwahrscheinlichkeit, die proportional zur Größe der Wahlkreise war, zufällig gezogen. Außerdem kamen noch 80 weitere Wahlkreise in die Stichprobe, in denen es 1992 eine knappe Entscheidung zwischen Labour und den Konservativen gegeben hatte. Schließlich wurden in jedem Wahlkreis zwei Wahllokale ausgewählt. Dazu wurden die Wahlbezirke, aus denen jeder Wahlkreis besteht, nach dem Stimmanteil der Konservativen bei der letzten lokalen Wahl sortiert. Anschließend wurde ein Stichprobenelement aus der oberen Hälfte und eines aus der unteren Hälfte gezogen (Brown, Firth und Payne 1999).

Da zu Beginn ihrer Zusammenarbeit 2004 sowohl ITV als auch BBC ein Paneldesign verwen-

deten, wurde dieses Vorgehen auch weiterhin beibehalten. Ein solches Design war die Lösung für zwei elementare Probleme: der höhere Anteil an Antwortverweigerern unter den Wählern der Konservativen und das Fehlen der Wahlergebnisse auf Stimmbezirksebene. Bei der Ziehung der Stichprobenbezirke wird auf der ersten Stufe in den meisten Ländern nach Merkmalen geschichtet, die als relevant für die Wahlentscheidung angesehen werden. In Großbritannien liegen aber für die Wahlbezirke, die als Untersuchungseinheiten verwendet werden sollen, weder die Ergebnisse der letzten Wahl vor, noch decken sich diese Gebiete mit den Bezirken, für die Zensusdaten vorliegen würden. Daher kann auch nicht nach anderen Merkmalen geschichtet werden, die beim Zensus erhoben werden und für die Wahlentscheidung von Bedeutung sind, weshalb die Wahrscheinlichkeit sehr hoch ist, dass man keine repräsentative Auswahl an Wahlbezirken erhält. Dies könnte bei einer Betrachtung der absoluten Werte zu einer starken Verzerrung führen, weil es dabei recht große Unterschiede zwischen den Wahlkreisen gibt. „For example, the constituency level standard deviation of the percentage of the vote that was won by the three main parties in Britain in 2005 was Labour, 15.1, Conservative, 14.0, and Liberal Democrat, 10.4.“ Allerdings hat man festgestellt, dass die Veränderungen der Stimmanteile wesentlich weniger variieren. „In contrast, the equivalent statistic for the change in their share of the vote since 2001 was only Labour, 4.0, Conservative, 3.0, and Liberal Democrat, 4.4“ (Curtice und Firth 2008: 511). Man hat nun versucht, die Schwierigkeiten, die eine nicht repräsentative Stichprobe bereitet, dadurch zu verringern, dass man nicht mehr die Stimmanteile selbst schätzt, sondern nur noch die Differenzen zwischen den Stimmanteilen bei verschiedenen Wahlen. Jedoch werden zur Berechnung der Veränderungen die Ergebnisse der letzten Wahl benötigt, die auf Wahlbezirksebene nicht existieren. Stattdessen verwendete man 2005 die Exit-Poll-Daten, die bei der Wahl 2001 erhoben wurden, als Ausgangswert und versuchte 2005 wieder dieselben Wahllokale mit Interviewern zu besetzen. Ziel war es, die lokalen Veränderungen der Stimmanteile der einzelnen Parteien zu schätzen. Mit diesen geschätzten Veränderungen wurden ausgehend von den Wahlkreisergebnissen von 2001 die Ergebnisse für 2005 berechnet, indem man die Veränderungen zu den Ausgangswerten addierte. Diese Vorgehensweise hilft auch beim zweiten Problem, nämlich der höheren Antwortbereitschaft von Labourwählern. Wenn nur die Differenzen zwischen den Ergebnissen von zwei Wahltagsbefragungen berechnet werden, reicht es, wenn sich der Anteil der Antwortverweigerer bei den Wählern der jeweiligen Parteien nicht verändert hat, um den korrekten Wert zu erhalten.

Für die Wahl 2005 hatten sich BBC und ITV entschieden, in 120 Wahllokalen Wahltagsbefragungen durchzuführen. 2001 führte BBC Befragungen in 90 Wahllokalen durch, wobei Wahlkreise mit knappen Wahlausgängen, vor allem zwischen Labour und den Konservativen, besonders stark vertreten waren. ITV dagegen hatte seine 100 Wahllokale nur in Wahlkreisen, bei denen es in früheren Wahlen knappe Entscheidungen zwischen Labour und den Konservativen oder den Liberalen und den Konservativen gegeben hatte. Nun musste man sich auf 120 Wahlbezirke einigen. Allerdings gestaltete sich dieses Unterfangen als relativ einfach, da sich viele Wahlbezirke seit der letzten Wahl so stark verändert hatten, dass kein sinnvoller Vergleich der Daten

von 2001 und 2005 möglich gewesen wäre. So blieben schlussendlich nur 116 Orte übrig und es wurden noch vier weitere Wahlbezirke in London hinzugenommen, die 2001 noch nicht besucht wurden. Diese wurden nach ihrer Repräsentativität für den jeweiligen Wahlkreis ausgewählt, da man hoffte, die Veränderungen in diesen Bezirken durch den Vergleich der Exit-Poll-Daten von 2005 und den Wahlkreisergebnissen von 2001 berechnen zu können (Curtice und Firth 2008).

#### 4.6.2 Durchführung der Wahltagsbefragung

Moon und McGregor, die die BBC Exit Polls maßgeblich mitentwickelt hatten, schlugen vor, die Wahllokale nicht den ganzen Tag über mit Interviewern zu besetzen, sondern immer nur für drei Stunden. Dabei sollten aber alle Tageszeiten abgedeckt werden. Moon und McGregor wollten auf diese Weise die zur Verfügung stehenden Interviewerstunden möglichst effizient nutzen. Die Überlegung dahinter war, dass das Wahlverhalten während eines Tages in einem Wahlbezirk nicht so sehr variiert, wie das Wahlverhalten zwischen den Bezirken. 1997 wurden zum Beispiel in jedem der ausgewählten Wahllokale innerhalb von drei Stunden 40 Interviews durchgeführt (Moon und McGregor 1992).

2005 dagegen wurden die Wahllokale den ganzen Tag über besetzt. Es sollten in jedem der ausgewählten Wahllokale in etwa 225 Interviews durchgeführt werden, wobei der Anteil der befragten Wähler umgekehrt proportional zur Anzahl der registrierten Urnenwähler ist. Auch in Großbritannien sind die Antwortverweigerer ein Thema, obwohl nur einer von sechs ausgewählten Briten nicht an der Wahltagsbefragung teilnimmt. Damit liegt die Partizipationsrate deutlich über der in den USA, aber sobald die Antwortverweigerer sich sehr ungleichmäßig auf die verschiedenen Parteien verteilen, kann es trotzdem zu einem verzerrten Ergebnis kommen. Der Umgang mit Non-Response unterscheidet sich vom Vorgehen in Amerika und Deutschland. So sollen in Großbritannien die Interviewer angeben, wie sie vermuten, dass die Leute, die die Teilnahme verweigert haben, gewählt haben. Diese Vermutungen werden „adjusted by a matrix of interviewer guess by reported vote among respondents“ (Curtice und Firth 2008: 516).

Durch den Übergang zur Schätzung der Differenzen zwischen den Ergebnissen von zwei Wahltagsbefragungen ist es allerdings nicht mehr unbedingt nötig, eine Methode zu finden, mit der die Antwortverweigerer ausgeglichen werden können. Allerdings handelt es sich dabei um eine Möglichkeit zur Kontrolle, falls sich wider Erwarten eine höhere Verweigerungsrate bei den Labour-Wählern einstellen sollte oder ein noch größerer Anteil der konservativen Wähler eine Teilnahme an der Wahltagsbefragung ablehnt als bei der letzten Wahl (Curtice und Firth 2008).

#### 4.6.3 Prognose

Auch bei der Prognose verfolgt man in Großbritannien einen etwas anderen Ansatz. Die Veränderungen der Stimmanteile werden mit Hilfe von Regressionsmodellen geschätzt. Dazu werden die Ergebnisse aus der Wahltagsbefragung mit Informationen, die man über diesen Bezirk gesammelt hat, in Verbindung gebracht. Mögliche Prädiktoren sind zum Beispiel das Abschneiden

einzelner Parteien bei der letzten Wahl, aber auch die Unterscheidung, ob es sich um einen ländlichen oder städtischen Bezirk handelt. In Großbritannien wurden die Prognosen am Wahlabend (für BBC) schon seit Mitte der 1970er Jahre auf Grundlage von Regressionsmodellen erstellt, allerdings verwendete man zu dieser Zeit noch keine Daten aus Exit Polls. Stattdessen versuchte man, mit den Ergebnissen aus Wahlkreisen, in denen der Sieger bereits verkündet wurde und mit Daten aus Vorwahlbefragungen Prognosen für die Wahlkreise zu erstellen, in denen der Ausgang noch ungewiss war. Zur Schätzung der Regressionskoeffizienten wurde seit 1975 meist eine Ridge-Regression verwendet, da auf diese Weise schon bei sehr geringen Datenmengen die Schätzung von Koeffizienten möglich ist. Im Laufe der Jahre wurden zahlreiche Veränderungen, zum Beispiel an den Zielvariablen, vorgenommen und die Exit Polls wurden ein wichtiger Bestandteil der Wahlprognose.

Bei der Wahl 2005 hatte man als Zielvariablen die Veränderungen in den Stimmanteilen für:

- $100 \text{ LD}/(\text{LD}+\text{Lab}+\text{Con})$
- $100 \text{ Lab}/(\text{Lab}+\text{Con})$

Dabei steht „LD“ für Liberal Democrats, „Lab“ für Labour und „Con“ für Conservative Party. Um eine einfache Interpretation zu gewährleisten, wurden die Responsevariablen ohne Transformation verwendet. Die stetigen Prädiktoren dagegen wurden entweder dichotomisiert oder in drei geordnete Kategorien umgewandelt, um Stimmanteile über 100% oder unter 0% zu verhindern. Für jede Zielgröße wurden verschiedene multiple lineare Regressionsmodelle gefittet. Die Schätzung der Koeffizienten erfolgte über eine gewichtete KQ-Schätzung, wobei starke Ausreißer automatisch heruntergewichtet wurden. Allerdings zeigte sich, dass mit dem verwendeten Verfahren teilweise sehr extreme Parameter geschätzt werden. Daher schlagen Curtice und Firth weitere Untersuchungen in Richtung Shrinkageverfahren vor. Auch in anderen Aspekten unterschied sich das Vorgehen 2005 von dem in früheren Jahren. So wurden die Exit-Poll-Daten in diesem Jahr lange als einzige Grundlage für die Prognosen verwendet. Die tatsächlich ausgezählten Ergebnisse wurden dagegen erst spät am Abend eingesetzt. Damit hat man solange kein Problem, wie die Ergebnisse der Exit Polls frei von größeren Fehlern sind. Dies war 2005 der Fall, als man allein auf Grundlage der Wahltagsbefragungen die Verteilung der Sitze exakt prognostizieren konnte. Sollten sich allerdings Ungenauigkeiten bei den Exit Polls ergeben, dann würde man diese nur schwer entdecken.

Die geschätzten Veränderungen wurden dann zu den Wahlkreisergebnissen von 2001 addiert. Auf diese Weise erhielt man die geschätzten Stimmanteile für die einzelnen Parteien, welche als Grundlage für eine Wahrscheinlichkeitsvorhersage der Sitze dienten. Anstatt immer der Partei den Sitz zuzuschlagen, die den höchsten Stimmanteil in einem Wahlkreis hatte, berechnete man die Wahrscheinlichkeit, mit der sie den Sitz bekommen würde. So trug man der Tatsache Rechnung, dass es sich bei den Stimmanteilen nur um Schätzungen handelt, die mit einem gewissen Fehler behaftet sein können. Die Wahrscheinlichkeiten, die man auf diese Weise berechnet hatte, wurden aufsummiert und schließlich auf eine ganze Zahl gerundet. Diese Zahl entsprach der



Anzahl der Sitze, die man für die jeweilige Partei erwartete.

Die Wahrscheinlichkeit, dass Partei  $j$  Wahlkreis  $i$  gewinnt ist  $r_{ij} / \sum_j r_{ij}$ . Die Formel, die zur Berechnung der Wahrscheinlichkeiten verwendet wurde, lautet:

$$r_{ij} = \exp \left[ - \left( \frac{\max_j \{s_{ij}\} - s_{ij}}{\sigma} \right)^\lambda \right]$$

Dabei ist  $s_{ij}$  der prognostizierte Stimmanteil.  $\lambda$  und  $\sigma$  sind Tuning-Parameter.  $\lambda$  gibt an, welche Form der Verteilung man für die Fehler der geschätzten Anteile  $s_{ij}$  annimmt.  $\lambda = 1$  steht für eine Verteilung mit „heavy-tails“. Größere Werte von  $\lambda$  stehen für leichtere Ränder. Das sind Verteilungen, bei denen große Fehler eher seltener auftreten.  $\sigma$  ist der Dispersionsparameter der Fehlerverteilung. Größere Werte von  $\sigma$  stehen für eine größere Unsicherheit bei den vorhergesagten Anteilswerten. Bei der Wahl 2005 wurde  $\sigma = 4$  und  $\lambda = 1,5$  gesetzt. Für diese Werte entschied man sich nach einer Analyse der Exit-Poll-Daten und der tatsächlich ausgezählten Stimmen von 2001 und zusätzlichen Überlegungen zur Variation bei der Wahl 2005.

Am Wahlabend wurden für die beiden Zielvariablen mehrere Regressionsmodelle berechnet. Es wäre eine Möglichkeit gewesen, die Ergebnisse der verschiedenen Regressionsmodelle zu kombinieren, jedoch entschied man sich jeweils für eins der Modelle, um die Prognose möglichst einfach und nachvollziehbar zu halten. Bei dem Modell für die Veränderung des Stimmanteils von Labour wurden nur die Prädiktoren „Turnout 2001“ und „Country dwellers“ verwendet. Das Regressionsmodell für die Veränderungen des Stimmanteils der Liberal Democrats ist etwas komplexer als das für Labour:

$$\frac{LD}{LD + Lab + Con} = 3,9 + 13,7 \cdot Scotland - 9,5 \cdot Confirst; LDsecond - 6,6 \cdot LDfirst + 5,7 \cdot Con - LDmarginal$$

In diesem Regressionsmodell ist ein einflussreicher Faktor in positiver Richtung, ob sich der Wahlkreis in Schottland befindet. In Wahlkreisen, in denen die Liberal Democrats einen Sitz verteidigten oder gegen die Konservativen antraten, mussten sie mit Verlusten bei den Stimmanteilen rechnen. Dieser negative Effekt war allerdings in Wahlkreisen, in denen es bei der letzten Wahl ein recht knappes Ergebnis zwischen den Konservativen und den Liberal Democrats gegeben hatte, weniger stark ausgeprägt.

Da Briefwähler bei Wahltagsbefragungen naturgemäß nicht berücksichtigt werden können, muss man auf Schätzungen über das Verhalten und den Anteil der Briefwähler zurückgreifen. In Großbritannien wurde 2001 vom Meinungsforschungsinstitut MORI bei Vorwählerhebungen unter anderem gefragt, wer beabsichtigt, per Brief zu wählen. Eine solche Befragung wurde auch 2005 wieder durchgeführt und es zeigte sich, dass der Anteil der konservativen Wähler unter den Briefwählern etwas zurückgegangen ist. Dies konnte dann bei den Prognosen berücksichtigt werden, indem man bei den Regressionsmodellen Veränderungen am Intercept vornahm. Beim Modell, das die Veränderungen des Stimmanteils für Labour beschreibt, wurde der Intercept von -4,3 auf -3,9 reduziert (Curtice und Firth 2008).

## 5 Verschiedene Fehlertypen und Fehlerquellen

Obwohl durch Wahltagsbefragungen einige Fehlerquellen umgangen werden können, von denen Vorwählerhebungen und andere politische Umfragen betroffen sind (wie beispielsweise der Unterscheidung zwischen Personen, die tatsächlich wählen werden und solchen, die das nur angeben), haben auch sie ihre Schwierigkeiten. So hat man bei Exit Polls beispielsweise nicht die Möglichkeit, einen zweiten Befragungsversuch an einem anderen Tag bei derselben Person zu unternehmen, wie es bei Telefonumfragen möglich ist.

Grundsätzlich gilt für Wahltagsbefragungen genau wie für jede andere Stichprobe, mit der auf die Grundgesamtheit geschlossen werden soll, dass diese mit einem Zufallsfehler behaftet ist. Die Höhe des Stichprobenfehlers unterscheidet sich je nachdem, wie stark ein bestimmtes Merkmal in der Stichprobe geclustert ist. So weisen Merkmale wie Rasse, die sich oft in einigen Gegenden konzentrieren, größere Zufallsfehler auf als bei Variablen wie Geschlecht, die relativ gleichmäßig auf alle Wahlbezirke verteilt sind.

Neben Fehlern, die durch die Stichprobe verursacht werden, gibt es auch noch zahlreiche andere Fehlerquellen. Im Gegensatz zum Stichprobenfehler kann die Höhe des systematischen Fehlers jedoch nicht zuverlässig kalkuliert werden. Dies hat zur Folge, dass in manchen Fällen ein Stimmvorsprung, der basierend auf dem Zufallsfehler ausreichen würde, doch zu gering für einen Sieg ist. Für den systematischen Fehler gibt es zahlreiche Ursachen. So können Probleme durch Wähler entstehen, die entweder vom Interviewer verpasst werden oder nicht antworten wollen. Es ist ein bekanntes Phänomen in Großbritannien, dass Wähler der Konservativen es ablehnen, an Wahltagsbefragungen teilzunehmen („Shy Torys“); in den USA existiert dieselbe Problematik mit den Anhängern der Republikaner. Auch die verschiedenen Altersgruppen zeigen oftmals eine unterschiedliche Bereitschaft, an den Exit Polls teilzunehmen. Meist lehnen mehr ältere Menschen das Ausfüllen des Fragebogens ab. Obwohl man versucht, die Folgen des Non-Response durch Gewichtung in den Griff zu bekommen, kann man nie ausschließen, dass Wähler, die ein Interview ablehnen, andere Charakteristiken aufweisen als solche, die teilnehmen. Dies kann zur Folge haben, dass man durch Gewichtung noch größere Verzerrungen konstruiert als bereits in den Daten vorhanden sind.

Außerdem kann die Reihenfolge der Fragen zu einer Verzerrung der Antworten führen. Wenn die Frage nach der Beliebtheit des Präsidenten auf die Frage nach der Einschätzung der aktuellen Wirtschaftslage folgt, wird es nicht ausbleiben, dass viele Personen bei einer guten Wirtschaftslage auch eine positivere Einstellung zum Präsidenten angeben.

Ein weiteres Problem sind Wahllokale, in denen mehrere Bezirke abstimmen („co-location“). Da nicht klar ist, zu welchem Bezirk der Wähler gehört, wenn er das Wahlgebäude verlässt, ist es schwierig, die festgelegten Auswahlwahrscheinlichkeiten einzuhalten. In der Regel wird davon

ausgegangen, dass die Wähler, die in den gemeinsamen Wahllokalen abstimmen, sich nicht allzu sehr voneinander unterscheiden und das Befragungsintervall wird mit der Anzahl der dort abstimmenden Wahlbezirke multipliziert. Dabei handelt es sich nicht um ein vernachlässigbares Problem, da bis zu einem Drittel der Abstimmungsbezirke in den USA hiervon betroffen sind (Levy 1983). Ein weiterer Grund, aus dem Befragungsintervalle nicht eingehalten werden können, sind die Abstandsvorschriften zwischen Interviewer und Wahllokal in einigen Bundesstaaten.

Weitere Fehlerquellen sind die falsche Interpretation der Fragen durch die Wähler, Fehler bei der Übertragung der Antworten und Probleme bei der Eingabe der Daten. Außerdem werden die Interviewer bei ihrer Arbeit nicht kontrolliert. Dies kann dazu führen, dass sie das Befragungsintervall nicht einhalten, sondern Wähler befragen, bei denen sie es für wahrscheinlich halten, dass diese antworten. Auch daraus kann leicht eine Verzerrung der Ergebnisse resultieren (Scheuren und Alvey 2008).

Ein Fehler, der immer mehr an Bedeutung gewinnt, sind die fehlenden Briefwähler in den Exit-Poll-Daten. Man kann zwar versuchen, über Telefonumfragen Informationen von den Briefwählern zu erhalten und dafür Anpassungen durchzuführen. Allerdings kann vor der Wahl nur geschätzt werden, welcher Anteil der Wähler seine Stimme nicht persönlich am Wahltag abgegeben wird. Obwohl die Anzahl der Briefwähler immer mehr zugenommen hat, wurde ihr Anteil an der Wählerschaft immer wieder unterschätzt. Wenn sich die Brief- und Urnenwähler voneinander klar unterscheiden, kann dies auch für die Abweichung der Schätzung vom tatsächlichen Ergebnis verantwortlich sein.

## 6 Die Exit Polls in den Jahren 2000 und 2004 in den USA

Bei der Bundestagswahl 2002 gaben die Meinungsforschungsinstitute ungewöhnlich deutlich voneinander abweichende 18-Uhr-Prognosen heraus. Während die Forschungsgruppe Wahlen ein Kopf-an-Kopf-Rennen zwischen Rot-Grün und Schwarz-Gelb vorhersagte, prognostizierte Infratest dimap lange Zeit eine Mehrheit von Union und FDP. Im Laufe des Abends korrigierte aber auch Infratest dimap seine Prognose und schließlich wurde Rot-Grün zum Gewinner der Bundestagswahl 2002 erklärt. Das Problem, dass durch die Exit Polls der Wahlausgang falsch prognostiziert wird, gab es nicht nur in Deutschland, sondern auch in den USA im Jahr 2000 und zum Teil auch 2004. Doch während dieses Thema in Deutschland in der Literatur nicht thematisiert wird, kam es in den USA zu detaillierten Analysen, wie es zu diesem Versagen kommen konnte.

### 6.1 Die Präsidentschaftswahl 2000

Bei der Präsidentschaftswahl 2000 stellte sich relativ bald heraus, dass die Wahl in Florida entschieden werden würde. Wer von den beiden Präsidentschaftskandidaten, George W. Bush oder Al Gore, diesen Staat und damit seine 25 Wahlmänner gewinnen würde, sollte der nächste Präsident der Vereinigten Staaten werden. Es dauerte allerdings über einen Monat, bis schließlich der Supreme Court entschied, dass die anberaumten Nachzählungen beendet werden und damit Georg Bush der neue Präsident der USA wird.

Genau in diesem Staat unterliefen dem Voter News Service einige Fehler, was zu falschen Prognosen führte und damit die allgemeine Verwirrung und Diskussion noch befeuerte. Die Fehler von VNS und den führenden Fernsehanstalten begannen, als um kurz vor 20 Uhr Ostküstenzeit Al Gore zum Sieger erklärt wurde. Um 19 Uhr hatten alle Wahllokale, die sich in der östlichen Zeitzone befinden, geschlossen. Dabei handelte es sich um 95% der Wahllokale Floridas und damit um die große Mehrheit der Wahllokale in diesem Bundesstaat. Daher war es den Medien erlaubt, eine Prognose über den Wahlausgang abzugeben. Da eine Prognose nur auf Grundlage der Exit-Poll-Daten als zu riskant erschien, wartete man jedoch die ersten ausgezählten Stimmen ab. Diese bestätigten aber (fälschlicherweise) die Führung Gores aus den Exit Polls. Für sechs Wahlbezirke, in denen Exit Polls durchgeführt worden waren, lagen bereits die ausgezählten Stimmen vor. Aber der Vergleich der Exit-Poll-Daten mit den Ergebnissen in diesen Stimmbezirken deutete sogar eher noch auf eine Unterschätzung der Stimmen für Gore hin. Auch alle berechneten Schätzungen zeigten eine Führung Gores zwischen fünf und zehn Prozentpunkten, mit t-Werten zwischen 2,0 und 4,8. Wenn man den zu diesem Zeitpunkt besten Schätzer (den mit dem geringsten Zufallsfehler) anwandte, hatte Gore einen Vorsprung von 7,3

Prozentpunkten vor Bush mit einem t-Wert von 3,8. Damit war das Ergebnis hochsignifikant, da der kritische Wert 2,6 betrug. Bei diesem Schätzer handelte es sich um einen Verhältnisschätzer mit nachträglicher geographischer Schichtung, wobei die Gouverneurswahl von 1998 als Vergleichswahl diente. Bei Analysen nach der Wahl stellte sich allerdings heraus, dass dieser Schätzer insgesamt mit einem Fehler von 7,3 Prozentpunkten behaftet war (Mitofsky und Edelman 2002).

Für diesen enormen Fehler gab es mehrere Ursachen. Eine große Schwierigkeit stellten zum Beispiel die vorzeitigen Wähler und die Briefwähler dar. Es wurden zwar Telefonumfragen durchgeführt, um Informationen über die Wähler zu erhalten, die nicht am Wahltag persönlich zur Wahl gehen. Allerdings gab es solche telefonischen Befragungen nur in Kalifornien, Oregon und Washington. Außerdem war der Umfang der Telefonstichproben recht gering. Dies hatte zur Folge, dass man den Anteil der Menschen, die nicht am eigentlichen Wahltag wählten, nur sehr schwer schätzen konnte. Ferner musste geschätzt werden, wie stark sich die Urnenwähler und die Briefwähler unterscheiden. In Florida ging man davon aus, dass 7,5% vorzeitig oder per Brief wählen würden. Tatsächlich waren es aber 12%. Außerdem nahm man an, dass der Anteil der Republikaner an der Briefwahl um 15,3% höher ist als ihr Anteil an der Wahl an der Urne. Es zeigte sich, dass es 23,7% waren. Dadurch reduzierten die Briefwahlstimmen den Vorsprung Gores um 1,7 Prozentpunkte, angenommen hatte man lediglich einen Prozentpunkt (Konner 2003).

Außerdem gab es in Florida während der Wahl einige Komplikationen. So sorgten beispielsweise die „butterfly ballots“ in Palm Beach County für große Verwirrung. Bei den dort verwendeten Lochkarten war das dritte Loch für Gore vorgesehen, obwohl er als zweites auf der Liste stand. Dies führte dazu, dass der Kandidat Patrick Buchanan in diesem County ungewöhnlich viele Stimmen erhielt, da einige Wähler versehentlich für ihn stimmten (Scheuren und Alvey 2008). Außerdem hatte man bei der Berechnung des Verhältnisschätzers eine ungünstige Vergleichswahl verwendet. Diese wies zwar die höchste Korrelation mit dem aktuellen Rennen auf, der Wert war jedoch nur geringfügig besser als bei einer anderen Wahl, die einen deutlich geringeren Fehler geliefert und damit den verfrühten „Call“ verhindert hätte. Bei der Analyse nach der Wahl stellte sich zudem heraus, dass die Auswahl der Stichprobenbezirke für die Wahltagsbefragungen nicht repräsentativ für das Wahlergebnis in Florida war. Als immer mehr ausgezählte Stimmen zur Verfügung standen, die gegen einen eindeutigen Sieg Gores sprachen, nahmen die Nachrichtensender und VNS ihren „Call“ zurück.

Den nächsten Fehler begingen die Fernsehsender, als sie Bush zum Sieger erklärten. Während die Entscheidung für Gore auf Grund einer zu gering ausgefallenen Schätzung des Stichprobenfehlers zustande kam, wurde der „Bush-Call“ durch falsche Annahmen über die noch ausstehenden Stimmen und Fehler bei der Eingabe von ausgezählten Stimmen verursacht. Die Entscheidung Bush zum Sieger zu erklären, basierte weder auf Daten aus der Wahltagsbefragung noch auf Prognosemodellen, sondern ausschließlich auf den ausgezählten Stimmen der Countys. Allerdings gab es bei der Eingabe der Stimmen Komplikationen. So hatte die Eingabe

be des Ergebnisses für einen Wahlbezirk in Volusia County zur Folge, dass sich bei Bush die Anzahl der gesamten Stimmen um 16022 erhöhte, gleichzeitig wurde aber dieselbe Anzahl an Stimmen bei Gore abgezogen. Dieser Fehler wurde beim VNS zwar entdeckt, aber erst relativ spät korrigiert. Zudem wurden auch in Brevard County zu hohe Werte zu Gunsten von Bush eingegeben, was dazu führte, dass man eine deutlichere Führung Bushs annahm, als dies tatsächlich der Fall war. Außerdem hatte man berechnet, dass ungefähr 179.000 Stimmen noch nicht ausgezählt waren. In Wahrheit waren es aber 359.000. Der Grund für diesen Fehler waren Prognosemodelle für die noch ausstehenden Stimmen, die mit zu einfachen Annahmen arbeiteten. Durch die Kombination dieser beiden Fehler kam man zu dem Schluss, dass Gore 63% der noch ausstehenden Stimmen benötigt, um noch gewinnen zu können. Dies war jedoch unrealistisch und so beschlossen die Fernsehsender um 2:15 Uhr, Bush zum Sieger zu erklären. Als die Fehler schließlich korrigiert wurden, zeigte sich, dass der Ausgang sehr knapp werden würde, weshalb man auch den „Call“ für Bush zurücknahm und sich für den Status „too close to call“ entschied (Mitofsky 2003).

Obwohl die Fehler in Florida am einflussreichsten waren und damit auch am meisten beachtet wurden, gab es auch in anderen Bundesstaaten Probleme. Allerdings war die Entscheidung in diesen Staaten nicht so knapp wie in Florida, so dass die prognostizierten klaren Siege zwar in Wirklichkeit recht knappe Entscheidungen waren, die Prognose aber trotzdem zu Gunsten des siegreichen Kandidaten ausfiel (Frankovic 2004).

Nach den Fehlern bei der Wahl 2000 versuchte VNS, sein System zu überarbeiten, was dazu führte, dass 2002 das Computersystem in der Wahlnacht zusammenbrach. Deshalb konnten an diesem Abend keine Wahlanalysen mit Hilfe von Exit-Poll-Daten durchgeführt werden. Dies hatte zur Folge, dass der Voter News Service 2003 aufgelöst wurde (Guterbock und Daves 2003).

Seit 2004 werden die Exit Polls vom National Election Pool (NEP) durchgeführt. Doch im ersten Jahr von NEP kam es wieder zu Problemen am Wahltag.

## 6.2 Die Präsidentschaftswahl 2004

Auch 2004 verursachten die Ergebnisse der Exit Polls Kontroversen. Am Wahltag wurden im Internet nachmittags Zahlen veröffentlicht, die eine Führung des demokratischen Präsidentschaftskandidaten John Kerry zeigten. Im Laufe des Wahlabends verkündeten die Medien jedoch einen deutlichen Sieg von Amtsinhaber George Bush. Dies führte zur Vermutung, dass bei der Wahl nicht alles mit rechten Dingen zugegangen ist. So versuchten beispielsweise Steven Freeman und Joel Bleifuss (2006) in ihrem Buch „Was the 2004 presidential election stolen?“ mit verschiedenen Verfahren nachzuweisen, dass es einen Wahlbetrug gegeben haben musste. Edison Media Research und Mitofsky International, die im Auftrag des National Election Pools (NEP) die Wahltagsbefragungen durchgeführt hatten, veröffentlichten im Januar 2005 einen Bericht, in dem sie zu den Vorkommnissen am Wahltag Stellung bezogen. Die Autoren schrei-

ben, dass die Daten, die im Internet veröffentlicht wurden, lediglich Rohdaten waren und als solche keine Aussagekraft hatten. Außerdem wurde bereits im Laufe des Nachmittags deutlich, dass die Exit-Poll-Daten für einige Staaten fehlerhaft waren. So teilte der NEP den großen Fernsehanstalten um 16:30 Uhr Ostküstenzeit mit, dass die Daten in neun Bundesstaaten möglicherweise mit größeren Fehlern behaftet waren. In sieben Staaten vermutete man eine deutliche Überschätzung des Anteils der Stimmen für Kerry und in zweien nahm man einen zu hohen Stimmanteil für Bush an.

In dem Bericht werden verschiedene Faktoren für die Fehler genannt. Als Hauptursache für die Abweichung zwischen den Exit-Poll-Ergebnissen und den tatsächlichen Wahlergebnissen machte man einen überdurchschnittlichen WPE (within precinct error) ausfindig. Der WPE ist definiert als Differenz zwischen dem Stimmenanteil in einem Wahlbezirk, den der führende Kandidat laut Exit Polls bekommt, und dem Stimmanteil, den er entsprechend den ausgezählten Stimmen erhält. Dieser erhöhte WPE wurde sehr wahrscheinlich von Kerry-Wählern verursacht, die bereitwilliger an den Exit Polls teilnahmen als Bush-Wähler. Man hat festgestellt, dass besonders in Jahren mit hoher Wahlbeteiligung und großem Interesse am Wahlkampf, wie 1992 und 2004, die Wähler des demokratischen Kandidaten in der Exit-Poll-Stichprobe überrepräsentiert sind. Die Wahlbeteiligung in den Jahren 1992 und 2004 lag um rund fünf Prozentpunkte höher als bei den beiden Präsidentschaftswahlen dazwischen und gleichzeitig lag auch der WPE mit -5,0 1992 und -6,5 2004 deutlich über den Werten von 1996 und 2000 (-2,2 bzw. -1,8). Allen gemein ist das negative Vorzeichen, dass die Überschätzung des Stimmanteils für den demokratischen Kandidaten anzeigt. Insgesamt überschätzten die Exit-Poll-Daten 2004 in 26 Staaten den Stimmenanteil für Kerry und in vier Staaten wurden die Stimmen für Bush um mehr als eine Standardabweichung überschätzt.

Eine weitere Ursache für die teilweise schlechte Qualität der Exit-Poll-Daten waren laut Bericht die Interviewer. Diese hätten beispielsweise die vorgegebenen Intervalle nicht eingehalten. Auch das jugendliche Alter vieler Interviewer gilt als Ursache, da bei diesen besonders ältere Wähler seltener bereit waren, den Fragebogen auszufüllen. Schließlich trug auch die geringe Partizipationsrate zu den fehlerhaften Daten bei. Die Verantwortlichen bei Edison Media Research und Mitofsky International zogen daraus die Konsequenzen, dass die Interviewer besser geschult werden müssten und eine größere Anzahl älterer Personen für die Befragungen rekrutiert werden sollte. Schließlich wollte man untersuchen, ob sich durch veränderte Fragebögen die Anzahl der Personen, die den Fragebogen ausfüllen, erhöhen lässt. Weiterhin ist es in den USA zum Teil problematisch, dass die Interviewer sehr weit entfernt von den Wahllokalen stehen müssen, was die Einhaltung des Befragungsintervalls erschwert. Besonders problematisch war die Wahltagsbefragung in Ohio, denn fünf Tage vor der Wahl entschied der Staatssekretär von Ohio, dass die Interviewer beim Exit Poll mindestens 30 Meter vom Wahllokal entfernt stehen müssen. Gegen diese Vorschrift reichte der National Election Pool eine Klage ein, die in der Nacht vor der Präsidentschaftswahl zu seinen Gunsten entschieden wurde. Während die Interviewer noch rechtzeitig davon in Kenntnis gesetzt werden konnten, wussten viele Wahlhelfer vor Ort nichts

von diesem Urteil und so konnte ein großer Teil der Interviewer seine Arbeit erst mit deutlicher Verspätung aufnehmen. Der letzte Interviewer konnte sogar erst um 17 Uhr beginnen. Auch dabei könnte es sich um eine mögliche Ursache für den zu hohen Anteil an Kerry-Wählern in der Stichprobe handeln, da man beobachtet hat, dass demokratische Wähler oftmals eher später am Wahltag ihre Stimme abgeben.

Da es auch 2004 wieder technische Probleme gab, versuchte man auch auf diesem Gebiet Verbesserungen zu erreichen. Schließlich wollte man zukünftig stärkere Anstrengungen unternehmen, die Exit-Poll-Ergebnisse länger unter Verschluss zu halten, um so der unkontrollierten Veröffentlichung von unbearbeiteten Rohdaten entgegenzuwirken. Dieser Plan wurde bei den folgenden Wahlen auch in die Tat umgesetzt. Bei der Präsidentschaftswahl 2008 erhielten nur einige wenige Analysten des NEP vor 17 Uhr Zugang zu den Exit-Poll-Daten und diese wurden vollständig von der Außenwelt abgeschirmt (Blumenthal 2008).

Es wurde auch die zunehmende Anzahl der Wähler, die bereits vor dem eigentlichen Wahltag abstimmen, als Problem für die Qualität der Wahltagsbefragung erkannt. Daher wird man in Zukunft noch stärker auf die telefonische Befragung von Briefwählern und anderen „frühen“ Wählern setzen. Obwohl inzwischen fast ein Fünftel der Wähler nicht regulär am Wahltag seine Stimme abgibt, wurden 2004 nur in den 13 Bundesstaaten mit dem höchsten Anteil an vorzeitigen Wählern und Briefwählern bei der Präsidentschaftswahl 2001 Telefoninterviews durchgeführt (Edison Media Research und Mitofsky International 2005).



## 7 Genauigkeit von Prognosen auf Grundlage von Exit Polls

In diesem Kapitel soll die Qualität der Prognosen auf der Basis von Wahltagsbefragungen bei amerikanischen und deutschen Wahlen beurteilt werden. Dazu werden die amerikanische Präsidentschaftswahl von 2004 und die Bundestagswahlen der Jahre 2002 und 2005 in Deutschland betrachtet. Bevor auf die amerikanischen Ergebnisse eingegangen werden kann, sollen noch die bei der Wahl 2004 verwendeten Fragebögen näher erläutert werden.

### 7.1 Analyse der Exit-Poll-Daten zur amerikanischen Präsidentschaftswahl 2004

#### 7.1.1 Beschreibung des Datensatzes

Der National Election Pool (NEP), bestehend aus ABC News, Associated Press, CBS News, CNN, Fox News und NBC News, finanzierte 2004 die Exit Polls für die amerikanische Präsidentschaftswahl. Edison Media Research und Mitofsky International wurden mit der Durchführung der Exit Polls beauftragt. Außerdem erstellen sie am Wahlabend Analysen und Prognosen für den NEP. Dazu werden neben den Daten aus der Wahltagsbefragung auch die ausgezählten Stimmen einiger Stichprobenbezirke verwendet, welche ebenfalls von Edison Media Research und Mitofsky International zusammengetragen werden. Schließlich liefert die Associated Press noch die ausgezählten Stimmen der einzelnen Countys. Über die Fragen, die bei der Wahltagsbefragung gestellt werden, entscheiden allein die Mitglieder des National Election Pools.

Der Datensatz zur Wahl 2004 enthält die Ergebnisse der Befragungen am Wahltag nach Verlassen des Wahllokals und zusätzliche Daten aus telefonischen Befragungen, die vor dem Wahltag stattgefunden haben, um Informationen über die vorzeitigen Wähler und die Briefwähler zu erhalten. Es wurden in allen 50 Staaten sowie im District of Columbia Befragungen durchgeführt, wobei in Oregon nur telefonische Befragungen stattgefunden haben, da man in diesem Staat nur per Brief wählen kann. Für die Exit Polls wurden 1.480 Wahlbezirke ausgewählt, in denen die Befragungen durchgeführt wurden. Aus diesen 1.480 Wahlbezirken wurde eine Teilstichprobe im Umfang von 250 Wahlbezirken gezogen, die sogenannte nationale Stichprobe. In diesen Wahlbezirken erhielten die Wähler eine der vier Varianten des nationalen Fragebogens. Die Wähler in den Wahlbezirken, die nicht in die nationale Stichprobe gezogen wurden, erhielten staatspezifische Fragebögen. In diesen wurde nicht nur nach dem Stimmverhalten bei der Präsidentschaftswahl gefragt, sondern auch danach, wen der Wähler bei den Gouverneurs-

und Kongresswahlen gewählt hat. In vielen Staaten wurde am Wahltag auch über Referenden abgestimmt. Dazu gab es ebenfalls Fragen im Fragebogen, zum Beispiel in Georgia: „How did you vote on Amendment 1, defining marriage only as a union between a man and a woman?“. Um aus den 51 Staaten-Umfragen eine nationale Umfrage zu machen, bekommt jeder Staat ein Gewicht zugewiesen, entsprechend dem Anteil, den die Stimmen eines Staates an allen Stimmen amerikaweit ausmachen. Mit diesem Wert wird dann die Gewichtung jeder Einzelperson des jeweiligen Staates multipliziert.

Die Wähler in der nationalen Stichprobe erhielten eine von vier verschiedenen Fragebogenversionen, wobei manche Fragen nur auf einem Teil der Fragebögen standen. Die ersten beiden Versionen unterschieden sich nur dadurch, dass bei den Fragen „Which ONE candidate quality mattered most in deciding how you voted for president?“ und „Which ONE issue mattered most in deciding how you voted for president?“ die Reihenfolge der Antwortmöglichkeiten vertauscht wurde. Bei solchen Fragen besteht immer die Gefahr, dass die befragten Personen den ersten Punkt in der Reihe nehmen, der ihnen einigermaßen wichtig erscheint, obwohl möglicherweise weiter unten noch ein wichtigerer Faktor stehen würde. Indem man die Antwortmöglichkeiten vertauscht und vergleicht, wie häufig die einzelnen Antwortmöglichkeiten genannt wurden, kann diese potentielle Verzerrung kontrolliert werden. Bei der vierten Fragebogenversion war es möglich, auf eine aktuelle Entwicklung zu reagieren. Sie beinhaltet die Frage „In your vote for president today, how would you rate the importance of“, worauf eine Leerzeile folgt. Am Sonntag vor der Wahl hat man entschieden, dass die Interviewer in diese Zeile „the Osama bin Laden video“ schreiben sollten. Es gab jedoch auch elf Fragen, die auf jedem Fragebogen zu finden waren. Dabei handelt es sich zum Beispiel um Geschlecht, Rasse, Alter, Religion und für welchen Präsidentschaftskandidaten man gestimmt hat. Bei der Frage nach der Rasse standen Weiß, Schwarz, Hispanic, Asiatisch oder andere Rasse zur Auswahl und in einigen Staaten gab es noch zusätzliche Antwortmöglichkeiten wie American Indian. Bei der nationalen Befragung und in 24 Staaten mit einem Latino-Anteil von mehr als fünf Prozent wurde zusätzlich die Frage gestellt, ob der Wähler „of Hispanic descent“ ist. Diese Frage diente dazu „Hispanic voters“, die als Rasse weiß oder schwarz angegeben haben, als Hispano-Amerikaner zu erkennen.

Der Datensatz enthält zusätzliche Variablen, die nicht in den Fragebögen abgefragt wurden. Dazu gehört eine Backside-Variable, die angibt, ob der Befragte den beidseitig bedruckten Fragebogen umgedreht hat. Sobald der Wähler eine Frage auf der Rückseite beantwortet hat, steht bei dieser Variable ein „Ja“. Außerdem werden die Gewichte für jeden befragten Wähler angegeben. In acht Staaten gab es in den Wahlbezirken mit mehr als 20% Hispanics auch Fragebögen in spanischer Sprache. Diese Staaten sind Arizona, California, Colorado, Florida, New Jersey, New Mexico, New York and Texas. Zusätzlich gab es noch spanische Fragebögen in einem Wahlbezirk von Louisiana, der in der nationalen Stichprobe enthalten war. Für diese Staaten gibt es eine Spalte, die angibt, ob der Fragebogen in englischer oder spanischer Sprache war. Außerdem enthält der Datensatz noch die Variable „call“. Die Interviewer geben ihre Ergebnisse dreimal am Tag an die Zentrale weiter und die Variable „call“ gibt an, bei welchem dieser drei

Berichte eine bestimmte Person durchgegeben wurde.

Die Telefonbefragung für die vorzeitigen Wähler und die Briefwähler wurde in den zwölf Staaten durchgeführt, die bei der Wahl 2000 den größten Anteil der frühen Wähler und Briefwähler stellten und in Oregon, wo nur per Brief gewählt werden kann. Für die meisten dieser Staaten liegen zwischen 300 und 500 telefonische Interviews vor. Lediglich in Washington und Oregon sind es mehr. Die Haushalte für die Telefonbefragung wurden mit Hilfe von Random Digit Dialing ausgewählt und innerhalb der Haushalte wurde anschließend zufällig eine Person ausgesucht. Die Befragung fand in der Woche vor der Wahl statt und es wurden diesselben Fragen gestellt wie bei den Befragungen am Wahltag. Am Wahltag wurden diese Ergebnisse dann mit den Daten der Exit Polls kombiniert. Dazu wurde an den Daten aus der telefonischen Befragung eine Gewichtung für die Auswahlwahrscheinlichkeiten durchgeführt und zusätzlich wurde eine demographische Gewichtung vorgenommen. Diese demographische Gewichtung basierte auf allen Befragten über 18 Jahren, egal ob sie vorzeitig oder per Brief gewählt haben. Für die nationale Stichprobe wurden 500 Telefoninterviews pro Fragebogen durchgeführt, also insgesamt 2000 (National Election Pool 2004).

### 7.1.2 Ergebnisse für den Exit Poll 2004

Tabelle 1 enthält die Ergebnisse, die man bekommt, wenn man mit dem veröffentlichten Datensatz die Wahlergebnisse berechnet. (Personen, die angegeben hatten, bei der Präsidentschaftswahl nicht gewählt zu haben oder bei denen der entsprechende Wert fehlte, wurden aus der Analyse ausgeschlossen.) Dabei stellt man fest, dass die Ergebnisse bis auf einige wenige Ausnahmen sehr präzise sind, teilweise sogar auf mehrere Stellen hinter dem Komma genau. Lediglich in Montana hat man eine deutlichere Abweichung von zwei Prozentpunkten beim Stimmanteil für Bush. Der Grund für dieses Ergebnis sind die Gewichtungen, die an den Daten vorgenommen werden. Neben einer Gewichtung für die Auswahlwahrscheinlichkeiten und die Antwortverweigerer werden die Daten noch an die tatsächlich ausgezählten Stimmen einer Region angepasst. So steht im Bericht von Edison Media Research und Mitofsky International (2005: 5):

*„As in past elections, the final exit poll data used for analysis in 2004 was adjusted to match the actual vote returns by geographic region within each state. Thus, the discrepancy due to differing response rates was minimized and did not significantly affect the analysis of the vote. The exit polls reliably describe the composition of the electorate and how certain demographic subgroups voted.“*

Am Wahlabend kann man auf diese Weise eine Verzerrung zu Gunsten eines Kandidaten ausgleichen und für die Wahlanalyse hofft man, dass durch die Anpassung der Stimmanteile aus den Exit Polls an die ausgezählten Stimmen auch die Verteilung der übrigen Variablen an die Grundgesamtheit angepasst wird. Wenn man jedoch nicht am Wahlverhalten bestimmter

Gruppen interessiert ist, sondern die Prognosegenauigkeit der Wahltagsbefragungen beurteilen möchte, dann ist ein solcher Datensatz nicht zielführend. Es gibt jedoch keine frei zugänglichen Datensätze, bei denen nur eine Gewichtung für die Auswahlwahrscheinlichkeiten und die Antwortverweigerer vorgenommen wurde. In wissenschaftlicher Hinsicht ist es bedauerlich, dass man als Außenstehender nur einen sehr begrenzten Einblick in die Vorgehensweisen und Ergebnisse bei Wahltagsbefragungen erhält. Auf diese Weise wird eine kritische Auseinandersetzung mit den Methoden und eine eventuelle Verbesserung der verwendeten Verfahren von außen sehr erschwert, wenn nicht sogar verhindert.

Tabelle 1: Vergleich der Exit-Poll-Ergebnisse mit den tatsächlichen Resultaten

State	Bush: Exit Poll	Bush: Ergebnis	Differenz	Kerry: Exit Poll	Kerry: Ergebnis	Differenz
Alabama	0.62034	0.62490	-0.00456	0.36936	0.36862	0.00074
Alaska	0.61791	0.61220	0.00571	0.35043	0.35607	-0.00564
Arizona	0.54640	0.54949	-0.00309	0.44629	0.44461	0.00168
Arkansas	0.53959	0.54306	-0.00347	0.45208	0.44548	0.00660
California	0.44130	0.44432	-0.00302	0.53636	0.54397	-0.00761
Colorado	0.52254	0.51711	0.00543	0.46492	0.47038	-0.00546
Connecticut	0.44438	0.43948	0.00490	0.54771	0.54314	0.00457
D.C.	0.09344	0.09361	-0.00017	0.89475	0.89383	0.00092
Delaware	0.45695	0.45753	-0.00058	0.53876	0.53347	0.00529
Florida	0.51678	0.52098	-0.00420	0.47437	0.47091	0.00346
Georgia	0.58615	0.58029	0.00586	0.41019	0.41414	-0.00395
Hawaii	0.45463	0.45265	0.00198	0.54537	0.54010	0.00527
Idaho	0.68420	0.68519	-0.00099	0.30411	0.30321	0.00090
Illinois	0.44231	0.44516	-0.00285	0.55176	0.54869	0.00307
Indiana	0.59611	0.59981	-0.00370	0.39343	0.39287	0.00056
Iowa	0.50023	0.49838	0.00185	0.49136	0.49171	-0.00035
Kansas	0.62783	0.62006	0.00777	0.36088	0.36625	-0.00537
Kentucky	0.59583	0.59550	0.00033	0.39765	0.39688	0.00077
Louisiana	0.56717	0.56722	-0.00005	0.42138	0.42216	-0.00078
Maine	0.45251	0.44577	0.00674	0.52727	0.53573	-0.00846
Maryland	0.42962	0.42979	-0.00017	0.56167	0.55972	0.00195
Massachusetts	0.37069	0.36928	0.00141	0.61736	0.62188	-0.00452
Michigan	0.47919	0.47812	0.00107	0.50570	0.51231	-0.00661
Minnesota	0.47432	0.47656	-0.00224	0.51375	0.51135	0.00240
Mississippi	0.59052	0.59453	-0.00401	0.40574	0.39732	0.00842
Missouri	0.52785	0.53321	-0.00536	0.46358	0.46122	0.00236
Montana	0.57048	0.59068	-0.02020	0.39830	0.38565	0.01265
Nebraska	0.65641	0.65978	-0.00337	0.33289	0.32721	0.00568
Nevada	0.50359	0.50470	-0.00111	0.48371	0.47878	0.00493
New Hampshire	0.48713	0.48983	-0.00270	0.50213	0.50355	-0.00142
New Jersey	0.46248	0.46239	0.00009	0.52908	0.52923	-0.00015
New Mexico	0.49853	0.49838	0.00015	0.48697	0.49047	-0.00350
New York	0.40578	0.40085	0.00493	0.58127	0.58374	-0.00247
North Carolina	0.56063	0.56053	0.00010	0.42752	0.43611	-0.00859
North Dakota	0.62754	0.62863	-0.00109	0.35319	0.35500	-0.00181
Ohio	0.50809	0.50817	-0.00008	0.48776	0.48710	0.00066
Oklahoma	0.65325	0.65570	-0.00245	0.34675	0.34430	0.00245
Oregon	0.47734	0.47424	0.00310	0.51523	0.51600	-0.00077
Pennsylvania	0.48274	0.48456	-0.00182	0.51264	0.50958	0.00306
Rhode Island	0.38770	0.38746	0.00024	0.59642	0.59539	0.00103
South Carolina	0.57968	0.57981	-0.00013	0.41410	0.40903	0.00507
South Dakota	0.60125	0.59911	0.00214	0.38215	0.38444	-0.00229
Tennessee	0.56703	0.56800	-0.00097	0.42205	0.42526	-0.00321
Texas	0.61576	0.61188	0.00388	0.37931	0.38288	-0.00357
Utah	0.71086	0.71539	-0.00453	0.26411	0.25997	0.00414
Vermont	0.38861	0.38921	-0.00060	0.58742	0.59119	-0.00377
Virginia	0.53962	0.53774	0.00188	0.45484	0.45562	-0.00078
Washington	0.45737	0.45640	0.00097	0.52765	0.52821	-0.00056
West Virginia	0.55526	0.56071	-0.00545	0.43603	0.43205	0.00398
Wisconsin	0.49080	0.49369	-0.00289	0.50097	0.49749	0.00348
Wyoming	0.68933	0.68998	-0.00065	0.29182	0.29132	0.00050

Die verwendeten Daten stammen aus dem vom Roper Center veröffentlichten Datensatz zur Präsidentschaftswahl 2004 (National Election Pool 2004).

## 7.2 Genauigkeit der 18-Uhr-Prognose in Deutschland

In Deutschland sind die 18-Uhr-Prognosen in den meisten Fällen sehr genau. Ausnahmen stellen beispielsweise die Landtagswahlen 1994 in Sachsen und Sachsen-Anhalt dar. Die kumulierte Abweichung betrug für Infas in Sachsen 11,8 Prozentpunkte und in Sachsen-Anhalt 8 Prozentpunkte. Mit 11,2 Prozentpunkten Abweichung in Sachsen schnitt auch die Forschungsgruppe Wahlen nicht viel besser ab. Bei vielen Wahlen waren die Prognosen jedoch wesentlich besser, wie man Tabelle 6 im Anhang entnehmen kann. Diese enthält die kumulierten Abweichungen für 34 deutsche Wahlen zwischen 1993 und 2000.

Die folgenden Tabellen erhalten die 18-Uhr-Prognosen von Infratest dimap und der Forschungsgruppe Wahlen für die Bundestagswahlen 2002 und 2005 und die amtlichen Endergebnisse.

Tabelle 2: Bundestagswahl 2002

Partei	Endergebnis	Infratest dimap		Forschungsgruppe Wahlen	
		18-Uhr-Prognose	Abweichung	18-Uhr-Prognose	Abweichung
SPD	38,5	37,0	1,5	38,0	0,5
CDU/CSU	38,5	39,0	0,5	38,0	0,5
Grüne	8,6	9,5	0,9	9,0	0,4
FDP	7,4	7,0	0,4	7,5	0,1
PDS	4,0	4,3	0,3	4,0	0,0
Sonstige	3,0	3,2	0,2	3,5	0,5

Von Infratest dimap wurden in 400 Stimmbezirken 20.482 Wähler am Wahltag befragt. Die Forschungsgruppe Wahlen befragte 20.561 Personen und in sieben ausgewählten Wahlkreisen wurden zusätzlich 9.435 Wähler zur Ermittlung von möglichen PDS-Direktmandaten interviewt. Die Stichprobe enthielt bei der FGW 244 Wahllokale (Infratest dimap 2002, Forschungsgruppe Wahlen 2002).

Tabelle 3: Bundestagswahl 2005

Partei	Endergebnis	Infratest dimap		Forschungsgruppe Wahlen	
		18-Uhr-Prognose	Abweichung	18-Uhr-Prognose	Abweichung
SPD	34,3	34,0	0,3	33,0	1,3
CDU/CSU	35,2	35,5	0,3	37,0	1,8
Grüne	8,1	8,5	0,4	8,0	0,1
FDP	9,8	10,5	0,7	10,5	0,7
Linke	8,7	7,5	1,2	8,0	0,7
Sonstige	3,9	4,0	0,1	3,5	0,4

In 580 Stimmbezirken wurden von Infratest dimap 102.713 Wähler nach ihrem Stimmverhalten, Alter und Geschlecht gefragt und 16.187 Personen erhielten einen ausführlichen Fragebogen mit

weiteren wahlrelevanten Fragen. Die Forschungsgruppe Wahlen führte in 249 Wahlbezirken Interviews mit 24.620 Wählern durch (Infratest dimap 2005, Forschungsgruppe Wahlen 2005).

2002 wich die 18-Uhr-Prognose von Infratest dimap zwar „nur“ um 3,8 Prozentpunkte vom tatsächlichen Ergebnis ab. Besonders schwerwiegend waren die Abweichungen jedoch, da es dadurch nach einem Sieg für die Unionsparteien aussah. Schließlich konnte sich jedoch die Regierungskoalition aus SPD und Grünen doch noch den Machterhalt sichern. Im Gegensatz zu Infratest dimap prognostizierte man bei der Forschungsgruppe Wahlen schon um 18 Uhr einen sehr knappen Wahlausgang.

Bei der Bundestagswahl 2005 dagegen konnte Infratest dimap mit einer weitaus besseren Prognose als die Konkurrenz von der FGW überzeugen. In diesem Jahr betrug die kumulierte Abweichung der 18-Uhr-Prognose bei der Forschungsgruppe Wahlen fünf Prozentpunkte, während Infratest dimap „lediglich“ um drei Prozentpunkte daneben lag, wobei ein Großteil auf Abweichungen bei der Linkspartei zurückzuführen war. Dieser Erfolg könnte möglicherweise auf die Veränderungen bei der Befragung zurückzuführen sein. So hat Infratest dimap 2005 eine sehr hohe Fallzahl bei der Wahltagsbefragung erreicht, indem sie bei den meisten Befragten den Fragebogen bis auf die elementarsten Fragen (Alter, Geschlecht und Wahlverhalten) gekürzt hat. Damit trotzdem noch eine ausführliche Wahlanalyse möglich ist, bekamen außerdem gut 16.000 Personen einen ausführlicheren Fragebogen.

Tabelle 4 und Tabelle 5 im Anhang enthalten Informationen zum Abschneiden von Infratest dimap und der Forschungsgruppe Wahlen bei der Landtagswahl 2000 in Schleswig-Holstein und bei der Landtagswahl 2009 in Hessen enthalten.

## 8 Aktuelle Forschungsfragen

### 8.1 Die Befragung von Latino-Wählern

Während zum Zeitpunkt der Entstehung von politischen Meinungsumfragen und Wahltagsbefragungen in den USA die meisten Befragten Weiße waren, hat sich das Bild der amerikanischen Gesellschaft in den letzten Jahren und Jahrzehnten stark gewandelt. In einigen Bundesstaaten liegt der Anteil der weißen Bevölkerung bereits unter 50%. Allerdings enthalten die Stichproben für die Exit Polls noch immer vornehmlich Bezirke mit mehrheitlich weißer Bevölkerung. Für die schwarzen Wähler hat man aber bereits in den 1970er Jahren erkannt, dass das Wahlverhalten stark davon abhängt, ob sie in einer eher heterogenen oder eher homogenen Nachbarschaft leben. Wenn die schwarzen Wähler in der Stichprobe jedoch hauptsächlich aus „weißen“ Wahlbezirken stammen, dann vermittelt das ein falsches Bild vom Wahlverhalten der schwarzen Bevölkerung. Diese Erkenntnis schlug sich dann auch bei der Stichprobenziehung nieder, indem man dafür sorgte, dass Wahlbezirke mit einem hohen Anteil schwarzer Bevölkerung verstärkt in der Stichprobe auftauchen. Inzwischen stellen jedoch die Lateinamerikaner einen immer größeren Teil der amerikanischen Bevölkerung. Es stellt sich nun die Frage, ob die Stimmen der Latino-Wähler (und anderer Minderheiten) bei Wahltagsbefragungen korrekt erfasst werden. Der National Election Pool kam 2004 zu dem Ergebnis, dass Bush 45% bei den Latinos erhalten hatte. Andere Untersuchungen, die nur Latinos befragten, kamen zu wesentlich geringeren Werten im Bereich von etwa 30%. Eine mögliche Erklärung für dieses Ergebnis wäre eine falsche Auswahl der Stichprobenbezirke. Dadurch, dass bei der Wahl 2004 nur fünf Bezirke ausgewählt wurden, die hauptsächlich von Latinos bewohnt werden, stammten die meisten befragten Latinos aus Gegenden mit heterogener Bevölkerung. Es ist jedoch sehr wahrscheinlich, dass dieses Vorgehen, genau wie bei den schwarzen Wählern, ein verzerrtes Bild vom Wahlverhalten der Latinos liefert. Ein großes Problem ist, dass die Zunahme von ethnischen und rassischen Minderheiten nicht mit einer Vermischung, sondern eher mit einer noch stärkeren Trennung von ethnischen und rassischen Gruppen einhergegangen ist. Dies macht es umso bedeutsamer, auch Bezirke auszuwählen, die vornehmlich von Schwarzen, Latinos oder anderen Gruppen bewohnt werden.

Francisco Pedraza und Matt Barreto beschreiben in ihrem Artikel eine Studie der University of Washington. Für diese Studie wurde eine Wahltagsbefragung in Seattle-King County durchgeführt, wobei die Stichprobenbezirke einmal zufällig ausgewählt wurden und einmal nach Rasse geschichtet wurde. Es wurden bewusst Bezirke ausgewählt, die von Schwarzen, Latinos, Asiaten oder Weißen dominiert werden und zusätzliche einige Bezirke, die gemischtrassig waren. Bei dieser Studie zeigte sich eine deutliche Verbesserung der Schätzungen, wenn die Bezirke zuerst



nach Rassen geschichtet werden. Die Verbesserung betrifft sogar die Ergebnisse für weiße Wähler. Dies liegt vermutlich daran, dass es auch in Bezirken, in denen eine eigentliche Minderheit die Mehrheit stellt, Weiße leben. Deren Wahlverhalten könnte sich wiederum von dem solcher Weißen unterscheiden, die in rein weißen Bezirken leben. Wenn Gebiete, die viele Bewohner aus Minderheitengruppen haben, in der Studie unterrepräsentiert sind, betrifft das auch die Weißen, die in solchen Gebieten wohnen. Zusätzlich zur Schichtung benutzte die University of Washington Fragebögen in fünf verschiedenen Sprachen. Dabei handelt es sich um eine gute Idee, die geringen Partizipationsraten zu erhöhen. Allerdings ist diese Idee auf nationaler Ebene nur mit einem relativ hohem Mehraufwand realisierbar, da es bereits sehr viele Fragebogenvarianten gibt. Wenn diese neben Englisch und Spanisch auch noch in weiteren Sprachen zur Verfügung gestellt werden sollen, benötigt man dafür zusätzliche finanzielle Mittel (Pedraza und Barreto 2008).

## 8.2 Vorzeitige Wähler und Briefwähler

In den vorhergehenden Punkten sind die Folgen der zunehmenden Bedeutung der Briefwahl bereits angeklungen. Besonders gravierend sind die Auswirkungen, wenn sich die Urnenwähler systematisch von den Briefwählern unterscheiden. Diesen Fall hatte man zum Beispiel in Florida bis zum Jahr 2001. Bis dahin gab es in Florida nämlich relativ strikte Bestimmungen, wer per Briefwahl wählen durfte. Daher waren ein Großteil der Briefwähler Militärangehörige und Geschäftsleute auf Reisen. Da diese beiden Personengruppen tendenziell eher republikanisch wählen, war es üblich, dass die republikanischen Kandidaten bei den Briefwählern einen über 20 % höheren Anteil erreichten als bei den Urnenwählern. Seit diese Bestimmungen gelockert wurden, gibt es jedoch die berechtigte Vermutung, dass sich auch das Wahlverhalten von Brief- und Urnenwählern angenähert hat. Zu diesem Thema wurde zum Beispiel in Kalifornien eine Untersuchung durchgeführt und man kam zu der Erkenntnis, dass sich das Wahlverhalten der Briefwähler nicht sonderlich von dem der Urnenwähler unterscheidet. Allerdings handelt es sich dabei lediglich um Untersuchungen für einen Staat mit lockeren Briefwahlgesetzen (Barreto et al. 2006). In anderen Staaten dagegen kann es nach wie vor Unterschiede geben.

Allerdings stellten Barreto et al. auch fest, dass sich die Brief- und die Urnenwähler in anderen Charakteristiken als dem Wahlverhalten durchaus unterscheiden. So sind Briefwähler in der Mehrzahl besser ausgebildet und älter als der Durchschnittswähler. Bei der nationalen Stichprobe zur Präsidentschaftswahl 2004 unterscheiden sich sowohl gewichtet als auch ungewichtet die Urnenwähler von den Briefwählern beziehungsweise den vorzeitigen Wählern um weniger als einen Prozentpunkt, wobei man bei Verwendung der ungewichteten Daten eine leichte Bevorzugung Bushs durch die Briefwähler erhält.

Allerdings könnte es Probleme bei der wahlanalytischen Arbeit mit den Daten geben, wenn sich die Briefwähler und die Urnenwähler in ihren soziodemographischen Merkmalen unterscheiden. Wenn man die Briefwähler also ganz außer acht lässt, könnte man den Eindruck bekommen,

dass zum Beispiel die Wahlbeteiligung unter den älteren Menschen gesunken ist. Sie könnten aber auch per Brief gewählt haben, anstatt persönlich zur Wahl zu gehen. Daher wird es in Zukunft nötig sein, einen immer größeren Anteil der Wähler auf anderen Wegen als durch die klassischen Wahltagsbefragungen nach Verlassen des Wahllokals zu befragen. Auf welchem Wege dies geschehen soll, ist eine Forschungsfrage, die große Aufmerksamkeit verdient. Momentan werden die Befragungen in den USA mit Hilfe von Telefoninterviews durchgeführt. Allerdings handelt es sich dabei um ein recht kostspieliges Unterfangen, was zur Folge hat, dass die Fallzahlen noch relativ gering sind. So wurden bei der amerikanischen Präsidentschaftswahl 2004 nur in den 13 Bundesstaaten mit dem höchsten Briefwähleranteil bei der Wahl 2000 Telefonumfragen durchgeführt. Wenn sich jedoch die Briefwähler zwischen den Bundesstaaten unterscheiden, zum Beispiel aufgrund von restriktiven Wahlgesetzen, so ist das nur ein sehr unbefriedigender Ansatz.

## 9 Eine kritische Auseinandersetzung mit Exit Polls

In dieser Arbeit wurden unter anderem die Stärken und Schwächen von Wahltagsbefragungen thematisiert. Diese Stärken und Schwächen sind der Grund, aus dem es immer wieder kritische Stimmen zu den Exit Polls gab. Einerseits wurde sie aufgrund ihrer exakten Prognosen und den damit verbundenen möglichen Gefahren angegriffen, auf der anderen Seite gab es aber auch einen großen Aufschrei in den Jahren, in denen die Exit Polls alles andere als zuverlässig waren. Ordentlich durchgeführte Exit Polls sind in vielen Fällen eine verlässliche Grundlage für die Prognose des Wahlausgangs. Gerade das führt aber zu dem „Problem“, dass es sich die Fernsehanstalten erlauben können, auf der Basis von Exit-Poll-Daten Prognose zu betreiben. In den USA kann das allerdings bedeuten, dass der Sieger der Präsidentschaftswahl verkündet werden kann, bevor bereits alle Wahllokale geschlossen sind. Dass für eine vorzeitige Prognose des Wahlausgangs aber nicht unbedingt Daten aus einer Wahltagsbefragung nötig sind, zeigt das Beispiel der amerikanischen Präsidentschaftswahl von 1964. In diesem Jahr verkündete ein CBS-Nachrichtensprecher bereits vor der Schließung aller Wahllokale den Sieg Lyndon Johnsons. In diesem Fall beruhte der „Call“ für Johnson ausschließlich auf ausgezählten Stimmen in Staaten, die ihre Wahllokale bereits geschlossen hatten (Mitofsky 1991). Allerdings machen Exit Polls eine vorzeitige Prognose des Siegers noch einfacher. Seit 1964 gibt es Diskussionen darüber, ob eine verfrühte Bekanntgabe des Wahlsiegers die Stimmen der Amerikaner im Westen weniger wert macht und die Wahl in irgendeiner Weise beeinflusst. Daher braucht es nicht zu verwundern, dass es nach der Wahl 1980, als Ronald Reagan vorzeitig zum Sieger erklärt wurde, Forderungen nach einem Verbot von Exit Polls gab.

Exit Polls haben jedoch nicht nur mit ihrer Genauigkeit „zu kämpfen“. Da die Wahltagsbefragungen in vielen Jahren sehr genaue Ergebnisse lieferten, kann man leicht vergessen, dass es sich dabei um eine Stichprobe handelt, die mit einer gewissen Fehlertoleranz versehen ist. Aufgrund der hohen Fallzahlen sind starke Abweichungen vom tatsächlichen Ergebnis zwar sehr unwahrscheinlich, aber nicht unmöglich. Hinzu kommt, dass nur der Zufallsfehler berechnet werden kann. Wie stark die Stichprobe durch andere Faktoren verzerrt ist, kann dagegen nur vermutet werden.

Es stellt sich also die Frage, ob es sinnvoll ist, eine Menge Geld in Exit Polls zu investieren. Besonders in Deutschland kann man am Sinn von Wahltagsbefragungen zweifeln, da bereits um circa 18:20 Uhr Hochrechnungen auf der Grundlage von ausgezählten Stimmen vorliegen, die in etwa dieselben Ergebnisse liefern wie die 18-Uhr-Prognose. In anderen Ländern wie zum Beispiel Indien gibt es wesentlich bessere Rechtfertigungen für die Exit Polls, da hier einige Tage vergehen, bis der Wahlsieger fest steht. Doch auch für Deutschland gibt es gute Gründe, an der Wahltagsbefragung festzuhalten. Denn nur der große Stichprobenumfang einer Wahl-

tagsbefragung erlaubt eine genaue Analyse der deutschen Wählerschaft und bietet so eine solide empirische Basis für die sonst doch recht theoriegeleitete Politikwissenschaft. Allerdings könnte man überlegen, ob nicht auch in Deutschland eine Zusammenarbeit der Fernsehsender sinnvoll wäre, um sich so den finanziellen Aufwand und das methodische „Know-How“ zu teilen.

## Literatur

- Barreto Matt A., Matthew J. Streb, Mara Martis und Fernando Guerra (2006). Do Absentee Voters Differ from Polling Place Voters? *Public Opinion Quarterly*, 70, 224-234.
- Bautista René, Mario Callegaro, Jose Alberto Vera und Francisco Abundis (2008). Excerpt from: Nonresponse in Exit Poll Methodology: A Case Study in Mexico-Mixed Modes. In *Elections and Exit Polling*, Scheuren Fritz J. und Wendy Alvey (Hg.). Hoboken: John Wiley & Sons, Inc.
- Bishop George F. und Bonnie S. Fisher (1995). Secret Ballots and Self-reports in an Exit-Poll Experiment. *Public Opinion Quarterly*, 59, 568-588.
- Blumenthal Mark (2004). *What About Those German Exit Polls?* Online im Internet unter: Mystery Pollster. [http://www.mysterypollster.com/main/2004/12/what\\_about\\_thos.html](http://www.mysterypollster.com/main/2004/12/what_about_thos.html), [Stand: 11.05.2009].
- Blumenthal Mark (2008). *Looking for Presidential Exit Polls?* Online im Internet unter: Pollster.com. [http://www.pollster.com/blogs/looking\\_for\\_presidential\\_exit.php](http://www.pollster.com/blogs/looking_for_presidential_exit.php), [Stand: 11.05.2009].
- Brown, Philip J., David Firth und Clive D. Payne. (1999). Forecasting on British Election Night 1997. *Journal of the Royal Statistical Society A*, 162, 211-226.
- Curtice John und David Firth (2008). Exit polling in a cold climate: the BBC-ITV experience in Britain in 2005. *Journal of the Royal Statistical Society A*, 171, 509-539.
- Edison Media Research und Mitofsky International (2005). *Evaluation of Edison/Mitofsky Election System 2004*. Online im Internet. <http://www.electionmathematics.org/em-exitpolls/EvaluationJan192005.pdf>, [Stand: 17.05.2009]
- Forschungsgruppe Wahlen (2000). *Wahl in Schleswig-Holstein: eine Analyse der Landtagswahl vom 27. Februar 2000*. Mannheim: Forschungsgruppe Wahlen.
- Forschungsgruppe Wahlen (2002). *Bundestagswahl: eine Analyse der Wahl vom 22. September 2002*. Mannheim: Forschungsgruppe Wahlen.
- Forschungsgruppe Wahlen (2005). *Bundestagswahl: eine Analyse der Wahl vom 18. September 2005*. Mannheim: Forschungsgruppe Wahlen.
- Forschungsgruppe Wahlen (2009). *Wahl in Hessen: eine Analyse der Landtagswahl vom 18. Januar 2009*. Mannheim: Forschungsgruppe Wahlen.
- Frankovic Kathleen A. (1992). Technology and Changing Landscape of Media Polls. In *Media Polls in American Politics*, Thomas E. Mann und Gary R. Orren (Hg.). Washington: The Brookings Institution.
- Frankovic Kathleen A. (2004). Election Reform: The U.S. News Media's Response to the Mistakes of Election 2000. In *Rethinking the vote*, Ann N. Crigler, Marion R. Just und Edward J. McCaffery (Hg.). New York: Oxford University Press.

- Freeman Steven F. und Joel Bleifuss (2006). *Was the 2004 presidential election stolen?* New York: Seven Stories Press.
- Gawiser Sheldon R. und G. Evans Witt (1994). *A journalist's guide to public opinion polls*. Westport: Praeger.
- Gibowski Wolfgang G. (1985). Hochrechnung. In *Pipers Wörterbuch zur Politik 1*, Dieter Nohlen (Hg.). München: Piper.
- Guterbock Thomas M. und Robert P. Daves (2003). The Polls-Review Symposium Election Night 2000 in Perspective: Fluke or Normal Accident? Introduction to the Symposium. *Public Opinion Quarterly*, 67, 1-4.
- Hilmer Richard (2008). Exit Polls - A Lot More than Just a Tool for Election Forecasts. In *Public Opinion Polling in a Globalized World*, Marita Carballo und Ulf Hjelm (Hg.). Berlin, Heidelberg: Springer Verlag.
- Hilmer Richard und Jürgen Hofrichter (2001). Wahltagsbefragungen in den neunziger Jahren. In *Wahlen und Wähler. Analysen aus Anlass der Bundestagswahl 1998*, Hans-Dieter Klingemann und Max Kaase (Hg.). Wiesbaden: Westdeutscher Verlag.
- Hofrichter Jürgen (1999). Exit Polls and Elections Campaigns. In *Handbook of Political Marketing*, Bruce I. Newman (Hg.). Thousand Oaks: Sage.
- Infratest dimap (2000). *WahlREPORT Landtagswahl Schleswig-Holstein 2000*. Berlin: Infratest dimap.
- Infratest dimap (2002). *WahlREPORT Bundestagswahl 2002*. Berlin: Infratest dimap.
- Infratest dimap (2005). *WahlREPORT Bundestagswahl 2005*. Berlin: Infratest dimap.
- Infratest dimap (2009). *WahlREPORT Landtagswahl Hessen 2009*. Berlin: Infratest dimap.
- Klorman Ricardo (1976). Chronopolitics: What Time Do People Vote? *Public Opinion Quarterly*, 40, 182-193.
- Konner Joan (2003). The Case for Caution: This System Is Dangerously Flawed. *Public Opinion Quarterly*, 67, 5-18.
- Levy Mark R. (1983). The Methodology and Performance of Election Day Polls. *Public Opinion Quarterly*, 47, 54-67.
- Matthias Glyn und David Cowling (1995). The ITN exit poll. In *Political communications: the general election campaign of 1992*, Ivor Crewe und Brian Gosschalk (Hg.). Cambridge: Cambridge University Press.
- Merkle Daniel M. und Murray Edelman (2002). Nonresponse in Exit Polls: A Comprehensive Analysis. In *Survey Nonresponse*, Robert M. Groves, J.L. Eltinge und R.J.A. Little (Hg.). New York: John Wiley & Sons, Inc.
- Mitofsky Warren J. (1991). A Short History of Exit Polls. In *Polling and Presidential Election Coverage*, Paul J. Levrakas und Jack K. Holley (Hg.). Newbury Park: Sage Publications.
- Mitofsky Warren J. (2003). Voter News Service after the Fall. *Public Opinion Quarterly*, 67, 45-58.

- Mitofsky Warren J. und Murray Edelman (1995). A Review of the 1992 VRS Exit Polls. In *Presidential Polls and the News Media*, Paul J. Lavrakas, Michael W. Traugott und Peter V. Miller (Hg.). Boulder: Westview Press.
- Mitofsky Warren J. und Murray Edelman (2002). Election Night Estimation. *Journal of Official Statistics*, 18, 165-179.
- Moon Nick (1999). *Opinion polls: History, theory and practice*. Manchester: Manchester University Press.
- Moon Nick und Robin McGregor (1992). Exit Polls - developing a technique. *Journal of the Market Research Society*, 34, 257-268.
- National Election Pool (2004). *National Election Pool Election Exit Polls* [computer file]. Roper Center version 2. New York: Edison Media Research /Mitofsky International [producer], 2004. Storrs: The Roper Center for Public Opinion Research [distributor], 2005.
- Pedraza Francisco und Matt Barreto (2008). Excerpt from: Exit Polls and Ethnic Diversity: How to Improve Estimates and Reduce Bias Among Minority Voters. In *Elections and Exit Polling*, Scheuren Fritz J. und Wendy Alvey (Hg.). Hoboken: John Wiley & Sons, Inc.
- Roth Dieter (2008). *Empirische Wahlforschung: Ursprung, Theorien, Instrumente und Methoden*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Scheuren Fritz J. und Wendy Alvey (Hg.) (2008). *Elections and Exit Polling*. Hoboken: John Wiley & Sons, Inc.

## Anhang

Tabelle 4: Landtagswahl in Schleswig-Holstein 2000

Partei	Infratest dimap			Forschungsgruppe Wahlen	
	Endergebnis	18-Uhr-Prognose	Abweichung	18-Uhr-Prognose	Abweichung
SPD	43,1	41,5	1,6	42,5	0,6
CDU	35,2	35,5	0,3	34,5	0,7
Grüne	6,2	7,5	1,3	7,0	0,8
FDP	7,6	8,5	0,9	8,0	0,4
SSW	4,1	4,0	0,1	4,0	0,1
Sonstige	3,7	3,0	0,7	4,0	0,3

Von Infratest dimap wurden 6257 Wähler in 150 Wahlbezirken befragt. Die kumulierte Abweichung betrug 4,9 Prozentpunkte. Die Forschungsgruppe Wahlen führte Interviews mit 6411 Wählern (in 177 Stimmbezirken) durch. Die kumulierte Abweichung betrug 2,9 Prozentpunkte und war damit deutlich geringer als bei Infratest dimap (Infratest dimap 2000, Forschungsgruppe Wahlen 2000).

Tabelle 5: Landtagswahl in Hessen 2009

Partei	Infratest dimap			Forschungsgruppe Wahlen	
	Endergebnis	18-Uhr-Prognose	Abweichung	18-Uhr-Prognose	Abweichung
CDU	37,2	37,5	0,3	37,5	0,3
SPD	23,7	23,5	0,2	23,5	0,2
FDP	16,2	16,0	0,2	17,0	0,8
Grüne	13,7	14,0	0,3	13,5	0,2
Die Linke	5,4	5,1	0,3	5,0	0,4

Von Infratest dimap wurden in 150 Stimmbezirken 22661 Personen nach Stimmverhalten, Geschlecht und Alter gefragt und zusätzlich erhielten 3613 Wähler einen ausführlicheren Fragebogen. Die kumulierte Abweichung betrug 1,3 Prozentpunkte. Die Forschungsgruppe Wahlen führte in 140 Wahlbezirken Interviews mit 12.414 Wählern nach Verlassen des Wahllokals durch. Die kumulierte Abweichung betrug 1,9 Prozentpunkte (Infratest dimap 2009, Forschungsgruppe Wahlen 2009).



Tabelle 6: Genauigkeit der 18-Uhr-Prognosen in den Jahren 1993 bis 2000

	Infas	FGW
Hamburg 1993	5,8	6,4
Niedersachsen 1994	2,7	2,1
Europawahl 1994	5,8	5,0
Sachsen-Anhalt 1994	8,0	6,2
Brandenburg 1994	2,9	5,1
Sachsen 1994	11,8	11,2
Bayern 1994	4,4	4,6
Bundestagswahl 1994	2,1	1,6
Saarland 1994	7,2	3,8
Thüringen 1994	7,5	4,0
Mecklenburg-Vorpommern 1994	6,0	4,2
Hessen 1995	2,8	3,6
Nordrhein-Westfalen 1995	3,1	2,4
Bremen (Land) 1995	4,4	3,7
Berlin 1995	6,4	6,6
Baden-Württemberg 1996	5,0	2,6
Rheinland-Pfalz 1996	4,4	3,0
Schleswig-Holstein 1996	5,7	3,6
Durchschnitt 1993-1996	5,3	4,4
	Infratest dimap	FGW
Hamburg 1997	3,0	7,2
Niedersachsen 1998	4,4	0,8
Sachsen-Anhalt 1998	7,4	7,2
Bayern 1998	4,6	4,0
Bundestagswahl 1998	2,4	1,0
Mecklenburg-Vorpommern 1998	3,0	5,2
Hessen 1999	2,0	3,2
Bremen (Land) 1999	3,1	3,2
Europawahl 1999	1,2	2,2
Saarland 1999	3,4	5,7
Brandenburg 1999	2,0	2,4
Thüringen 1999	5,2	6,2
Sachsen 1999	8,2	5,8
Berlin 1999	4,6	2,4
Schleswig-Holstein 2000	4,9	2,9
Nordrhein-Westfalen 2000	3,6	1,4
Durchschnitt 1997-2000	3,9	3,8

Aus: „Wahltagsbefragung in den neunziger Jahren“ (Hilmer und Hofrichter 2001)

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

München, den 18. Mai 2009

.....

*(Manuela Hölzlwimmer)*