

Methoden und Verfahrensordnung

Entwurf vom 1. November 2004

Erstellt durch das Steuergremium^a des Instituts

Kontakt:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)

Dillenburger Straße 27

D-51105 Köln

Fon: +49 (0)221 – 35685-0

Fax: +49 (0)221 – 35685-1

Email: methoden@iqwig.de

^a Bastian H, Bender R, Kaiser T, Kirchner H, Lange S, Müller de Cornejo G, Noelle G, Sawicki PT

Inhaltsverzeichnis

1.	Präambel	5
2.	Allgemeine Organisation und Abläufe	7
2.1	Aufbau des Institutes	7
2.2	Organisation der EDV	11
2.3	Internes Qualitätsmanagement	13
2.4	Öffentlichkeitsarbeit	16
2.5	Struktur der externen Auftragsvergabe	17
2.6	Gewährleistung der fachlichen Unabhängigkeit	18
2.7	Struktur der Berichtserstellung mit Auftrag	20
2.8	Struktur der Berichtserstellung ohne Auftrag	21
2.9	Veröffentlichung von Berichten	23
2.10	Struktur der Recherche und Literaturscreening	23
2.11	Entscheidungsfindung	27
2.12	Verteilung der Beweislast bei Fragestellungen	29
2.13	Priority-Setting	30
2.14	Zeitpunkte der Berichtserstellung	31
2.15	Rapid Response	32
3.	Wissenschaftliche Methoden und Biometrie	33
3.1	Darstellung von Effekten und Risiken	33
3.2	Beurteilung statistischer Signifikanz	34
3.3	Beurteilung klinischer Relevanz	35
3.4	Subgruppen-Analysen	36
3.5	Ermittlung patienten-relevanter Effekte und Größen	37
3.6	Aspekte der Beurteilung der Studienqualität	39
3.7	Ermittlung von unerwünschten Wirkungen	40
3.8	Beurteilung von Studien mit „überholten“ Methoden	43
3.9	Beurteilung verschiedener Studienarten	43
3.10	Rangordnung verschiedener Studienarten / Evidenzgrade	45
3.11	Bezug zwischen Studientyp / -art und Fragestellung	45
3.12	Beurteilung nicht oder nicht vollständig publizierter Daten	46
3.13	Umgang mit Rohdaten	46
3.14	Objektivität der Studiendurchführung und -evaluation	47
3.15	Beurteilung von Ergebnisunsicherheiten	47

3.16	Beurteilung nicht verblindbarer Verfahren	48
3.17	Berücksichtigung datenschutzrechtlicher Aspekte / Vertraulichkeit	48
3.18	Berücksichtigung ethischer Aspekte	49
3.19	Darstellung und Beschreibung von Biasarten	50
3.20	Nachweis der Verschiedenheit	52
3.21	Nachweis der Gleichheit	53
3.22	Meta-Analysen	54
3.23	Adjustierung und multifaktorielle Verfahren	55
3.24	Social science and qualitative research	56
4.	Spezielle Beurteilung medizinischer und versorgungsrelevanter Inhalte	61
4.1	Medikamentöse und nicht-medikamentöse Interventionen	61
4.2	Diagnostische Verfahren	63
4.3	Screening	67
4.4	Gesundheitsökonomie	70
4.5	Leitlinien und Disease Management Programme	75
4.6	Systematische Übersichten und HTA-Berichte	81
4.7	Prognose	84
4.8	Individuelle Risikoabschätzung	86
4.9	Beurteilungsmethoden in Abhängigkeit von der Prävalenz und der Art der Erkr.	87
4.10	Beurteilung von nicht erkrankungsassoziierten Beschwerden	88
4.11	Beurteilung von komplementären Verfahren	88
4.12	Beurteilung bevölkerungsweiter Interventionen / Prävention	89
4.13	Qualitätsmanagement in der Versorgung	91
4.14	Beschreibung der Art und Größe des Placeboeffektes	91
5.	Patient Information and Research	93
5.1	Goal	93
5.2	Information products and contents	93
5.3	Priority-setting and topic selection	94
5.4	General editorial and management mechanisms	95
5.5	Evidence basis for the Institute's production processes	96
5.6	Research and development of comprehensive patient information reports	97
5.7	Short summaries of systematic reviews and key 'mega trials'	100
5.8	Communication standards	101
5.9	Translation	103

Allgemeine Anmerkungen:

Dieser Entwurf der Methoden und Verfahrensordnung wird zunächst einem externen Review-Prozess unterworfen. Das Dokument wird regelmäßig überarbeitet, zunächst bei Bedarf, mindestens jedoch in halbjährlichem Abstand.

1. Präambel

Der Gesetzgeber hat mit der Gesundheitsreform 2003 (Gesundheitssystem-Modernisierungsgesetz, GMG) die Etablierung eines neuen staatsunabhängigen Instituts im deutschen Gesundheitswesen vorgesehen. Der Gemeinsame Bundesausschuss (G-BA) hat diese wissenschaftliche Institution in Form einer privaten Stiftung errichtet. Der ausschließliche Zweck der Stiftung ist die Errichtung und Unterhaltung des „Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen“ (IQWiG). Organe der Stiftung sind der Stiftungsrat, sowie ein fünfköpfiger Vorstand. Das Institut ist eine Einrichtung der Stiftung unter verantwortlicher wissenschaftlich unabhängiger Leitung. Beratende Gremien des Institutes sind ein Kuratorium sowie ein wissenschaftlicher Beirat. Das Kuratorium besteht aus 30 Mitgliedern. Der wissenschaftliche Beirat wird durch den Vorstand bestellt, er besteht aus mindestens sechs und bis zu zwölf Wissenschaftlern. Der Sitz des Institutes für Qualität und Wirtschaftlichkeit im Gesundheitswesen ist bis 2008 in Köln.

Dem Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen obliegt künftig die wissenschaftliche Bewertung des medizinischen Nutzens, der Qualität und der Wirtschaftlichkeit von Leistungen. Dazu gehören unter anderem die Bewertung von Behandlungsleitlinien, die Abgabe von Empfehlungen zu strukturierten Behandlungsprogrammen (DMP), die Nutzenbewertung von Arzneimitteln sowie die Herausgabe von Patienteninformationen.

Das Institut wird zu Fragen grundsätzlicher Bedeutung für die Qualität und Wirtschaftlichkeit der im Rahmen der gesetzlichen Krankenversicherung erbrachten Leistungen als fachlich unabhängige wissenschaftliche Einrichtung der Stiftung unter Berücksichtigung alters-, geschlechts- und lebenslagenspezifischer Besonderheiten insbesondere auf folgenden Gebieten tätig:

- Recherche, Darstellung und Bewertung des aktuellen medizinischen Wissensstandes zu diagnostischen und therapeutischen Verfahren bei ausgewählten Krankheiten,
- Erstellung von wissenschaftlichen Ausarbeitungen, Gutachten und Stellungnahmen zu Fragen der Qualität und Wirtschaftlichkeit der im Rahmen der gesetzlichen Krankenversicherung erbrachten Leistungen,
- Bewertung evidenzbasierter Leitlinien für die epidemiologisch wichtigsten Krankheiten,
- Abgabe von Empfehlungen zu Disease-Management-Programmen (DMP),
- Bewertung des Nutzens von Arzneimitteln,
- Bereitstellung von für alle Bürgerinnen und Bürger verständlichen allgemeinen Informationen zur Qualität und Effizienz in der Gesundheitsversorgung.

Aufgabe des Instituts ist es, auf diesen Gebieten durch die Abgabe von Empfehlungen den Gemeinsamen Bundesausschuss in der Wahrnehmung seiner gesetzlichen Aufgaben zu unterstützen.

Das Institut soll durch seine Arbeit zu einer fortwährenden Verbesserung der Qualität der medizinischen Versorgung der Bevölkerung beitragen. Das Institut soll eine eigenständige unabhängige wissenschaftliche Kompetenz entwickeln sowohl in der Beantwortung von herangetragenen Fragestellungen als auch in der Bewertung von versorgungsrelevanten medizinischen Inhalten und Konzepten sowie für patienten-relevanten Forschungsbedarf einschließlich der Weitergabe dieser Information an BMGS, G-BA und die Bevölkerung.

Die Aufgaben des Institutes beziehen sich auf die Erstellung von themenbezogenen Berichten auf Anforderung des G-BA sowie des BMGS, sowie - nach Zustimmung des Vorstandes - auf die Initiierung, Koordination und Publikation von wissenschaftlichen Untersuchungen in Bereichen, die einer Vervollständigung des versorgungsrelevanten medizinischen Wissens bedürfen. Das Institut führt ein regelmäßiges Screening der Literatur durch nach versorgungsrelevanten medizinischen Innovationen einschließlich ihrer Bewertung und gibt diese Informationen allgemeinverständlich weiter. Aufgrund der internationalen Literatur und eigener Recherchen kann das Institut Vorschläge für innovative versorgungsrelevante Forschung unterbreiten, diese initiieren, in der Durchführung begleiten und publizieren.

Das Steuergremium des Institutes, das sich aus der Institutsleitung und den Ressortleitern zusammensetzt, erstellt und modifiziert die Verfahrenweisen und Methoden der Institutsarbeit. Die Publikation und permanente Diskussion dieser Methoden ist ausdrücklich gewollt und soll neben der Transparenz der Institutsarbeit zu einer fortlaufenden dynamischen Verbesserung dieser Methoden beitragen.

2. Allgemeine Organisation und Abläufe

2.1 *Aufbau des Institutes*

Das Institut wird in seiner Errichtungsphase (2004 bis Ende 2005) neu gegründet und aufgebaut, gefolgt von einer Ausbauphase bis Ende 2006, in der die Strukturen verfestigt und erweitert werden. Die Personalgewinnung wird über Ausschreibungen erfolgen. Die Ressortleiter haben ein Mitspracherecht bei der Einstellung der wissenschaftlichen Mitarbeiter für ihr Ressort.

Das Institut besteht von Anfang an aus einer Institutsleitung, einer Institutsverwaltung, einem Institutssteuergremium sowie sechs bis acht feststehenden Ressorts mit je einer Ressortleiterin oder einem Ressortleiter, die sich aus wissenschaftlichen und nicht-wissenschaftlichen Mitarbeitern zusammensetzen, sowie aus wechselnden Projektgruppen, die je nach Bedarf und Auftrag aus den Mitgliedern der einzelnen Ressorts von der Institutsleitung zusammengesetzt werden. Das Steuergremium setzt sich aus der Institutsleitung und den Ressortleitern zusammen. Die Ressorts erarbeiten interne Methoden und Ablaufstrukturen für die Inhalte ihrer Arbeit und stimmen diese mit der Institutsleitung und dem Steuergremium ab. Sie haben ferner die Aufgabe die in ihrem Ressort relevante Literatur zu screenen, einen Bericht über relevante Themen zu verfassen und diese Informationen weiterzugeben. Die Projektgruppen und die jeweiligen Projektgruppenleiter werden je nach Art und Inhalt der Aufträge und nach Bedarf von der Institutsleitung in Abstimmung mit dem Steuergremium gebildet. Ob sich diese Strukturen zu einem eigenen Ressort gruppieren und verfestigen, ist den Erfahrungen in der Zusammenarbeit überlassen.

Der Sitz des Institutes ist Köln. Eine eigene Dependence in Berlin wird zunächst nicht eingerichtet. Ende 2008 ist ein Umzug von Köln nach Berlin geplant.

2.1.1 Ressorts und Institutsleitung

Institutsleitung

Die Institutsleitung besteht aus einer Person, die das Institut leitet, und einer/einem Stellvertreter/in. Die Institutsleitung wird vom Vorstand des Institutes bestellt. Der Institutsleitung ist ein Institutsleitungsreferat zugeordnet. Der Institutsleitung wird ein/eine Verwaltungsleiter/Verwaltungsleiterin zur Seite gestellt. Der Institutsleitung und der Verwaltungsleitung ist ein Sekretariat einschließlich Dokumentations- und Archivabteilung sowie Textverarbeitung beigestellt.

Institutsverwaltung und Projektcontrolling

Die Institutsverwaltung wird von der Institutsleitung bestimmt und soll über eine ökonomische und juristische Kompetenz verfügen. Zusätzliche juristische Kompetenz kann ggf. durch eine externe Beratung gewährleistet werden. Die Institutsverwaltung betreut die internen und externen Geschäfte des Institutes. Der Verwaltungsleiter kann bei entsprechender Eignung gleichzeitig die Aufgaben des Ressortleiters „Gesundheitsökonomie“ wahrnehmen. Ferner obliegt der Verwaltungsleitung die Aufgabe des internen und externen Projektcontrollings einschließlich der Abschlüsse der diesbezüglichen Verträge. Zu den Aufgaben gehört auch, zusammen mit dem Steuergremium einen Haushaltsplan zu erstellen, die Kontrolle der Haushalts- und Wirtschaftsführung und die Erstellung entsprechender Berichte für die Institutsleitung. Weiterhin obliegt der Verwaltung die technische Leitung der EDV, das Personalwesen, Finanz- und Personalbuchhaltung, Bestandsverwaltung und Inventur, Beschaffung, die Organisation von internen und externen Weiterbildungsmaßnahmen und Kongressen sowie die Kommunikation und Koordination mit den Institutsghremien. Die Institutsverwaltung erarbeitet in enger Absprache mit dem Steuergremium und auf Weisung der Institutsleitung verbindliche Ablaufanweisungen zur detaillierten Gestaltung der internen organisatorischen Institutsabläufe (Standard Operating Procedures – SOPs), die von der Institutsleitung erlassen werden.

Ressort „Biometrie“

Dieses Ressort hat die primäre biometrisch-statistische sowie methodische Kompetenz in evidenzbasierter Medizin im Institut. Zu den methodischen Aufgaben dieses Ressorts wird die enge inhaltliche Kooperation mit dem wissenschaftlichen Beirat des Institutes gehören. Die Aufgaben umfassen die Erstellung und fortlaufende Anpassung der biometrischen Arbeitsmethoden des Institutes, die gezielt projektbezogene Beratung der übrigen Ressorts und der Institutsleitung, sowie die Publikation von neuen Erkenntnissen, die in diesem Zusammenhang gewonnen werden.

Ressort „Bewertung von Arzneimitteln“

Dieses Ressort verfügt über eine pharmakologische sowie methodisch-therapeutische Kompetenz in evidenzbasierter Medizin. Die Aufgaben umfassen die Erstellung von Berichten zu therapeutisch-pharmakologischen Fragestellungen und die gezielt projektbezogene Beratung der übrigen Ressorts und der Institutsleitung, sowie die Publikation von neuen Erkenntnissen, die in diesem Zusammenhang gewonnen werden.

Ressort „Bewertung von nicht-medikamentösen therapeutischen Verfahren, diagnostischen Verfahren und Screening-Maßnahmen“

Dieses Ressort verfügt über eine breite biometrische und medizinisch-methodische Kompetenz in evidenzbasierter Medizin in den Bereichen Screening, Diagnostik und Therapie. Die Aufgaben umfassen die Erstellung von Berichten zu nicht-pharmakologischen therapeutischen Fragestellungen, zu diagnostischen Verfahren sowie zu Screening-Maßnahmen und die gezielt projektbezogene Beratung der übrigen Ressorts und des Vorstandes, sowie die Publikation von neuen Erkenntnissen, die in diesem Zusammenhang gewonnen werden.

Ressort „Gesundheitsökonomie“

Die Aufgaben dieses Ressorts umfassen die Erstellung von Berichten zu ökonomischer Bewertung medizinischer Verfahren und anderer Abläufe im Gesundheitswesen auf der Grundlage bereits erarbeiteter valider Effekte dieser Maßnahmen, sowie die Publikation von neuen Erkenntnissen, die in diesem Zusammenhang gewonnen werden. Dieses Ressort soll an der Verwaltung des Institutes angesiedelt sein. Der Verwaltungsleiter soll gleichzeitig sein Ressortleiter sein.

Ressort „Leitlinien und Disease Management Programme“

Die Aufgaben dieses Ressorts umfassen die Bewertung bzw. die Auftragsvergabe zur Bewertung von erstellten Leitlinien sowie die wissenschaftliche Beratung des Gemeinsamen Bundesausschusses bei der Erstellung von Empfehlungen zu Aspekten von Disease Management Programmen (DMP), die Modifikation bestehender Inhalte der Programme und eine entsprechende Kommunikation mit betroffenen Verbänden und Gremien. Die Aufgaben umfassen darüber hinaus eine Kooperation mit der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF), den einzelnen Fachgesellschaften, sowie weiteren Expertengruppen. In Abstimmung mit den Ressorts „Patienteninformation“ und „Öffentlichkeitsarbeit“ werden für wesentliche Empfehlungen der ärztlichen Leitlinien allgemeinverständliche Informationen erstellt und publiziert.

Ressort „Öffentlichkeitsarbeit und Qualitätsmanagement“

Dieses Ressort übernimmt die Organisation und Durchführung der internen und externen Kommunikation. Hierzu zählt u.a. die Information der Bevölkerung in allgemeinverständlicher Sprache über die Aktivitäten des Institutes und die Information der Mitarbeiter des Institutes über aktuelle relevante Entwicklungen im Gesundheitswesen. Dieses Ressort übernimmt die Funktion des Sprechers / der Sprecherin. Es ist verantwortlich für die Publikation von Berichten und Verlautbarungen der übrigen Ressorts. Das Ressort übernimmt die Vorbereitung einer Publikationsplattform in deutsch und englisch für versorgungsrelevante wissenschaftliche

Publikationen. Es ist zuständig für das interne Qualitätsmanagement und organisiert in Zusammenarbeit mit dem Steuergremium die hierfür notwendigen Prozesse inklusive der regelmäßigen Überprüfung des Methodenpapiers. Darüber hinaus übernimmt dieses Ressort die Erarbeitung von Methoden zur Erfassung und Bewertung von Qualitätsaspekten im Gesundheitswesen.

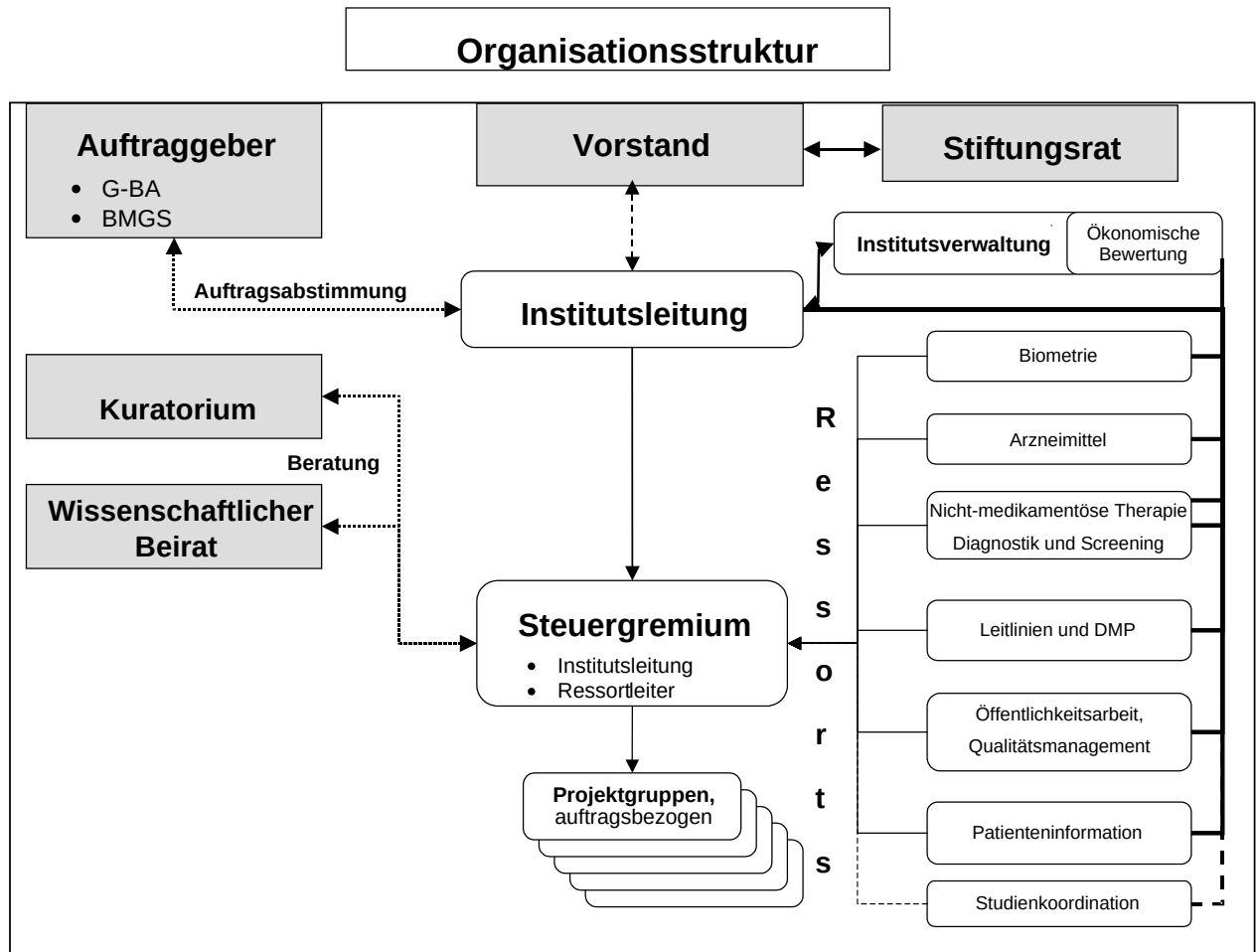
Ressort „Patienteninformation – Erstellung und Methodik“

Dieses Ressort verfügt über eine Kompetenz in wissenschaftlicher Recherche und Beurteilung von Originalpublikationen und zusammenfassenden wissenschaftlichen Berichten, Übersichten und Metaanalysen patienten-relevanter Inhalte. Die Hauptaufgabe dieses Ressorts ist die Publikation patienten-relevanter und verständlicher Information, die ständige Durchsicht und Überwachung der wissenschaftlichen Literatur bezüglich neuer patienten-relevanter Erkenntnisse, sowie ihre Bewertung und Darstellung nach den Kriterien evidenzbasierter Medizin und evidenzbasierter Patienteninformation. Die Zusammenfassung und Publikation dieser Inhalte geschieht nach Rücksprache mit dem Steuergremium und in Kooperation mit dem Ressort „Öffentlichkeitsarbeit“. Darüber hinaus ist das Ressort zuständig für die Erstellung und Verbreitung von Methoden zur allgemeinen Wissensvermittlung wissenschaftlicher medizinischer Zusammenhänge.

Ressort „Studienkoordination“

Dieses Ressort wird nicht in der Aufbauphase des Institutes gebildet. In der Ausbauphase wird das Ressort über einen Mitarbeiterstamm verfügen, der sich nach der Anzahl und der Größe der wissenschaftlichen Projekte richtet. Die Aufgabe wird in der Erstellung von Studienprotokollen zu versorgungsrelevanten Themen mit fehlender zuverlässiger Datenlage bestehen, die Ausschreibung dieser Themen zur Durchführung, die Koordination und Überwachung der Studiendurchführung, sowie die Evaluation der Ergebnisse und ihre Publikation.

Abb.: Organisationsstruktur



2.2 Organisation der EDV

2.2.1 Netz- und Datentechnik

Das Institut verfügt über ein kabelgebundenes sowie mehrstufiges drahtloses Netzwerk für mobile Telefonie über Wireless-LAN Handys, mobile Datenkommunikation im internen Netzwerk mit besonderen Sicherheitsfunktionen sowie mobilen Internetzugang für Besucher des Institutes.

Der Zugang vom internen Netzwerk nach außen wird durch ein Firewall-System gesichert. Dieses System wird auch für den gesicherten Zugang von Extern (z.B. Homeoffice, Einwahl über GSM/UMTS Netze, usw.) genutzt. Zu diesem Zweck wird ein VPN-Zugang („Virtual Private Network“) eingerichtet.

Die Server werden zentral gesichert und sind jeweils als RAID System ausgelegt. Telefonanlage, Server und Teile der Peripherie sind über eine unterbrechungsfreie Stromversorgung (USV) gesichert.

2.2.2 Telefontechnik

Die Telefonanlage arbeitet mit „Voice over IP“ Technologie, d.h. für die Telefonie wird das normale Datennetzwerk (auch das Drahtlose) genutzt. Für die Verbindung zum öffentlichen Netz stehen bis zu 60 ISDN-Kanäle per Primärmultiplex-Anschluss zur Verfügung. Weiterhin können IP-Softphones verwendet werden (Softwaretelefon). Für den Betrieb von analogen Endgeräten, z.B. Fax, stehen Adapter für analoge Endgeräte zur Verfügung. Die Funktionen Fax senden/empfangen sowie Voicemail empfangen werden über eine Integration in das Mail-System realisiert.

2.2.3 Arbeitsplatz

Als Client Rechner werden wahlweise Notebooks oder geräuscharme Miniworkstations verwendet. Als Bildschirme kommen Flatscreens zum Einsatz. Weiterhin steht pro Büro ein kleinerer lokaler Laserdrucker zur Verfügung

2.2.4 Netzdrucker/Kopierer/Scanner

Die Netzdrucker/Kopierer sind als Multifunktionsgeräte konzipiert und verfügen über eine Scan-to-PDF-to-Email Funktionalität. Die Funktionen lassen sich komfortabel in das Netzwerk integrieren; wesentliche Routineaufgaben werden Zeit- und Ressourcen sparend erledigt.

2.2.5 Software

Die Softwarebeschaffung und Lizenzverwaltung findet zentralisiert über die Institutsverwaltung statt.

2.2.6 An-/Beschaffung

Die Anschaffung erfolgt entweder über vierjährige Leasingverträge oder Kauf.

2.2.7 Wartung

Für die Wartung der Netzwerk- und Telefoniekomponenten wurde ein vierjähriger Wartungsvertrag abgeschlossen. Dieser umfasst auch einen kostenlosen Austausch der wesentlichen zentralen Bausteine.

Für die Wartung der Multifunktionsdrucker wurde ein vierjähriger Wartungsvertrag abgeschlossen. Er beinhaltet auch die kostenlose Bereitstellung aller Verbrauchsmaterialien (außer Papier).

2.2.8 Interne Betreuung

Die interne Betreuung der Arbeitsplätze sowie gesamten Netz- und Telefoninfrastruktur erfolgt durch einen Systemadministrator, der über für alle im Institut eingesetzten Komponenten ausreichende und langjährige theoretische und praktische Erfahrungen besitzt.

2.2.9 Internetauftritt

Es wurden verschiedene Domänen (u.a. „IQWG.de“ und „IQWIG.de“) für das Institut reserviert. Weitere, z.B. für eine separate Publikation der Inhalte im Bereich Patienteninformation, folgen ggf. Ein erster Internetauftritt des Institutes ist realisiert. Die Weiterentwicklung des Webauftritts wird mit eigenen Personalressourcen umgesetzt. Hierzu wird ein Content Management System beschafft.

2.2.10 Datenschutz

Die Institutsleitung bestellt einen Datenschutzbeauftragten. Zunächst wird dies die oder der Justiziar(in) des Instituts sein. Auch wenn zunächst nicht davon auszugehen ist, dass innerhalb des Institutes personenbezogene oder –beziehbare Patientendaten verarbeitet werden, müssen die notwendigen Voraussetzungen hierfür alsbald gewährleistet sein. Regelungen durch die jeweils aktuelle Bundesdatenschutzgesetzgebung sind hierbei strikt zu beachten, einzuhalten und zu kontrollieren. Dazu gehört u.a. auch die Erstellung von Weisungen für Angehörige der Verwaltungs- und EDV-Abteilung, die z.B. Einzelheiten im Umgang mit Personaldaten näher regeln.

2.3 Internes Qualitätsmanagement

Grundsätzlich sind viele Prozesse zur Erstellung der Arbeit des Instituts in diesem Papier geregelt. Hier soll der Aspekt der internen Qualitätsentwicklung über die wissenschaftlichen Fragestellungen auf der Ebene der Organisationsentwicklung behandelt werden.

2.3.1 Ziel

Qualitätssicherung und Qualitätsmanagement (QM) wollen die Qualität der Arbeitsprozesse und –ergebnisse wahren und erhöhen. Dies ist nur möglich, wenn eine Bereitschaft besteht, Probleme rechtzeitig zu erkennen, zu analysieren und praktikable Verbesserungsvorschläge zu unterbreiten und umzusetzen.

Mit einem Qualitätsmanagement sollen folgende Ziele erreicht werden:

- Wahrung und Erhöhung der Qualität von Arbeitsprozessen und –ergebnissen
- Identifizierung, Analyse und Behebung von Problemen
- Evaluation, Sicherung und Verbesserung der Struktur-, Prozess- und Ergebnisqualität
- Personal- und Organisationsentwicklung
- Erzeugung nicht monetärer Anreizsysteme
- Personalbindung und stärkere Betriebszugehörigkeit
- Erhöhung der Arbeitszufriedenheit.

Qualitätsmanagement ist in diesem Zusammenhang als eine Methode zu nennen, mit der die Aufbau- und Ablauforganisation kontinuierlich angepasst und sowohl nach außen wie auch nach innen dokumentiert und dargestellt werden, unabhängig davon, ob ein bestimmtes QM-System angewandt wird oder nicht. Dies betrifft vor allem die inneren Prozesse und Abläufe des Institutes sowie im Weiteren die Konzeption und die regelmäßige Verfassung und Veröffentlichung von „Qualitätsberichten“ zur Tätigkeit des Institutes.

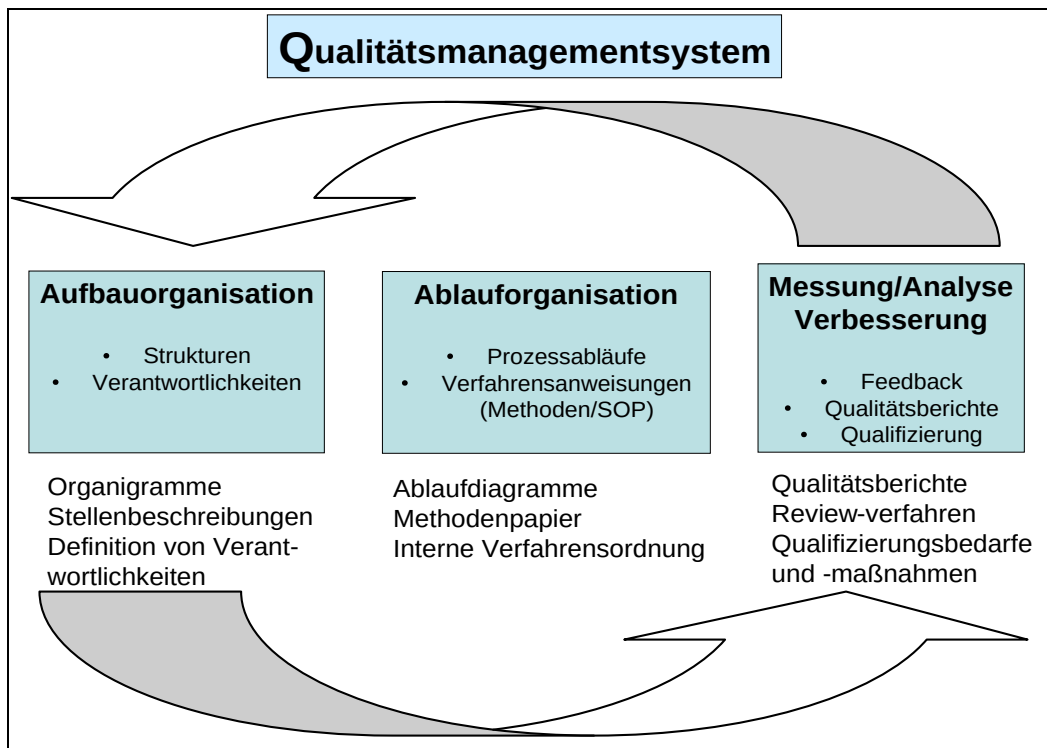
2.3.2 Methode

Im Institut wird ein QM-System implementiert. Um den Prozess der Etablierung dieses Systems voranzutreiben, und auch für die Entscheidung, ob für ein bestimmtes Modell votiert wird, bietet es sich an, eine/n Qualitätsbeauftragte/n festzulegen. Diese/r würde in Zusammenarbeit mit dem Steuergremium und der Leitung Vorschläge für die einzelnen QM-Instrumente erarbeiten und abstimmen sowie für die Umsetzung sorgen. Folgende Aspekte sind dabei vorrangig zu definieren:

- Strukturelle Elemente einschließlich der Definition der Qualifikationsanforderungen und Qualifizierungsmaßnahmen,
- Definition von Abläufen und Schnittstellen innerhalb der Ressorts und zwischen den Ressorts,
- Definition von Vereinbarungen zu Abläufen und Prozessen, die über dieses Papier hinausgehen.

Im nächsten Schritt sind Verfahren für die „Messung und Analyse“ zu definieren, abzustimmen und umzusetzen. Auch dies geschieht in enger Zusammenarbeit mit dem Steuergremium. Der/die Qualitätsbeauftragte hat dafür Sorge zu tragen, dass die Prinzipien des Qualitätsmanagements allen Mitarbeitern des Institutes zugänglich sind und von allen mitgetragen werden.

Abb.: Modell für ein Qualitätsmanagementsystem



2.3.3 Fortbildung und Qualifikation

Fortbildung und Qualifikation gehören zur Personalentwicklung und haben einen bedeutenden Stellenwert für die Leistungsfähigkeit und Arbeitszufriedenheit der Mitarbeiter des Institutes. Es handelt sich dabei um eine innerinstitutionelle Angelegenheit. Der organisatorische Teil wird dabei von der Institutsverwaltung übernommen.

Grundsätzlich lassen sich Fortbildung und Qualifikation in verschiedene Bereiche gliedern. Der ressortspezifische Qualifikationsbedarf ist von den Ressortleitern zu identifizieren und in Absprache mit dem Verwaltungsleiter / Steuergremium zu organisieren.

Ressortübergreifende Qualifizierungsmaßnahmen können z.B. in folgenden Bereichen erfolgen:

- Moderation und Präsentation
- EDV-technische Maßnahmen/Schulungen
- Projektmanagement
- Krisenmanagement
- Konfliktmanagement/Coaching

- EbM-Kurse für Anfänger/Fortgeschrittene
- Fortbildung über statistische Methoden
- Führungskräftetraining
- Medientraining
- Sprachen

Ferner kann das Institut prüfen, inwieweit Fortbildungsangebote / Qualifikationen auch in Zusammenarbeit mit Universitäten oder anderen wissenschaftlichen Einrichtungen formuliert werden können.

Organisationstechnisch sollte die Durchführung dieser Maßnahmen bei der Institutsverwaltung angesiedelt sein. Einmal oder zweimal im Jahr kann eine spezifische Abfrage für die Institutsmitarbeiter durchgeführt werden. Je nach Ergebnis kann dann im Steuergremium die Durchführung dieser Qualifikationen besprochen und Verantwortliche für die Durchführung bestimmt werden.

Darüber hinaus ist zu prüfen, welche institutsinternen Qualifikationen vorhanden sind, auf- oder ausgebaut werden können, um Weiterbildungsmaßnahmen nach Vorgaben der Ärztekammern oder anderer Gremien auch nach außen dokumentieren zu können. Dazu zählen etwa Weiterbildungsermächtigungen (z. B. „Medizinische Informatik“) oder die Unterstützung beim Erwerb von Zertifikaten der GMDS („Biometrie in der Medizin“, „Medizinische Informatik“, „Epidemiologie“ und „Medizinische Dokumentation“). Private Fortbildungs- und Qualifizierungsmaßnahmen sind weitgehend zu unterstützen, soweit diese den institutsinternen Interessen nicht entgegenstehen bzw. diesen entgegenkommen.

2.4 Öffentlichkeitsarbeit

2.4.1 Zielsetzung

Ziel der Öffentlichkeitsarbeit ist die Information der verschiedenen Zielgruppen über die Aktivitäten des Institutes. Dies umfasst die allgemeine Information über die Rolle und die Arbeitsweise des Institutes wie auch die spezifische Information, wie z.B. die Bekanntgabe von Publikationen, Berichten oder auch relevanten Informationen, die anlassbezogen herausgegeben werden (s.a. Abschnitt 2.9). Öffentlichkeitsarbeit bedeutet aber auch die Information der Mitarbeiter des Institutes über wichtige gesundheitspolitische oder wissenschaftliche Diskussionen in den Medien.

Ein wichtiger Baustein für die Öffentlichkeitsarbeit wird die Pressearbeit sein. Hier wird eine Journalistin tätig sein, die mit ihrer Expertise den Aufbau der Pressearbeit unterstützen wird. Die effektive Öffentlichkeitsarbeit wird gewährleistet durch:

- Konzeption, Recherche, Text und Redaktion von Informations- und Präsentationsmaterialien des Institutes
- Repräsentation des Institutes auf wichtigen Veranstaltungen, Kongressen, etc.
- Aufbau und Pflege von Kontakten zu den Medien
- Verfassen von Presstexten
- Vorbereiten von Pressekonferenzen
- Erstellen eines Pressespiegels
- Führen und Betreuen von Fachbesuchern
- Organisation von Informations- und Presseveranstaltungen
- Contentplanung und -pflege des Internetauftrittes

Für die Erstellung eines täglichen oder wöchentlichen Pressespiegels könnte eine Zusammenarbeit mit der Pressestelle des G-BA oder/und anderen organisiert werden.

Der Aufbau eines wirksamen Konzeptes für die interne und externe Kommunikation wird in der ersten Hälfte der Aufbauphase des Institutes zu leisten sein. Hierzu gehört die Erstellung eines Verteilers für wichtige Meldungen vom Institut nach außen, inklusive der Pflege des Verteilers. Für den Aufbau eines „Pressespiegels“ ist die effektive Nutzung von Service-Angeboten oder die Zusammenarbeit mit anderen Gremien zu prüfen und zu organisieren.

2.4.2 Methode

Neben den organisatorischen Aufgaben übernimmt das Ressort Öffentlichkeitsarbeit die Erarbeitung von Textvorschlägen und/oder Sprachregelungen in Zusammenarbeit mit den zuständigen Ressorts. Eine Abstimmung oder Freigabe wird durch das Steuergremium oder durch die Institutsleitung vorgenommen (s.a. Abschnitt 2.15).

2.5 Struktur der externen Auftragsvergabe

Die externe Auftragsvergabe erfolgt nach sach- und themenbezogenen Ausschreibekriterien, die innerhalb des Steuergremiums erarbeitet werden unter Beachtung der rechtlichen Vorgaben. Als

Vorbedingung zur Teilnahme an einer Ausschreibung ist die schriftliche Offenlegung potentieller Tätigkeiten, die geeignet sein könnten, die fachliche Unabhängigkeit zu beeinflussen, im Rahmen einer eidesstattlichen Versicherung unabdingbar (s.a. Abschnitt 2.6).

Ausschreibungen können projektbezogene Teilaufträge beinhalten. Über den Zuschlag berät das Steuergremium nach Ablauf einer angemessenen Ausschreibungsfrist. Die Vertragsgestaltung erfolgt weitgehend einheitlich und transparent nach noch detailliert festzulegenden Kriterien.

Die Beauftragung oder Teilbeauftragung orientiert sich in der Regel an den in den Abschnitten 2.7 und 2.8 festgelegten Abläufen. Die Vergütung erfolgt in zuvor festgelegten Schritten nach Abnahme einzelner Arbeitspakete durch ein internes oder externes Reviewerteam. Die interne und externe Koordination von Autoren und Reviewern sowie das Projektcontrolling finden dabei durch die jeweiligen Projektgruppen nach allgemein noch festzulegenden Maßstäben durch die Institutsleitung und –verwaltung statt.

Die formale Ausgestaltung und Vergabe von Aufträgen sowie die Mittelzuweisung nach Abnahme von Arbeitsergebnissen obliegt der Institutsverwaltung. Weiterhin berät die Institutsverwaltung die Projektgruppen im Einzelfall, z.B. durch die Erarbeitung und Ausgestaltung von Projektplänen und Arbeitsanweisungen. Darüber hinaus wird ein internes Projektcontrolling von der Verwaltung durchgeführt. Die Autorenrechte für alle im Namen des Institutes erbrachten externen Leistungen liegen – sofern nicht ausdrücklich anders vereinbart – beim Institut.

Auftragnehmer, die Aufträge oder Teile hiervon durch Dritte erfüllen lassen, müssen diese im Rahmen der Ausschreibung verbindlich benennen. Für diese gelten insbesondere hinsichtlich ihrer fachlichen Unabhängigkeit die gleichen Maßregeln wie für den Auftragnehmer selbst (s.a. Abschnitt 2.6). Für die ordnungsgemäße Erfüllung und Abwicklung der Aufträge ist unabhängig vom jeweiligen Innenverhältnis immer der Auftragnehmer gegenüber dem Institut verantwortlich.

Über die Vertragsbestandteile, insbesondere jedoch über die wissenschaftlichen Inhalte oder Arbeits(teil)ergebnisse ist gegenseitiges Stillschweigen zu vereinbaren. Der Auftragnehmer verpflichtet sich zusätzlich zur Einhaltung ggf. anzuwendender datenschutzrechtlicher Bestimmungen.

2.6 Gewährleistung der fachlichen Unabhängigkeit

2.6.1 Zielsetzung

Die wissenschaftliche und fachliche Unabhängigkeit des Institutes und der von ihm vertretenen und veröffentlichten Inhalte ist gesetzlich in §139a SGB V sowie in der Satzung der Institutsstiftung

verankert. Der Begriff „Unabhängigkeit“ kann dabei naturgemäß nur annähernd real Anwendung finden, da die Einschätzung der fachlichen Unabhängigkeit individuell unterschiedlich sein kann. Insofern „Unabhängigkeit“ nur ein relatives Ziel sein kann, gilt es gleichzeitig, den Begriff „Transparenz“ einzuführen, um etwaige Entscheidungsprozesse und deren Ergebnisse vor dem Hintergrund einer „relativen Unabhängigkeit“ nach außen und innen nachvollziehbar zu gestalten.

2.6.2 Gewährleistung der externen fachlichen Unabhängigkeit

Vor dem Abschluss eines externen Vertrages zwischen dem Institut und einem externen Experten oder einer externen Institution zur fachlichen Beratung, Studiendurchführung oder Erstellung eines Berichtes muss das Steuergremium entscheiden, ob Bedenken wegen einer potentiellen fachlichen Abhängigkeit bestehen. Dafür müssen alle externen Experten und Institutionen eine Liste mit allen potentiellen Tätigkeiten, die geeignet sein könnten, ihre fachliche Unabhängigkeit zu beeinflussen, beim Steuergremium einreichen. Darüber hinaus muss eine eidesstattliche Erklärung abgegeben werden, dass diese Liste korrekt und vollständig alle potentiellen Abhängigkeiten angibt. Auf ausdrücklichen Wunsch des externen Experten bzw. der externen Institution kann die Liste geheim gehalten werden. Die Zusage der Geheimhaltung seitens des Instituts wird jedoch hinfällig, wenn dem Institut im Nachhinein andere als die auf der Liste genannten und die Unabhängigkeit potenziell beeinflussende Tätigkeit des externen Experten bzw. der externen Institution bekannt werden.

2.6.3 Gewährleistung der internen fachlichen Unabhängigkeit

Die interne fachliche Unabhängigkeit wird soweit möglich durch die Auswahl der Mitarbeiter gewährleistet. Diese müssen im Rahmen ihrer Einstellung glaubhaft ihr bisheriges Tätigkeitsumfeld darlegen als auch versichern, (Neben-) Tätigkeiten, die ihre fachliche Unabhängigkeit potentiell in Frage stellen könnten, im Rahmen einer Tätigkeit im Institut einzustellen. Den wissenschaftlichen Mitarbeitern des Institutes sind bezahlte Nebentätigkeiten, die im weitesten Sinne in einem beruflichen Kontext stehen können, untersagt. Grundsätzlich sind von allen Mitarbeitern alle Nebentätigkeiten gegenüber der Institutsleitung oder –verwaltung anzeigepflichtig. Zu Nebentätigkeiten gehören im weitesten Sinne auch unentgeltliche Ehrenämter wie Aufsichtsrats- oder Vorstandsämter in Vereinen oder Gesellschaften. Zuwiderhandlungen können im Einzelfall zu einer Abmahnung führen bzw. im wiederholten oder schwerwiegenden Fall zu einem Kündigungsgrund. Im Einzelfall entscheidet das Steuergremium, ob ein Mitarbeiter von einer bestimmten Tätigkeit oder einem bestimmten Projekt aus Gründen der Befangenheit ausgeschlossen werden muss.

2.7 Struktur der Berichtserstellung mit Auftrag

Der Ablauf der Übernahme eines Auftrags, der Projektablauf und die Publikation des aufgrund eines externen Auftrags zu erstellenden, vollständigen wissenschaftlichen Berichtes geschieht in fünf Schritten:

1. Formulierung des Auftrags und der Fragestellung.
 - a. Formulierung des Auftrags durch den Auftraggeber.
 - b. Überprüfung der Fragestellung und der Frageinhalte durch das Steuergremium des Institutes, ggf. unter Einbeziehung externer Expertise.
 - c. Abstimmung der Änderungsvorschläge und der neu formulierten Fragestellung mit dem Auftraggeber.
2. Erstellung des Berichtsplans und des ersten Vorberichtes.
 - a. Bildung der auftragsbezogenen Projektgruppe.
 - b. Definition patienten-relevanter Endpunkte.
 - c. Erstellung eines Berichtsplans, in welchem vorab Ziel und Methodik themenspezifisch festgeschrieben werden und welcher die Grundlage der weiteren Arbeitsschritte bildet.
 - d. Externe Vergabe des Projektes oder interne Bearbeitung.
 - e. Hinterlegung des Berichtsplans an einer unabhängigen Stelle.
 - f. Recherche.
 - g. Literaturbewertung.
 - h. Ggf. eigene statistische Arbeiten und Auswertungen.
 - i. Formulierung des ersten Vorberichtes durch die Projektgruppe.
3. Erstellung des zweiten Vorberichtes (internes und externes Review).
 - a. Internes Review des ersten Vorberichtes durch das Steuergremium des Institutes.
 - b. Abstimmung des ersten Vorberichtes mit dem Auftraggeber bezüglich der Vollständigkeit und ggf. Überarbeitung.
 - c. Externes Fachreview des korrigierten ersten Vorberichtes.
 - d. Formulierung des zweiten Vorberichtes durch die Projektgruppe und erneute Vorlage beim Steuergremium.

4. Erstellung des dritten Vorberichtes (Einholen von Stellungnahmen).
 - a. Publikation des zweiten Vorberichtes sowie des Berichtsplans im Internet.
 - b. Einholen von Stellungnahmen zum zweiten Vorbericht seitens des Vorstands, des Kuratoriums, des Stiftungsrats, des wissenschaftlichen Beirats, betroffener Organisationen und Verbände, Patientenorganisationen, der Beauftragten der Bundesregierung für die Belange der Patienten sowie Patientenvertretern im G-BA.
 - c. Zusammenfassung und Bewertung der Ergebnisse des externen Reviews, der Stellungnahmen und der übrigen Kommentare.
 - d. Formulierung des dritten Vorberichtes durch die Projektgruppe.
5. Erstellung des Abschlussberichtes und Publikation.
 - a. Formulierung des fertigen Berichtes durch das Steuergremium des Instituts.
 - b. Weiterleitung des Berichtes an den Auftraggeber durch die Institutsleitung.
 - c. Publikation des vollständigen wissenschaftlichen Berichtes einschließlich der Zusammenfassung von externen Stellungnahmen und Kommentaren.

2.8 Struktur der Berichtserstellung ohne Auftrag

Der Ablauf der Initiierung eines Auftrags, der Projektablauf und die Publikation des ohne eines externen Auftrags zu erstellenden, vollständigen wissenschaftlichen Berichtes geschieht in vier Schritten:

1. Formulierung des Auftrags und der Fragestellung durch das Steuergremium, ggf. unter Einbeziehung externer Expertise.
2. Erstellung des ersten Vorberichtes
 - a. Bildung der auftragsbezogenen Projektgruppe.
 - b. Definition patienten-relevanter Endpunkte.
 - c. Erstellung eines Berichtsplans, in welchem vorab Ziel und Methodik themenspezifisch festgeschrieben werden und welcher die Grundlage der weiteren Arbeitsschritte bildet.
 - d. Hinterlegung des Berichtsplans an einer unabhängigen Stelle.

- e. Externe Vergabe des Projektes oder interne Bearbeitung.
 - f. Recherche.
 - g. Literaturbewertung.
 - h. Ggf. eigene statistische Arbeiten und Auswertungen.
 - i. Formulierung des ersten Vorberichtes durch die Projektgruppe.
3. Erstellung des zweiten Vorberichtes (internes und externes Review).
- a. Internes Review des ersten Vorberichtes durch das Steuergremium des Institutes.
 - b. Externes Fachreview des korrigierten ersten Vorberichtes.
 - c. Formulierung des zweiten Vorberichtes durch die Projektgruppe und erneute Vorlage beim Steuergremium.
4. Erstellung des zweiten Vorberichtes (Einholung von Stellungnahmen)
- a. Publikation des zweiten Vorberichtes sowie des Berichtsplans im Internet.
 - b. Externes Fachreview des ersten Vorberichtes.
 - c. Einholen von Stellungnahmen zum zweiten Vorbericht seitens des Vorstands, des Kuratoriums, des Stiftungsrats, des wissenschaftlichen Beirats, betroffener Organisationen und Verbände, Patientenorganisationen, der Beauftragten der Bundesregierung für die Belange der Patienten sowie Patientenvertretern im G-BA.
 - c. Zusammenfassung und Bewertung der Ergebnisse des externen Reviews, der Abstimmungen und der übrigen Kommentare.
 - d. Formulierung des dritten Vorberichtes durch die Projektgruppe.
5. Erstellung des Abschlussberichtes und Publikation.
- a. Formulierung des fertigen Berichtes durch das Steuergremium des Instituts.
 - b. Publikation des vollständigen wissenschaftlichen Berichtes einschließlich der Zusammenfassung von externen Stellungnahmen und Kommentaren.

2.9 *Veröffentlichung von Berichten*

Wesentliche Aufgabe des Institutes ist es, vorhandene Evidenz einer sorgfältigen Prüfung zu unterziehen und die Ergebnisse dieser Prüfung mit dem Ziel der Verbesserung der individuellen und globalen Gesundheitsversorgung nutzbar zu machen.

Das Institut hat das Selbstverständnis, alle Ergebnisse, die im Institut oder im Namen des Institutes erarbeitet wurden, transparent zu machen und zu veröffentlichen.

Zur Wahrung der Unabhängigkeit des Institutes muss ausgeschlossen werden, dass die Auftraggeber oder andere interessierte Dritte Einfluss auf die Inhalte der Berichte nehmen können. Dies könnte zu einer Vermengung der wissenschaftlichen Ergebnisse einerseits und politischen oder interessengesteuerten Aspekten andererseits führen. Gleichzeitig muss vermieden werden, dass das Institut seinerseits bestimmte Ergebnisse zurückhält. Um dies auszuschließen, werden alle Ergebnisse, die das Institut im Rahmen seiner gesetzlichen Aufgaben erarbeitet, einschließlich des Berichtsplans und einer Zusammenfassung der zugrunde liegenden Materialien, die zur Ergebnisfindung beigetragen haben, zeitnah veröffentlicht.

Die Rechte an der Publikation liegen, falls nicht anders vereinbart, beim Institut.

Der Zeitrahmen für die Veröffentlichung ist Bestandteil des Berichtsplans.

2.10 *Struktur der Recherche und Literaturscreening*

2.10.1 *Struktur der themenbezogenen Literaturrecherche*

Die folgenden Ausführungen beziehen sich sowohl auf von Mitarbeitern des Instituts selbst durchgeführte Recherchen als auch auf Recherchen, die von einem externen Experten oder Mitarbeitern einer externen Institution im Rahmen eines durch das Institut erteilten Auftrags durchgeführt werden.

Allgemeine Grundsätze

Eine themenbezogene Literaturrecherche hat zum Ziel, alle für die jeweilige Fragestellung relevanten, d.h. zum Erkenntnisgewinn beitragenden, Publikationen zu identifizieren. Die Methodik der systematischen Literaturrecherche folgt daher dem generellen Grundsatz, dass die jeweilige themenbezogene Recherche mindestens alle die Kriterien erfüllen muss, die nach den Ergebnissen diesbezüglicher Forschung einen maßgeblichen Einfluss auf das Ergebnis, also auf die

Beantwortung der Fragestellung, haben oder haben könnten. Unter Kriterien sind dabei insbesondere zu verstehen

- die Auswahl der Datenquellen (z.B. öffentliche Datenbanken, nicht öffentliche Datenbanken, Handsuche in ausgewählten Fachzeitschriften, Kontaktaufnahme mit Fachexperten / Industrie / Patientenorganisationen etc.),
- die Suchtechnik in Bezug auf die Auswahl des Publikations- bzw. Studientyps (z.B. randomisierte, kontrollierte Studien, Fallbeobachtungen etc.),
- die Suchtechnik in Bezug auf die durch die Fragestellung vorgegebenen medizinischen Kriterien (z.B. Zielpopulation, Art der Intervention, Endpunkte etc.),
- die Suchtechnik in Bezug auf formale Merkmale der Publikation (z.B. Abstract-Publikationen, Sprache etc.),
- die Durchführung der Recherche selber (z.B. von einer Person oder von zwei Personen unabhängig voneinander).

Beispiele hierzu liefern die Publikationen [1-8].

Da die Relevanz dieser Kriterien für verschiedene Fragestellungen unterschiedlich sein kann, wird im Einzelfall geprüft, welche Kriterien zur Anwendung kommen. Liegen dabei für die einzelnen Kriterien in Bezug auf die Fragestellung keine validen Forschungsergebnisse vor, so wird primär geprüft, ob etablierte Strategien internationaler oder nationaler Arbeitsgruppen (z.B. der Cochrane Collaboration) zur Anwendung kommen können, für die bereits weit reichende Erfahrungen vorliegen.

Ablauf

Die Literaturrecherche gliedert sich in die drei Teile

1. Formulierung der Suchstrategie,
2. Anwendung der Suchstrategie (Primärrecherche),
3. Auswahl der relevanten Publikationen aus den Ergebnissen der Primärrecherche.

Die *Formulierung der Suchstrategie* erfolgt auf Basis der mit dem Auftraggeber abgestimmten Fragestellung und der Ergebnisse, die im Ermittlungsverfahren patienten-relevanter Effekte und Größen erlangt wurden (s.a. die Abschnitte 2.7, 2.8 und 3.5). Sie beinhaltet insbesondere

- a) die Auswahl der Datenquellen,
- b) die Auswahl der Suchwörter und -kriterien bei Datenbankrecherchen,

- c) die Festlegung, welche Studien als relevant erachtet werden unter Angabe von Ein- und Ausschlusskriterien.

Dabei wird im Einzelfall geprüft, ob die Hinzuziehung eines oder mehrer Fachexperten sinnvoll ist. Die vorläufig formulierte Suchstrategie ist dem Projektleiter des Instituts darzulegen und mit diesem abzustimmen. Liegt die Suchstrategie abschließend formuliert vor kann im Einzelfall überprüft werden, ob veröffentlichte Recherchen anderer Arbeitsgruppen ganz oder teilweise ausreichend sind für die eigene Fragestellung und auf die Ergebnisse dieser Recherchen zurückgegriffen wird.

Bei der *Anwendung der Suchstrategie* ist insbesondere vorab festzulegen, ob dies durch eine Person erfolgen soll oder durch zwei Personen unabhängig voneinander. In letzterem Fall sind bei Auftreten diskrepanter Ergebnisse diese dem Projektleiter des Instituts darzulegen und gemeinsam mit ihm die Ursachen zu eruieren. Diesbezüglicher Erkenntnisgewinn soll bei zukünftigen Recherchen berücksichtigt werden.

Ebenso ist bei der *Auswahl der relevanten Publikationen* aus den Ergebnissen der Primärrecherche festzulegen, ob dies durch eine Person oder zwei Personen unabhängig voneinander erfolgen soll. Nach Abschluss dieses Arbeitsschrittes ist vor Bewertung der Literatur das Rechercheergebnis dem Projektleiter sowie den von der Fragestellung tangierten Ressortleitern des Instituts vorzulegen, im Einzelfall kann ebenfalls ein Fachexperte hinzugezogen werden. Sofern sich hieraus die Anforderung einer Modifikation und/oder Erweiterung der Literaturrecherche ergibt, ist diese nach den genannten Grundsätzen durchzuführen und das Ergebnis vor Bewertung der Literatur wiederum vorzulegen.

Dokumentation

Alle Arbeitsschritte der Literaturrecherche sind in ihrer Gesamtheit zu dokumentieren. Hierzu gehören insbesondere

- die Suchstrategie einschließlich der Begründung für die Auswahl der Datenquellen und der angewandten Suchtechniken,
- bei Abfrage bibliographischer Datenbanken zusätzlich
 - o ein Recherche-Nachweis, vorzugsweise in Form einer Bildschirmkopie der Recherche-History,
 - o eine Auflistung der durch Anwendung der Suchstrategie erlangten Primärergebnisse (Zitate, Abstracts),

- o eine Auflistung der nach Sichtung der Primärergebnisse als für die Fragestellung relevant erachteten Publikationen,
 - o eine Auflistung der nicht als relevant erachteten aber im Primärergebnis enthaltenen Literaturstellen einschließlich einer Begründung für die Nicht-Berücksichtigung,
- bei persönlicher Kontaktaufnahme mit Produktherstellern, Fachleuten, Fachgesellschaften etc.
 - o Korrespondenznachweise,
- das Datum bzw. der Zeitraum der Recherche.

Die Form der Dokumentation (Papierkopie, Datenbankformat etc.) ist vorab mit dem Projektleiter des Instituts zu klären und sollte sich grundsätzlich an allgemeinen Vorgaben des Instituts orientieren.

Literatur

- [1] Pham B, Platt R, McAuley L, Klassen TP, Moher D. Is there a “best” way to detect and minimize publication bias? An empirical evaluation. *Eval Health Prof* 2001; 24: 109-125.
- [2] McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *Lancet* 2000; 356: 1228-1231.
- [3] Moher D, Pham B, Klassen TP et al. What contributions do languages other than English make on the results of meta-analyses? *J Clin Epidemiol* 2000; 53: 964-972
- [4] Gregoire G, Derderian F, Le Lorier J. Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? *J Clin Epidemiol* 1995; 48: 159-163
- [5] Juni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *Int J Epidemiol* 2002; 31: 115-123.
- [6] Fergusson D, Laupacis A, Salmi LR, McAlister FA, Huet C. What should be included in meta-analyses? An exploration of methodological issues using the ISPO meta-analyses. *Int J Technol Assess Health Care* 2000; 16: 1109-1119.
- [7] Sampson M, Barrowman NJ, Moher D et al. Should meta-analysts search Embase in addition to Medline? *J Clin Epidemiol* 2003; 56: 943-955.
- [8] MacLean CH, Morton SC, Ofman JJ, Roth EA, Shekelle PG; Southern California Evidence-Based Practice Center. How useful are unpublished data from the Food and Drug Administration in meta-analysis? *J Clin Epidemiol* 2003; 56: 44-51.

2.10.2 Literaturscreening

Neben der themenbezogenen, rückschauenden Recherche ist das frühzeitige Erkennen und Bewerten aktueller und relevanter Publikationen notwendig; Grundlage hierfür bildet die

systematische diesbezügliche Beobachtung der wesentlichen Datenquellen. Der Begriff „Datenquellen“ umfasst dabei nicht nur Fachzeitschriften, sondern auch die Laienpresse, Tages-, Wochen- und Monatspresse, elektronische Medien etc.

Als „relevant“ ist in diesem Zusammenhang eine Publikation insbesondere dann einzustufen, wenn

- die Publikation die aktuelle Versorgungslage mutmaßlich erheblich beeinflussen wird,
- die Publikation als „Meilensteinstudie“ zu werten ist,
- in der Publikation ein in der Öffentlichkeit vorrangig präsent Thema behandelt wird, oder
- die Publikation ein Thema der internen Prioritätenliste berührt (s.a. de Abschnitte 2.13 und 5.3).

Das Steuergremium legt fest, welche Datenquellen (Fachzeitschriften, Internetseiten etc.) wesentlich sind und daher regelmäßig beobachtet werden. Die Ressorts werten die Ihnen zugeteilten Datenquellen aus und leiten den jeweiligen anderen Ressorts für sie relevante Publikationen zu. Eine Publikation kann eine ad hoc Bewertung einschließlich einer offiziellen Stellungnahme des Instituts nach sich ziehen. Das diesbezügliche Vorgehen folgt festgelegten Standards (s.a. Abschnitt 2.15).

2.11 Entscheidungsfindung

2.11.1 Zielsetzung

Die Zielsetzung der Beratungen innerhalb des Steuergremiums ist es, ein maximales Maß an Sicherheit durch eine informierte Entscheidung zu erreichen, um daraus u.a. klare Empfehlungen für Dritte ableiten zu können. Wichtig ist es dabei, institutsintern und gegenüber den Auftraggebern sowie der fachlichen wie auch nicht-fachlichen Öffentlichkeit Glaubwürdigkeit aufzubauen und zu festigen. Primär handelt es sich bei der Entscheidungsfindung im Steuergremium um einen der wesentlichen innerorganisatorischen Prozesse.

Dazu müssen Entscheidungsprozesse wie deren Ergebnisse nach Außen und Innen transparent sein und sorgfältig abgewogen werden. Ziel ist es, einen Konsens im Steuergremium herauszuarbeiten, der nach außen als eine fundierte und einheitliche Entscheidung wahrgenommen wird. Dazu ist es u.a. notwendig, dass alle die Entscheidung des Instituts nach außen, aber auch gegenüber Mitarbeitern und Institutsangehörigen, mit verantworten und vertreten können. Durch Abstimmungen und Mehrheitsentscheidungen wird dies in den meisten Fällen nicht gelingen

können. Entscheidungen, die nicht von allen getragen werden, führen zu unterschiedlichen Darstellungen in der Öffentlichkeit und können als Inkompetenz wahrgenommen werden.

Die Verschiedenheit der Aspekte und Perspektiven gilt gleichzeitig aber auch als Human Ressource. Gegensätzliche Standpunkte oder konträre Meinungen müssen im Innenverhältnis geäußert werden (können) und bereichern die Diskussion. Dafür sind gegenseitige Wertschätzung und zielorientiertes Arbeiten eine unabdingbare Voraussetzung. Andernfalls droht die „innere Kündigung“.

Vermieden werden sollten daher insbesondere:

- Alleinentscheidungen Einzelner
- Untergang der Meinungsvielfalt durch erzwungenen Konsens
- Ungleichgewichtung von Einzelmeinungen
- Uninformierte Entscheidungen, die ohne ausreichenden Sachverstand getroffen werden
- Mehrheitsentscheidungen, die ohne adäquate Sachkompetenz getroffen werden
- Ablehnung unpopulärer Meinungen
- Entscheidungen, die die Leitung nicht gegenüber dem Vorstand vertreten kann
- Intransparente Entscheidungsprozesse
- das Verschleppen von Entscheidungen

2.11.2 Methoden der Entscheidungsfindung

Entscheidungen sollten zielorientiert getroffen werden. Die Entscheidungen werden in der Regel von dem zuständigen Ressort bzw. der eingesetzten Projektgruppe vorbereitet und im Steuergremium durch den Leiter der Projektgruppe oder des Ressorts oder einen zu benennenden Vertreter präsentiert. Die Darstellung sollte ausreichende Informationen, insbesondere über pro und contra sowie evtl. Risiken eines Sachverhaltes enthalten, die den Mitgliedern des Steuergremiums eine ausreichende Entscheidungsgrundlage bieten („informed consent“).

Jedes Ressort und die Leitung bekommt ausreichend Gelegenheit zur Kommentierung und zum Austausch verschiedener Standpunkte. Wesentliche Argumente werden im Protokoll der Sitzung festgehalten. Inhaltliche Änderungen werden in der Gruppe beraten und von den zuständigen Leitern (Projektgruppe oder Ressort) in das Dokument eingepflegt.

Der zeitliche Ablauf ist von der Dringlichkeit der Entscheidungen abhängig und macht unter besonderen Umständen beschleunigte Entscheidungsprozesse erforderlich.

2.11.3 Methoden bei Dissens

Kann das Steuergremium aufgrund der vorgelegten Argumente nicht zu einer Entscheidung kommen, die von allen Mitgliedern gegenüber Dritten vertreten werden kann, so geht der Entwurf an die Projektgruppe oder das Ressort zur Überarbeitung zurück. Die Überarbeitung orientiert sich dabei auf eine klar definierte Aufgaben- oder/und Fragestellung aus dem Steuergremium heraus.

Die Arbeitsgruppe kann sich dabei externen Sachverständes z.B. in Form des wissenschaftlichen Beirates bedienen. Der überarbeitete Entwurf ist dem Steuergremium erneut vorzulegen und dort zu diskutieren.

Die Institutsleitung hat bei allen Entscheidungsprozessen ein Vetorecht. Auch in diesem Fall geht der Entwurf zur weiteren Überarbeitung nach den oben ausgeführten Prinzipien in die Projektgruppe oder das Ressort zurück. Bei Abwesenheit der Institutsleitung kann das Vetorecht auf eine durch die Institutsleitung zuvor benannte und autorisierte Person übertragen werden. Besteht die Institutsleitung aus mehr als einer Person regelt sie jeweils im Innenverhältnis die Aus- bzw. Nichtausübung eines Vetos.

2.12 Verteilung der Beweislast bei Fragestellungen

Die Beweislast für einen behaupteten Nutzen von medizinischen Verfahren liegt immer bei denjenigen Institutionen oder Personen, die die Anwendung dieser Verfahren propagieren, sei es aus kommerziellen oder andersartig eigennützigen Motiven heraus oder auch aus Überzeugung. Nutzen soll in diesem Zusammenhang als Überwiegen von patienten-relevanten Vorteilen gegenüber Risiken durch die Anwendung des infrage stehenden medizinischen Verfahrens verstanden werden. Das heißt, die Beweislast für das Vorhandensein eines patienten-relevanten erwünschten Effekts und den Ausschluss von relevanten Risiken liegt bei denjenigen, die das Verfahren propagieren. Somit ist es im Interesse dieser Institutionen und Personen, für eine Bewertung des Verfahrens durch das Institut alle ihnen zur Verfügung stehenden Daten dem Institut ebenfalls zur Verfügung zu stellen. Es ist *nicht* Pflicht des Instituts, einen fehlenden patienten-relevanten Effekt oder das Vorhandensein von relevanten Risiken nachzuweisen. Das Institut *kann* allerdings – soweit die verfügbare Datenbasis dies zulässt – in seinen Bewertungen solche Nachweise (nach dem jeweils gültigen Stand der Wissenschaft) erbringen.

2.13 Priority-Setting

Priority-setting is not an exact process. Nevertheless, it is important to aim for fairness and transparency in the distribution of intellectual and financial resources in a public agency such as the Institute.

Priority-setting mechanisms in research and health technology assessment programs need to be transparent, with agreement within the agency about the rationale, values, methods and criteria being used [1-3]. These should enable the activities and priorities of the agency to be reviewed externally [1].

While much of the work of the Institute will be initiated in response to external requests, the Institute will need to set internal priorities for addressing the report requests. The Institute will also be setting in its own priorities in:

- Some internal work projects in all Departments;
- The production of patient information by the Patient Information Department;
- Projects commissioned from external researchers and agencies;
- Methodological and scientific work conducted by the Institute.

Priority-setting in each individual area may have particular issues that need to be addressed. Each Department may develop specific priority-setting procedures based on the specialised activities of that Department. However, priority-setting processes for the Steering Committee and individual Departments will aim to be explicit and consistent with the general rationale, values and methods set out below.

2.13.1 Rationale and values underlying the Institute's prioritisation

The Institute is concerned that its activities achieve a beneficial impact for the health of the German community and contribute to the development of community and scientific understanding.

In making choices about the work it undertakes, the Institute will consider the:

- Proportion of people likely to benefit from the work it chooses to undertake;
- Burden of disease (including cost), currently and in the future, both for individuals and the community, particularly disadvantaged groups;
- National health priorities;
- Ability for the activity of the Institute to affect practice and outcomes;
- Resources required by the Institute to address the activity effectively;

- Potential for beneficial impact on the community or health care system, and the impact on equity;
- Potential risks;
- Unique contribution the Institute could make, including assessing what others are doing or who else could potentially undertake the activity;
- Contribution that the activity could make to the quality of the Institute's work;
- Potential contribution to the evolution of scientific knowledge.

2.13.2 Processes for priority-setting

While there is a wide variety of relatively unvalidated models and processes in use for priority-setting in similar agencies to the Institute [1-3], they generally have several features in common. These have informed the basic format for the Institute's approach to priority-setting, as follows:

- Gather data and opinions to inform decision making, maintaining documentation;
- Apply the relevant criteria developed by the Institute or the particular Department;
- Supply a report of the data and opinions, together with a recommendation, to the Steering Committee, who will make the final decision, documenting the reason for the decision.

References

- [1] Hensall C et al. Priority setting for health technology assessment. Theoretical considerations and practical approaches. A paper produced for the EUR-ASSESS project. Int J Technol Assess Health Care, 1997; 13: 144-185.
- [2] Townsend J, Buxton M, Harper G. Prioritisation of health technology assessment. The PATHS model: methods and case studies. Health Technology Assessment 2003; Vol 7: No 20. NHS R&D HTA Programme.
- [3] Ghaffar A, de Francisco A, Matlin S. The Combined Approach Matrix: A priority-setting tool for health research. Geneva: Global Forum for Health Research, June 2004.

2.14 Zeitpunkte der Berichtserstellung

Die Bewertung einer Methode bzw. Intervention ist zu jeder Zeit möglich. Generelle Vorgaben dahingehend, dass die Erstellung einer systematischen Bewertung durch das Institut frühestens nach Verstreichen einer gewissen Zeit nach Zulassung oder Etablierung einer Methode oder Intervention erfolgt, werden nicht getroffen. Sofern im Falle einer frühzeitigen Bewertung auf

Grund fehlender Langzeitstudien eine hohe Ergebnisunsicherheit festgestellt wird, wird diese den allgemeinen Arbeitsmethoden folgend dargestellt.

2.15 *Rapid Response*

In besonderen Situationen kann es notwendig werden, auf tagesaktuelle Themen zeitnah und ohne großen zeitlichen Verzug zu reagieren. Dies setzt eine verkürzte Beratungsphase und eine schnelle Entscheidungsfindung innerhalb des Steuergremiums voraus. Darüber hinaus können auch Abstimmungen mit externen Einrichtungen erforderlich werden. Aufgrund des hohen zeitlichen Drucks und der damit verbundenen Gefahr einer ungenauen oder fehlerhaften Aussage in solchen Fällen ist die Orientierung an vorab definierten Prozessabläufen zur Fehlerminimierung notwendig. Entsprechende Verfahrensanweisungen, in welchen die Prozesse, Kommunikationswege und Verantwortlichkeiten festgelegt werden, werden vom Steuergremium erarbeitet. Bis dahin kommen die allgemein definierten Entscheidungswege zur Anwendung.

3. Wissenschaftliche Methoden und Biometrie

3.1 *Darstellung von Effekten und Risiken*

Die Darstellung von Interventions- oder Expositionseffekten ist zunächst eindeutig an eine klare Zielvariable zu knüpfen. Die Betrachtung einer alternativen Zielvariablen verändert auch die Darstellung und die Stärke eines möglichen Effekts. Die Wahl eines geeigneten Effektmaßes hängt grundsätzlich vom Messniveau der betrachteten Zielvariablen ab. Bei stetigen Variablen können Effekte in der Regel mit Hilfe von Mittelwerten sowie Differenzen von Mittelwerten – unter Umständen nach geeigneter Gewichtung – dargestellt werden. Bei kategoriellen Zielvariablen kommen die üblichen Effekt- und Risikomaße aus Vierfeldertafeln in Frage [1]. Nach Festlegung eines primären Effektmaßes zur Datenanalyse sollte zur deskriptiven Darstellung ein Effekt sowohl durch ein absolutes (z.B. absolute Risikoreduktion oder Number Needed to Treat) als auch durch ein relatives Maß (z.B. relatives Risiko oder Odds Ratio) beschrieben werden. Eine übersichtliche Zusammenfassung der Vor- und Nachteile der üblichen Effektmaße gibt Kapitel 8 des Cochrane Reviewer Handbooks [2]. Besondere Aspekte, die beim Vorliegen ordinaler Daten berücksichtigt werden müssen, beschreibt Agresti (1990,1999) [3,4].

Zwingend erforderlich ist es, bei jeder Effektschätzung den Grad der statistischen Unsicherheit zu beschreiben. Häufige Methoden hierfür sind die Berechnung des Standardfehlers sowie eines Konfidenzintervalls. Wann immer möglich soll ein adäquates Konfidenzintervall mit der Information, ob es sich um ein- oder zweiseitige Konfidenzgrenzen handelt sowie des gewählten Konfidenzniveaus, angegeben werden. In der medizinischen Forschung wird üblicherweise das zweiseitige Konfidenzniveau 95% verwendet, in manchen Situationen aber auch 90% oder 99%. Eine Übersicht über die häufigsten Methoden zur Berechnung von Konfidenzintervallen geben Altman et al. (2000) [5].

Je nach Datensituation (z.B. sehr kleine Stichproben) und Fragestellung ist zur Einhaltung des Konfidenzniveaus die Anwendung exakter Methoden bei der Intervallschätzung von Effekten und Risiken in Erwägung zu ziehen. Eine aktuelle Diskussion exakter Methoden gibt Agresti (2003) [6].

Literatur

- [1] Bender R. Interpretation von Effizienzmaßen der Vierfeldertafel für Diagnostik und Behandlung. Med Klin 2001; 96: 116-121.
- [2] Deeks JJ, Higgins JPT, Altman DG. Analysing and presenting results. In: Alderson P, Green A, Higgins JPT, editors. Cochrane Reviewers' Handbook 4.2.2 [updated March 2004]; Section 8. In: The Cochrane Library, Issue 1, 2004. Chichester: Wiley, 2004.

- [3] Agresti A. Categorical Data Analysis. New York: Wiley, 1990.
- [4] Agresti A. Modelling ordered categorical data: Recent advances and future challenges. Stat Med 1999; 18: 2191-2207.
- [5] Altman DG, Machin D, Bryant TM, Gardner MJ, editors. Statistics with Confidence: Confidence Intervals and Statistical Guidelines, 2nd ed. London: BMJ Books, 2000.
- [6] Agresti A. Dealing with discreteness: Making 'exact' confidence intervals for proportions, differences of proportions, and odds ratios more exact. Stat Meth Med Res 2003; 12: 3-21.

3.2 *Beurteilung statistischer Signifikanz*

Mit Hilfe statistischer Signifikanztests ist es möglich, auffällige Ergebnisse aufgrund echter Effekte von reinen Zufallsfunden, die auf normale Datenvariabilität zurückführbar sein können, abzugrenzen. Die Konvention, von einem statistisch signifikanten Resultat zu sprechen, falls $p < 0,05$ gilt, macht in vielen Fällen durchaus Sinn. Dennoch sollte man nicht stur auf der Forderung $p < 0,05$ beharren. Je nach Fragestellung und Aussage kann es erforderlich sein einen möglicherweise sehr viel kleineren p -Wert zu fordern. Umgekehrt gibt es Situationen, in denen man auch mit einem höheren Signifikanzniveau arbeiten kann.

Bei der Interpretation von p -Werten ist eine Reihe von Aspekten zu betrachten. Es muss im Detail klar sein, auf welche Fragestellung und welche Datensituation sich die Signifikanzangabe bezieht und wie genau die statistische Hypothesenformulierung lautet. Insbesondere muss klar sein, ob es sich um eine ein- oder zweiseitige Hypothese handelt [1] und ob die betrachtete Hypothese als Teil eines Mehrhypothesenproblems zu sehen ist [2]. Beide Aspekte, ob eine Hypothese ein- oder zweiseitig zu formulieren ist und ob für multiples Testen adjustiert werden muss, werden in der wissenschaftlichen Literatur immer wieder kontrovers diskutiert.

Bezüglich der Hypothesenformulierung ist traditionell von einem zweiseitigen Testproblem auszugehen. Ausnahmen hiervon bilden z.B. Nicht-Unterlegenheitsstudien (s.a. Abschnitt 3.21). Die Formulierung eines einseitigen Hypothesenproblems ist grundsätzlich immer möglich, bedarf jedoch einer genauen Begründung. Im Fall einer einseitigen Hypothesenformulierung sind die Anwendung einseitiger Signifikanztests sowie die Berechnung einseitiger Konfidenzgrenzen angebracht. Für eine bessere Vergleichbarkeit mit zweiseitigen statistischen Verfahren wird in einigen Guidelines für klinische Studien eine Halbierung des üblichen Signifikanzniveaus von 5% auf 2,5% gefordert [3]. Ein zentral wichtiges Grundprinzip ist in jedem Fall, dass die Hypothesenformulierung (ein- oder zweiseitig) und das Signifikanzniveau a priori klar festzulegen sind.

Ist eine betrachtete Hypothese eindeutig Teil eines Mehrhypothesenproblems, so ist eine adäquate Adjustierung für multiples Testen erforderlich. Eine Übersicht, in welchen Situationen dies der Fall ist, und welche Methoden hierfür zur Verfügung stehen, geben Bender & Lange (2001) [4].

Ein nicht signifikantes Ergebnis darf nicht per se als Evidenz für das Nicht-Vorhandensein eines Effekts gewertet werden [5]. Zum Nachweis einer Gleichheit sind Methoden für Äquivalenzhypothesen erforderlich (s.a. Abschnitt 3.21).

Literatur

- [1] Bland JM, Altman DG. One and two sided tests of significance. BMJ 1994; 309: 248.
- [2] Tukey JW. Some thoughts on clinical trials, especially problems of multiplicity. Sc 1977; 198: 679-684.
- [3] ICH E9 Expert Working Group. ICH Harmonised Tripartite Guideline: Statistical Principles for Clinical Trials. Stat Med 1999; 18: 1905-1942.
- [4] Bender R, Lange S. Adjusting for multiple testing - when and how? J Clin Epidemiol 2001; 54: 343-349.
- [5] Altman DG, Bland JM. Absence of evidence is not evidence of absence. BMJ 1995; 311: 485.

3.3 *Beurteilung klinischer Relevanz*

Die klinische Relevanz eines Effekts oder Risikos ist grundsätzlich nicht am p -Wert ablesbar. Die statistische Signifikanz ist eine Wahrscheinlichkeitsaussage, in die neben der Stärke eines möglichen Effekts auch die Variabilität der Daten und die Höhe des Stichprobenumfangs eingeht. So ist ein sehr kleiner p -Wert bei einem riesigen Stichprobenumfang anders zu beurteilen, als in einer kleinen Studie [1]. In einer kleinen Studie ist ein sehr kleiner p -Wert nur bei einem deutlichen Effekt zu erwarten, während in einer großen Studie auch bei verschwindend kleinem Effekt eine deutliche Signifikanz nicht ungewöhnlich ist.

Eine formale Beurteilung der klinischen Relevanz von Studienergebnissen ist ein noch weitgehend ungelöstes methodisches Problem. Nur in wenigen Leitlinien finden sich Hinweise auf die Definition von relevanten bzw. irrelevanten Unterschieden zwischen Gruppen. Ein erster Ansatz zur Beurteilung der klinischen Relevanz von Studienergebnissen ist die Bewertung des Effektschätzers und des dazugehörigen Konfidenzintervalls (s.a. Abschnitt 3.1) mit Hilfe medizinischer Sachkenntnis. Ein formales Relevanzkriterium kann die Beurteilung der (im Falle von günstigen Effekten) unteren Konfidenzgrenze für den Effektschätzer bzw. die Anwendung eines statistischen Tests mit Verschiebung der Nullhypothese zum Nachweis relevanter klinischer

Effekte sein. Eine weitere Möglichkeit liegt darin, ein Relevanzkriterium individuell zu formulieren, z.B. im Sinne einer Responderdefinition. Darüber hinaus spielt die individuelle Einschätzung der Betroffenen eine wesentliche Rolle. Anhaltspunkte hierzu wird im Einzelfall die Darlegung patienten-relevanter Endpunkte liefern (s.a. Abschnitt 3.5).

Literatur

- [1] Royall RM. The effect of sample size on the meaning of significance tests. Am Stat 1986; 40: 313-315.

3.4 Subgruppen-Analysen

Subgruppenanalyse werden in der methodischen Literatur sehr kritisch diskutiert [1,2]. Die Interpretation ihrer Ergebnisse wird im Wesentlichen durch drei Aspekte erschwert:

- Den post-hoc-Charakter, es sei denn Subgruppenanalysen wurden a priori geplant, sind wohl fundiert und sind Bestandteil des Studienprotokolls (ggf. auch von Amendments).
- Im Falle eines statistisch signifikanten Ergebnisses durch das Problem des multiplen Testens (s.a. Abschnitt 3.2), es sei denn, dieses wurde durch Anpassung des Signifikanzniveaus berücksichtigt.
- Im Falle eines statistisch nicht signifikanten Ergebnisses durch die in aller Regel geringe Power, die zumeist auch für das (inferenzstatistische) Aufdecken moderater Unterschiede in aller Regel zu gering ist, es sei denn, es wurde eine Fallzahlplanung mit angemessener Power und ggf. entsprechend erhöhtem Stichprobenumfang zusätzlich für eine Subgruppenanalyse vorgenommen [3].

Die Ergebnisse von Subgruppen-Analysen sollten bei Vorliegen eines der drei oben angesprochenen Aspekte unter starkem Vorbehalt in Bewertungen einfließen und nicht das Ergebnis der primären Analyse dominieren; dies gilt um so mehr, falls das primäre Studienziel nicht erreicht wurde.

Darüber hinaus sind Subgruppen-Analysen nicht interpretierbar, wenn das Subgruppen-bildende Merkmal nach Therapiebeginn (nach Randomisierung) erhoben wurde (bzw. erhoben werden kann), z.B. sogenannte „Responder-Analysen“.

Die statistische Analyse von Subgruppen-Effekten sollte anhand eines Interaktionstests vorgenommen werden. Der Befund, dass in einer Subgruppe ein statistisch signifikanter Effekt

beobachtet wurde, in einer anderen aber nicht, kann (inferenzstatistisch) nicht als das Vorliegen eines Subgruppen-Effekts interpretiert werden.

Ungeachtet der oben getroffenen Einschränkungen, können für manche Fragestellungen Subgruppen-Analysen die auf absehbare Zeit beste wissenschaftliche Evidenz für die Bewertung von Effekten in eben diesen Subgruppen darstellen [4], da z.B. ethische Überlegungen gegen eine Reproduktion der entsprechenden Ergebnisse in einer Validierungsstudie sprechen können.

Literatur

- [1] Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med 1992; 116: 78-84.
- [2] Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet 2000; 355: 1064-1069.
- [3] Brookes ST, Whitely E, Egger M, Davey Smith G, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: Risks of subgroup-specific analyses; power and sample size for the interaction test. J Clin Epidemiol 2004; 57: 229-236.
- [4] Freemantle N. Interpreting the results of secondary end points and subgroup analyses in clinical trials: should we lock the crazy aunt in the attic? BMJ 2001; 322: 989-991.

3.5 *Ermittlung patienten-relevanter Effekte und Größen*

Klinische Studien werden in der Regel so konzipiert, dass krankheitsbezogene Effekte gemessen werden. Diese Effekte können für Patienten Bedeutung haben oder auch nicht. Darüber hinaus kann das Ausmaß des Nutzens oder des Risikos von Patienten individuell unterschiedlich eingeschätzt werden. Beispielsweise können Menschen, die eine Erkrankung oder Verletzung (erlebt) haben, und diejenigen, die dies nicht persönlich erlebt haben, Nutzen und Risiken unterschiedlich empfinden und bewerten [1].

Viele Studien messen Surrogate als vermutete Indikatoren für ein patienten-relevantes Ergebnis. Der direkte Rückschluss von einer Veränderung einer Surrogatvariablen auf die Veränderung eines patienten-relevanten Ereignisses ist jedoch nicht ohne weiteres möglich [2-4]. Aus diesem Grund wird das Institut derartige Surrogate nur unter der Voraussetzung der hinreichend sicheren Prüfung der Kausalkette mit patienten-relevanten Effekten für die Bewertung eines medizinischen Verfahrens heranziehen.

Über klinisch relevante Ereignisse wie Mortalität, Morbidität und unerwünschte Wirkungen der medizinischen Intervention hinaus können auch andere Effekte oder Ereignisse für Patienten Bedeutung haben, wie z.B.

- Krankheitsbezogene Lebensqualität, einschließlich der Auswirkung auf die Alltagsaktivitäten und das Wohlbefinden,
- Annehmlichkeiten, Umstände und Aufwand der Behandlung,
- Behandlungszufriedenheit und –präferenzen,
- Auswirkungen auf Angehörige.

Ergebnisse, die für die Patienten wichtig sind, sollten berücksichtigt werden, bevor bewertende Schlussfolgerungen über Interventionen getroffen werden. Daten solcher Ergebnisse können die Aussagen einer systematischen Übersicht wesentlich ändern [5]. Zusätzlich muss berücksichtigt werden, ob Ergebnisse auf Patienten-Untergruppen (z.B. Geschlechtsunterschiede, verschiedene Altersgruppen) übertragbar sind, und ob Ergebnisse aus langfristigen Untersuchungen vorliegen.

Sofern ableitbar, sollte dargestellt werden, ob die Größenordnung des Effekts einer Intervention für Patienten relevant ist [5,6]. Diese Daten werden wenn möglich an Hand der Literaturrecherche inkl. der Auswertung der Ergebnisse qualitativer Forschung und Patienten- und Angehörigenbefragung (z.B. mittels Fokusgruppenbefragung) erlangt.

Die externe Validität von vielen patienten-relevanten Zielgrößen kann von kulturellen und sozialen Gegebenheiten beeinflusst werden und kann sich im Verlauf der Zeit ändern. Darüber hinaus kann die Einschätzung einer Intervention schon alleine dadurch beeinflusst werden, dass es sich um eine Neuentwicklung handelt.

Literatur

- [1] Slevin ML, Stubbs L, Plant HJ et al. Attitudes to chemotherapy: comparing views of patients with cancer with those of doctors, nurses, and general public. BMJ 1990; 300: 1458-1460.
- [2] Gøtzsche PC et al. Beware of surrogate outcome measures. Int J Tech Assess Health Care 1996; 12: 238-246.
- [3] Liberati A, Sheldon TA, Banta HD et al. Eur-Assess project subgroup report on methodology: Methodological guidance for the conduct of health technology assessment. Int J Tech Assess Health Care 1997; 13: 186-219.
- [4] The CAST investigators. Preliminary report: Effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. New Engl J Med 1989; 321: 406-412.
- [5] Dixon-Woods M et al. Integrative approaches to qualitative and quantitative evidence. London: NHS Health Development Agency; 2004.
- [6] Egger M, Davey Smith G, Phillips AN. Meta-analysis principles and procedures. BMJ 1997; 315: 1533-1537.

3.6 Aspekte der Beurteilung der Studienqualität

Bei der Beurteilung der allgemeinen Qualität von Studien spielt eine Reihe von Aspekten eine Rolle, die hier nicht vollständig im Detail wiedergegeben werden können. Grundsätzlich soll in einer Studie von der Planung über die Durchführung bis zur Auswertung und Präsentation nach einem anerkannten standardisierten Konzept vorgegangen werden. Hierzu gehört ein Studienprotokoll, in dem alle wichtigen Methoden und Vorgehensweisen beschrieben sind. Die üblichen Standards hierbei bilden bei randomisierten klinischen Versuchen die Grundprinzipien der „*Good Clinical Practice*“ (GCP) [1,2] und bei epidemiologischen Studien die Leitlinien und Empfehlungen zur Sicherung von „*Guter Epidemiologischer Praxis*“ (GEP) [3]. Ein zentrales Qualitätskriterium bei der Beurteilung von Studienergebnissen ist dabei, ob die Studie auch so ausgewertet wurde wie in der Planung vorgesehen.

Weitere wichtige Quellen, die grundlegende Aspekte zur Beurteilung der Studienqualität (bzw. der Qualität der entsprechenden Publikation) beschreiben, sind das CONSORT Statement für randomisierte klinische Versuche [4] und das dazugehörige Erläuterungsdokument [5], das CONSORT Statement für Cluster-randomisierte Studien [6], das QUOROM Statement für Meta-Analysen randomisierter Studien [7], das TREND Statement für nicht-randomisierte Interventionsstudien [8], das STARD-Statement für Diagnosestudien [9] und das dazugehörige Erläuterungsdokument [10] sowie verschiedene Standardwerke aus der Fachliteratur [11-13].

Zentrale Aspekte zur Beurteilung randomisierter klinischer Studien stellen ein adäquates Concealment, d.h. die Unvorhersehbarkeit und Verheimlichung der Gruppenzuteilung (z.B. durch externe Randomisierung bei nicht verblindbaren Studien), die verblindete Zielgrößenerhebung bei nicht verblindbaren Studien (siehe Abschnitt 3.16), die adäquate Anwendung des "Intention-to-treat"-Prinzips sowie die Festlegung auf ein eindeutiges Hauptzielkriterium bzw. die angemessene Berücksichtigung eines möglichen Multiplizitätsproblems (siehe Abschnitt 3.2) dar.

Die Beurteilung formaler Kriterien liefert wesentliche Anhaltspunkte für die Qualität von Studien. Es ist jedoch immer notwendig, eine über rein formale Aspekte hinaus gehende Beurteilung vorzunehmen, um beispielsweise Fehler, Widersprüche und Ungereimtheiten in Publikationen darzustellen und deren Relevanz für die Ergebnisinterpretation zu prüfen.

Literatur

- [1] Kolman J, Meng P, Scott G. *Good Clinical Practice. Standard Operating Procedures for Clinical Researchers*. Chichester: Wiley, 1998.
- [2] ICH Steering Committee. Official web site for the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). <http://www.ich.org/>. Zugriff am 20. Oktober 2004.

- [3] Arbeitsgruppe Epidemiologische Methoden der Deutschen Arbeitsgemeinschaft für Epidemiologie (DAE). Leitlinien und Empfehlungen zur Sicherung von Guter Epidemiologischer Praxis (GEP). <http://medweb.uni-muenster.de/institute/epi/dae/Empfehlungen.doc>. Zugriff am 21. Oktober 2004.
- [4] Moher D, Schulz KF, Altman DG, for the CONSORT Group. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med* 2001; 134: 657-662.
- [5] Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne DR et al. The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Ann Intern Med* 2001; 134: 663-694.
- [6] Campbell MJ, Elbourne DR, Altman DG for the CONSORT Group. CONSORT statement: Extension to cluster randomised trials. *Br Med J* 2004; 328: 702-708.
- [7] Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF et al. Improving the quality of reports of meta-analyses of randomised controlled trials: The QUOROM statement. *Lancet* 2000; 354: 1896-1900.
- [8] Des Jarlais DC, Lyles C, Crepaz N for the TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: The TREND statement. *Am J Public Health* 2004; 94: 361-366.
- [9] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med* 2003; 138: 40-44.
- [10] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM et al. The STARD statement for reporting studies of diagnostic accuracy: Explanation and elaboration. *Ann Intern Med* 2003; 138: W1-12.
- [11] Egger M, Davey Smith G, Altman DG, editors. *Systematic Reviews in Health Care: Meta-Analysis in Context*. London: BMJ Books, 2001.
- [12] Alderson P, Green S, Higgins JPT, editors. *Cochrane Reviewers' Handbook 4.2.2* (updated March 2004). In: *The Cochrane Library*, Issue 1, 2004. Chichester: Wiley, 2004.
- [13] Guyatt G, Rennie D, editors. *Users' Guide to the Medical Literature*. Chicago, IL: AMA Press, 2002.

3.7 *Ermittlung von unerwünschten Wirkungen*

Die Anwendung einer medizinischen Intervention gleich welcher Art (medikamentös, nicht-medikamentös, chirurgisch, diagnostisch, präventiv etc.) birgt das Risiko unerwünschter Wirkungen in sich. „Unerwünschte Wirkungen“ bezeichnet dabei alle diejenigen Ereignisse und Wirkungen, die einen individuell wahrgenommenen oder objektiv feststellbaren physischen oder psychischen Schaden darstellen, der zu einer mehr oder weniger starken, kurz- oder langfristigen Beeinträchtigung der Lebensqualität und/oder Verkürzung der Lebenserwartung führt oder führen kann.

Die Darstellung potenzieller Risiken einer medizinischen Intervention ist wesentlicher und gleichberechtigter Bestandteil der Berichtserstellung zu dieser Intervention. Sie gewährleistet eine informierte populationsbezogene, aber auch individuelle Nutzen-Risiko-Abwägung [1]. Voraussetzung dafür ist, dass analog zur Angabe der Effektstärke einer medizinischen Intervention auch Wahrscheinlichkeiten für das Eintreten der potenziellen unerwünschten Wirkungen aus den vorliegenden Daten ermittelt und angegeben werden.

Die Darstellung des Risikoprofils einer medizinischen Intervention ist jedoch in den allermeisten Fällen ungleich schwerer als die des Nutzens derselbigen [2]. Im Gegensatz dazu, dass durch eine medizinische Intervention in der Regel nur wenige patienten-relevante (Wirksamkeits-)Endpunkte positiv beeinflusst werden, ist das Auftreten unerwünschter Wirkungen vielschichtig. Darüber hinaus treten in klinischen Studien schwerwiegende Ereignisse (z.B. Tod, Organversagen) in der Regel eher selten auf. Schließlich ist festzustellen, dass die zielgerichtete Forschung zur Erfassung insbesondere der seltenen, aber schwerwiegenden unerwünschten Wirkungen einschließlich der Darstellung eines kausalen Zusammenhangs mit der medizinischen Intervention erheblich unterrepräsentiert ist [3-5]. Gründe hierfür liegen unter anderem in einer auf diesem Gebiet im Rahmen von klinischen Studien nicht ausreichend entwickelten Methodik und in einem hohen Aufwand an Ressourcen, um eine mehr als irrelevante Erhöhung von seltenen unerwünschten Ereignissen ausschließen zu können. Darüber hinaus verhindern möglicherweise verschiedene Interessenskonflikte eine ausgewogene Darstellung der Risiken.

Folge der oben genannten Hindernisse ist, dass in vielen Fällen trotz enormer Anstrengungen die Aussagenunsicherheit bei der Darstellung unerwünschter Wirkungen höher sein wird als bei der Darstellung positiver Effekte [6]. Notwendig ist, hier eine sinnvolle Balance zwischen Vollständigkeit der Aufarbeitung und Ressourcenaufwand zu finden. Daraus folgend wird es notwendig sein, die Aufarbeitung und Darstellung auf relevante unerwünschte Wirkungen zu beschränken. Als relevant können dabei insbesondere diejenigen unerwünschten Wirkungen bezeichnet werden, die

- dem ersten Anschein nach den Nutzen der Intervention fast oder vollständig oder mehr als aufheben könnten,
- dem ersten Anschein nach sich zwischen zwei oder mehr ansonsten gleichwertigen Therapieoptionen erheblich unterscheiden,
- dem ersten Anschein nach insbesondere bei denjenigen von mehreren Therapieoptionen auftreten, die besonders effektiv sind,
- dem ersten Anschein nach eine Dosis-Wirkungs-Beziehung haben,
- von Patienten als besonders wichtig angesehen werden (s.a. Abschnitt 3.5),

- von schwerwiegender Morbidität oder gar erhöhter Mortalität begleitet sind oder in Folge mit erheblicher Beeinträchtigung der Lebensqualität einhergehen.

Im Interesse der Patientensicherheit wird das Institut bei der Aufarbeitung und Darstellung unerwünschter Wirkungen folgende Grundsätze beachten.

- Grundlage der Auswahl relevanter unerwünschter Wirkungen nach den oben genannten Kriterien bildet die Zusammenstellung derjenigen unerwünschten Wirkungen und Ereignisse, die im Zusammenhang mit der Durchführung der betreffenden medizinischen Intervention gehäuft gesehen oder berichtet wurden. Die Zusammenstellung erfolgt insbesondere auf Grundlage der Daten kontrollierter Interventionsstudien, in denen zielgerichtet der Nutzen der Intervention untersucht wurde, sowie auf Grundlage epidemiologischer (z.B. Kohorten- oder Fall-Kontroll-Studien) und unkontrollierter Untersuchungen (z.B. aus Anwendungsbeobachtungen, Material von Pharmakovigilanzzentren etc.) sowie Tierexperimenten und Experimenten zur Überprüfung eines pathophysiologischen Konstrukts.
- Liegen aus den oben genannten Datenquellen begründete Verdachtsmomente für das Vorliegen einer unerwünschten Wirkung vor, so wird diese solange als möglich angesehen, bis dies durch die Ergebnisse zielgerichteter Forschung mit hinreichender Sicherheit ausgeschlossen werden kann (Beweislastumkehr). Dies gilt insbesondere für schwerwiegende unerwünschte Wirkungen. Die Evidenzhierarchie entspricht dabei derjenigen für Therapiestudien (s.a. die Abschnitte 3.10 und 3.11). Unter „hinreichend“ ist zu verstehen, dass
 - o die entsprechende(n) Untersuchung(en) in ihrem Design und ihrer Planung primär auf den Nachweis der Gleichwertigkeit zwischen der medizinischen Intervention und anderen Therapieoptionen oder Placebo bzw. keiner Intervention gerichtet waren und
 - o die (statistische) Definition der Gleichwertigkeit mit derjenigen vergleichbar ist, die vor Beginn des Bewertungsverfahrens durch das Institut z.B. durch Patientenbefragung ermittelt wurde (s.a. Abschnitt 3.5).^b

^b Bei der (statistischen) Definition der Gleichwertigkeit ist zu beachten, dass die Weite des Bereichs, der als Gleichwertigkeit akzeptiert wird, in der Regel mit der Durchführbarkeit der entsprechenden Studien korreliert. Z.B. wäre die Forderung, dass Gleichwertigkeit nur bei exakter Übereinstimmung der Ereignisraten akzeptiert wird, nicht umsetzbar.

Literatur

- [1] Ziegler DK, Mosier MC, Buenaver M, Okuyemi K. How much information about adverse effects of medication do patients want from physicians? Arch Intern Med 2001; 161: 706-713.
- [2] Derry S, Loke YK, Aronson JK. Incomplete evidence: the inadequacy of databases in tracing published adverse drug reactions in clinical trials. BMC Medical Research Methodology 2001; 1: 7.
- [3] Edwards JE, McQuay HJ, Moore RA, Collins SL. Reporting of adverse effects in clinical trials should be improved: lessons from acute postoperative pain. J Pain Symptom Manage 1999; 18: 427-437.
- [4] Ioannidis JPA, Lau J. Completeness of safety reporting in randomized trials – an evaluation of seven medical areas. JAMA 2001; 285: 437-443.
- [5] Bonhoeffer J, Zumbunn B, Heininger U. Reporting of vaccine safety data in publications: systematic review. Pharmacoepidemiol Drug Saf 2004 Jun 16 [Epub ahead of print].
- [6] Loke YK, Price D, Herxheimer A, on behalf of the Cochrane Adverse Effects Subgroup. Including adverse effects in your review. <http://www.dsru.org/wwwboard/latestdraft.pdf>. Zugriff am 6. Oktober 2004.

3.8 Beurteilung von Studien mit „überholten“ Methoden

Tritt der Fall ein, dass bei einer Projektarbeit des Instituts ältere Studien beurteilt werden müssen, die den heute gültigen Qualitätsansprüchen (s.a. Abschnitt 3.9) nicht genügen, weil sie in einer Zeit geplant und durchgeführt wurden, in denen diese Ansprüche noch nicht existierten, so wird dies bei der Beurteilung der Studien berücksichtigt. Es werden die Nachteile und Mängel der Studien aufgezeigt und mögliche Konsequenzen diskutiert, diese Mängel führen aber nicht automatisch zum Ausschluss der Studien aus der Bewertung.

3.9 Beurteilung verschiedener Studienarten

Es werden hier nur die wichtigsten Designs zusammengefasst, die bei der Beurteilung von Studien in der medizinischen Forschung eine Rolle spielen. Zunächst kann man unterscheiden zwischen Beobachtungsstudien und Interventionsstudien. Innerhalb der Beobachtungsstudien gibt es häufig den ersten Informationsgewinn aus Fallberichten oder Fall-Serien. Diese sind natürlich anfällig gegen Verzerrungen aller Art, so dass hier, je nach Fragestellung, nur sehr beschränkt eine Evidenz ableitbar ist (s.a. Abschnitt 3.10). Aus bevölkerungsbezogenen Querschnittsstudien lassen sich z.B. Krankheitsprävalenzen schätzen, aber keine Zusammenhänge zwischen veränderbaren Expositionen und Krankheiten ableiten. Weitere wichtige Studientypen der Epidemiologie sind

Fall-Kontroll-Studien [1], in denen retrospektiv nach Expositionen von Fällen und Kontrollen gefragt wird und Kohortenstudien [2], in denen bestimmte Gruppen (Kohorten) über einen Zeitverlauf hinweg beobachtet werden. Kohortenstudien sind in diesem Sinne prospektiv angelegt, allerdings gibt es auch retrospektive Kohortenstudien, in denen die Exposition aus der Vergangenheit erfasst wird (häufig in der Arbeits- oder auch Pharmakoepidemiologie anzutreffen). Grundsätzlich sind prospektive Studien retrospektiven Designs vorzuziehen. Allerdings sind z.B. Fall-Kontroll-Studien häufig die einzige praktikable Möglichkeit, Informationen über Zusammenhänge zwischen Expositionen und seltenen Erkrankungen zu gewinnen.

Im Bereich der Interventionsstudien ist zunächst eine Kontrollgruppe zu fordern. Aus einem reinen Vorher-Nachher-Vergleich in einem Design mit abhängigen Stichproben ohne Kontrollgruppe lässt sich in der Regel nicht der Effekt einer Intervention ableiten. Ausnahmen bilden Krankheitsbilder mit einem deterministischen (oder nahezu deterministischen) Verlauf (z.B. diabetisches keto-azedotisches Koma). Gütekriterien, welche die Aussagekraft kontrollierter Studien erhöhen, sind Randomisierung und Verblindung (s.a. die Abschnitte 3.6 und 3.19). Häufige Designs bei klinischen Studien bilden parallele Gruppen [3], Cross-Over-Studien [4] und Cluster-randomisierte Studien [5]. Sind in einer Studie Zwischenauswertungen geplant, so muss dies mit Hilfe eines adäquaten sequentiellen Designs berücksichtigt werden [6].

Diagnose- und Screening-Studien können sehr unterschiedliche Ziele haben, so dass hiervon die Wahl eines adäquaten Designs abhängt (s.a. die Abschnitte 4.2 und 4.3).

Zur Untersuchung von genetischen Faktoren, die die Entstehung und Verteilung von Krankheiten verursachen können, ist in den letzten Jahren die relativ neue Disziplin der genetischen Epidemiologie entstanden [7]. In diesem Bereich gibt es im Rahmen von Kopplungs- und Assoziationsstudien eine Reihe neuer spezieller Studiendesigns, auf die hier nicht weiter eingegangen werden kann.

Literatur

- [1] Breslow NE, Day NE. Statistical Methods in Cancer Research Vol. I: The Analysis of Case-Control Studies. Lyon: Int. Agency for Res. on Cancer, 1980.
- [2] Breslow NE, Day NE. Statistical Methods in Cancer Research Vol. II: The Design and Analysis of Cohort Studies. Lyon: Int. Agency for Res. on Cancer, 1987.
- [3] Pocock SJ. Clinical Trials: A Practical Approach. Chichester: Wiley, 1983.
- [4] Jones B, Kenward MG. Design and Analysis of Cross-Over Trials. London: Chapman & Hall, 1989.
- [5] Donner A, Klar J. Design and Analysis of Cluster Randomization Trials in Health Research. London: Arnold, 2000.
- [6] Whitehead J. The Design and Analysis of Sequential Clinical Trials. Chichester: Ellis Horwood, 1983.

- [7] Khoury MJ, Beaty TH, Cohen BH. Fundamentals of Genetic Epidemiology. New York: Oxford University Press, 1993.

3.10 Rangordnung verschiedener Studienarten / Evidenzgrade

Es gibt verschiedene Ansätze, bestimmten Studientypen bestimmte Evidenzgrade zuzuordnen, um so eine Rangfolge bezüglich der Evidenz verschiedener Studien im Rahmen einer systematischen Übersicht oder für die Entwicklung von Leitlinien zu erstellen [1,2]. Es gibt jedoch kein System, das allgemein akzeptiert und universell in allen systematischen Übersichten einsetzbar ist [3]. In Anlehnung an das System der Scottish Intercollegiate Guidelines Network Grading Review Group [2] lässt sich jedoch eine grobe Hierarchie von Studientypen erstellen, die weitgehend akzeptiert wird. Hiernach wird z.B. für Therapiestudien die höchste Evidenzstufe systematischen Übersichten randomisierter kontrollierter Studien zugeordnet. Auf dem nächsten Platz erscheinen randomisierte kontrollierte Studien. Diese werden in einigen Einteilungen noch in solche hoher oder weniger hoher Qualität eingestuft. Danach folgen nicht-randomisierte Interventionsstudien, prospektive Beobachtungsstudien, retrospektive Beobachtungsstudien, nicht-experimentelle Studien (Fallserien und Fallberichte) und mit niedrigster Evidenzstufe Expertenmeinungen ohne wissenschaftliche Begründung. Dieses grobe System ist der jeweiligen Situation und Fragestellung anzupassen und detaillierter darzustellen.

Literatur

- [1] Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ et al. Users' guides to the medical literature. IX. A method for grading health care recommendations. JAMA 1995; 274: 1800-1804.
- [2] Harbour R, Miller J, for the Scottish Intercollegiate Guidelines Network Grading Review Group. A new system for grading recommendations in evidence based guidelines. BMJ 2001; 323: 334-336.
- [3] Katrak P, Bialocerkowski AE, Massy-Westropp N, Kumar VSS, Grimme KA. A systematic review of the content of critical appraisal tools. BMC Med Res Meth 2004; 4: 22.

3.11 Bezug zwischen Studientyp / -art und Fragestellung

Der randomisierte kontrollierte Versuch gilt als Studientyp höchster Qualität. Allerdings muss dies in Bezug zur Fragestellung gesehen werden. Nur beim Wirksamkeitsnachweis einer Intervention ist in der Regel eine randomisierte kontrollierte Studie das zu fordernde Design. Bei anderen Fragestellungen kommen alternative Studiendesigns in Frage. Die wichtigsten Situationen hierbei

sind die folgenden. Zur Untersuchung diagnostischer Methoden genügt in vielen Fällen eine Querschnittsstudie. Das optimale Design zur Erforschung prognostischer Faktoren ist eine prospektive Kohortenstudie. Zur Untersuchung des Zusammenhangs zwischen Expositionen und sehr seltenen Erkrankungen kommen Fall-Kontroll-Studien in Frage. Sollen allerdings diagnostische Verfahren oder Prognose-Faktoren zusammen mit den aus dem Informationsgewinn resultierenden Konsequenzen (z.B. Einleitung einer Therapie) als Strategie geprüft werden, sind wiederum randomisierte Studien das Design der Wahl (siehe Abschnitte 4.2, 4.3 und 4.9).

3.12 Beurteilung nicht oder nicht vollständig publizierter Daten

In der Praxis ergibt sich häufig das Problem, dass zur Beurteilung von Publikationen notwendige Daten oder Informationen ganz oder teilweise fehlen. Das erste Ziel des Instituts ist es in solchen Fällen zumindest zu versuchen diese Informationen zu beschaffen, z.B. durch den Kontakt mit den Autoren. Ist es nicht möglich, die notwendigen Daten zu bekommen, so werden bei teilweise vorhandenen Informationen Sensitivitätsanalysen im Sinne von Best-Case- und Worst-Case-Szenarien durchgeführt, dargestellt und diskutiert. Beim vollständigen Fehlen relevanter Informationen, kann eine Publikation nicht beurteilt werden und es wird lediglich dargestellt, dass es zu einem bestimmten Thema weitere Daten gibt, die aber zur Beurteilung nicht zur Verfügung stehen.

3.13 Umgang mit Rohdaten

Im Rahmen der wissenschaftlichen Bewertung medizinischer Leistungen ist eine der Hauptaufgaben des Instituts das Sammeln und die Analyse bereits publizierter Daten aus systematischen Literaturrecherchen. Es ist jedoch nicht ausgeschlossen, dass das Institut für bestimmte Fragestellungen auch bisher noch nicht analysierte Rohdaten auswertet, die von externen Quellen zur Verfügung gestellt werden (z.B. Krankenkassen). Eine sinnvolle Analyse solcher Daten setzt voraus, dass die Rahmenbedingungen klar sind, unter denen diese Daten erhoben wurden und die Plausibilität und Qualität der Daten überprüfbar ist. Insbesondere ist darauf zu achten, dass wesentliche Qualitätskriterien eingehalten werden. Z.B. sollten die Daten bei Therapiestudien unter Einhaltung des GCP-Standards entstanden sein (s.a. Abschnitt 3.6). Darüber hinaus wird in den meisten Fällen die Vorlage eines Studienprotokolls zur adäquaten Beurteilung notwendig sein. Im Umgang mit Rohdaten werden datenschutzrechtliche Aspekte berücksichtigt (s.a. Abschnitt 3.17).

3.14 Objektivität der Studiendurchführung und -evaluation

Die wichtigsten Kriterien, mit denen die Objektivität der Durchführung von Studien beurteilt werden kann, sind bei Interventionsstudien das Einhalten formaler Gütekriterien wie Randomisierung, Verblindung und das Intention-to-Treat-Prinzip (s.a. Abschnitt 3.6). Eine Verletzung dieser Kriterien kann zu einer systematischen Verzerrung (Bias) im Ergebnis führen, da möglicherweise Entscheidungen oder Beurteilungen nicht mehr objektiv erfolgen (s.a. Abschnitt 3.19). Die Stärke eines möglichen Bias ist aber in Bezug zur Höhe eines geschätzten Effekts zu beurteilen. So würde bei einem sehr hohen Effekt mit schmalem Konfidenzintervall und einem zu erwartenden geringen Bias der wahre Effekt zwar möglicherweise verringert, aber die Richtung des Effekts würde sich nicht umdrehen. Solche Überlegungen sind im Einzelfall darzulegen und zu diskutieren.

3.15 Beurteilung von Ergebnisunsicherheiten

Grundsätzlich ist jedes Ergebnis einer empirischen Studie unsicher. Die statistische Unsicherheit, die bedingt ist durch die begrenzte Stichprobe, lässt sich in Form von Standardfehler und Konfidenzintervall quantifizieren und beurteilen (s.a. Abschnitt 3.1). Es darf jedoch nicht übersehen werden, dass diese Berechnungen unter der Annahme durchgeführt werden, dass das gewählte statistische Modell richtig ist und dass keine anderweitigen systematischen Fehler und Verzerrungen vorliegen. Die Unsicherheiten, die dadurch entstehen, dass die tatsächlichen Gegebenheiten mehr oder weniger stark vom gewählten statistischen Modell abweichen, bleiben hier unberücksichtigt [1]. Es gibt zwar formale Ansätze, auch diese allgemeinen Modellunsicherheiten zu berücksichtigen, die entsprechenden Verfahren sind aber noch nicht ausreichend untersucht bzw. ausgereift [2]. Zu fordern ist in jedem Fall eine qualitative Beurteilung der allgemeinen Ergebnisunsicherheit auf der Basis der aktuellen Literatur zum jeweiligen Thema. Nach wie vor sind die klassischen Kausalitätskriterien von Hill (1965) [3] hier eine Hilfe.

Literatur

- [1] Chatfield C. Model uncertainty, data mining and statistical inference (with discussion). J R Stat Soc A 1995; 158: 419-466.
- [2] Phillips CV. Quantifying and reporting uncertainty from systematic errors. Epidemiology 2003; 14: 459-466.

- [3] Hill AB. The environment and disease: Association or causation? Proc R Soc Med 1965; 58: 295-300.

3.16 Beurteilung nicht verblindbarer Verfahren

Zur Vermeidung systematischer Verzerrungen (Bias) bei der Effektschätzung (s.a. Abschnitt 3.19) werden kontrollierte Studien wenn möglich randomisiert und doppel-blind durchgeführt. In einigen Situationen ist jedoch eine Verblindung der Intervention nicht möglich. Auch nicht verblindete Studien können interpretierbare Ergebnisse liefern. Wichtige Instrumente zur Vermeidung von Bias sind hier wiederum Randomisierung und eine geeignete Wahl der Zielvariablen. In nicht verblindbaren Studien ist es zentral wichtig, dass ein adäquates Concealment gewährleistet wird (s.a. Abschnitt 3.6), und dass als Zielvariable ein „harter“ objektiver Endpunkt gewählt wird, oder zumindest dass dafür Sorge getragen wird, dass die Zielvariable unabhängig vom (unverblindeten) Behandler ist bzw. unabhängig vom Behandler verblindet erhoben wird (verblindete Zielgrößenerhebung).

3.17 Berücksichtigung datenschutzrechtlicher Aspekte / Vertraulichkeit

Die Verarbeitung personenbezogener oder personenbeziehbarer Daten findet innerhalb des Institutes nach Maßgaben der jeweiligen Bundesdatenschutzgesetzgebung statt. Für deren Einhaltung ist der vom Institut bestellte Datenschutzbeauftragte zuständig. Die infrastrukturellen Rahmenbedingungen für die Einhaltung der Datensicherheit ergeben sich aus den Ausführungen im Kapitel EDV (s.o.). Grundsätzlich umfasst die Verarbeitung datenschutzrelevanter Daten in erster Linie die Personalverwaltung des Institutes.

Darüber hinaus ist es zukünftig denkbar, dass das Institut auch im Rahmen von Forschungsarbeiten personenbeziehbare sowie in Ausnahmefällen personenbezogene Daten verarbeitet. Soweit diese ursächlich an dritter Stelle erfasst wurden oder werden, sind entsprechende Erklärungen über die Einhaltung der gesetzlichen Bestimmungen nach außen zu geben, andererseits ist vorab jeweils im Einzelfall sorgfältig zu prüfen, inwieweit die notwendigen gesetzlichen Voraussetzungen vorliegen (Einverständniserklärungen, Aufklärung, etc.).

Darüber hinaus sollte angestrebt werden, primär personenbezogene Daten in anonymisierter oder pseudonymisierter Form durch dritte anzunehmen und weiterzuverarbeiten, was in den meisten Fällen für die Forschungszwecke und einzelnen wissenschaftlichen Fragestellungen ausreichend sein dürfte. Insbesondere können damit evt. Vorbehalte zur Weitergabe von Daten an das Institut ausgeräumt werden.

Hinsichtlich der vertraulichen Behandlung von Unternehmensdaten sind ggf. geeignete Erklärungen gegenüber Dritten abzugeben, die eine Wahrung der Vertraulichkeit nach außen gewährleisten. Neben der notwendigen technischen Infrastruktur zur Wahrung der Datensicherheit finden sich in allen Arbeitsverträgen, die das Institut mit seinen Mitarbeitern abschließt, entsprechende Passagen, die die Mitarbeiter zur Vertraulichkeit verpflichten. Entsprechende Verpflichtungen müssen im Einzelfall auch extern beauftragte Personen oder Einrichtungen gegenüber dem Institut abgeben.

3.18 *Berücksichtigung ethischer Aspekte*

Das Institut wird getragen von dem primären Ziel, mit qualitativ hochwertiger Arbeit die Gesundheitsversorgung der Menschen in Deutschland zu verbessern. Dabei stehen für das Institut die Maximierung des globalen, aber auch des individuellen Patientennutzens sowie die Stärkung der Patientenautonomie durch Gesundheitsbildung und Information im Vordergrund. Die Methode der evidenzbasierten Medizin wird hierbei als notwendiges und hilfreiches Werkzeug verstanden. Das Institut wird dieses Werkzeug gewissenhaft anwenden und seine Grenzen berücksichtigen.

Das Institut ist sich darüber hinaus seiner Position im deutschen Gesundheitswesen und insbesondere seiner Verantwortung gegenüber den Personen und Institutionen, die Gesundheitsleistungen in Anspruch nehmen, durchführen, finanzieren oder entwickeln, bewusst. Auch wenn durch den Gesetzgeber eine strikte Trennung zwischen wissenschaftlicher Bewertung einerseits und Entscheidung für oder gegen die Aufnahme einer medizinischen Intervention in die Regelversorgung andererseits vorgesehen wurde, ist sich das Institut darüber im klaren, dass die Ergebnisse der Institutsarbeit mittel- oder unmittelbar Einfluss auf die medizinische Versorgung nehmen werden. Daraus folgend ist die Berücksichtigung der möglichen oder wahrscheinlichen Folgen, die die Berichterstattung des Institutes für einzelne Personen, Bevölkerungs- oder Berufsgruppen sowie Institutionen oder Unternehmen hat oder haben wird, elementarer Bestandteil der Institutsarbeit. Die frühzeitige Einbeziehung einzelner Vertreter der von der jeweiligen Institutsarbeit tangierten Gruppen und Institutionen wird hierbei unterstützend sein. Das Institut sieht sich dabei primär verantwortlich für die Belange aller Bürgerinnen und Bürger. Das Institut wird sich mit aller Kraft Einflüssen jedweder Herkunft widersetzen, die die Arbeit des Instituts weg von Transparenz und Unabhängigkeit hin zu von Interessen gesteuerten Aussagen lenken wollen.

Das Institut wird sich Fragen der Verteilungsgerechtigkeit nicht verschließen. Begrenzte Ressourcen führen dazu, dass die Wahlfreiheit auf der einen Seite mit einer Einschränkung derselbigen auf der anderen Seite einhergeht. Das Institut wird den Beteiligten im

Gesundheitswesen vermitteln, dass die Entscheidung für oder gegen eine medizinische Methode einem gewissenhaften Abwägungsprozesses des Für und Wider entspringen muss. In diesem Zusammenhang sieht sich das Institut insbesondere auch für Minderheiten und benachteiligte Bevölkerungsgruppen in der Verantwortung stehend.

Für das Institut stehen ethische Belange auch bei eigenen Forschungen im Vordergrund. Auch im Rahmen einer Berichtserstellung ist bei Forschungsaktivitäten die Abwägung des Für und Wider solcher Vorhaben für die Betroffenen sowie ggf. die Involvierung eines Ethikkomitees bei der Planung und Durchführung erforderlich.

3.19 Darstellung und Beschreibung von Biasarten

Unter Bias versteht man systematische Verzerrungen bei der Schätzung von Effekten aus Studiendaten. Es gibt eine Vielzahl von möglichen Ursachen, die einen Bias hervorrufen können [1]. Nur die wichtigsten Biasarten werden im Folgenden beschrieben; eine ausführliche Übersicht über verschiedene Biasarten in unterschiedlichsten Situationen gibt Feinstein (1985) [2].

Ein *Selection Bias* entsteht durch eine verzerrte Zuordnung der Probanden in zu vergleichende Gruppen, so dass systematische Unterschiede zwischen den Gruppen vorhanden sind. Sind dadurch wichtige Confounder in den Gruppen ungleich verteilt, so kommt es zum Selection Bias. Die Randomisierung ist die beste Methode zur Vermeidung eines Selection Bias, da die entstehenden Gruppen sich sowohl bezüglich bekannter als auch unbekannter Confounder nicht systematisch unterscheiden. In nicht randomisierten Studien kann der Effekt bekannter Confounder mit Hilfe multifaktorieller Methoden berücksichtigt werden (s.a. Abschnitt 3.23), es bleibt aber das Risiko eines systematischen Unterschieds zwischen den Gruppen durch nicht bekannte Confounder.

Neben der Vergleichbarkeit der Gruppen bezüglich potentieller prognostischer Faktoren spielen die Behandlungsgleichheit (mit Ausnahme der zu untersuchenden Intervention) und die Beobachtungsgleichheit aller Probanden eine entscheidende Rolle. Eine systematische Verzerrung durch unterschiedliche Behandlungen wird als *Performance Bias* bezeichnet. Eine Verletzung der Beobachtungsgleichheit kann zu einem *Detection Bias* führen. Die Verblindung ist ein wirksamer Schutz vor beiden Biasarten. In der Epidemiologie spricht man in diesem Zusammenhang von *Information Bias*.

Protokollverletzungen und Studienaustritte können das Studienergebnis systematisch verzerren, was als *Attrition Bias* bezeichnet wird. Zur Vermeidung von Attrition Bias kann das Intention to Treat Prinzip eingesetzt werden, das besagt, dass alle randomisierten Probanden in der Analyse

berücksichtigt werden und zwar in der durch die Randomisierung zugeordneten Gruppe, unabhängig von Protokollverletzungen.

In Diagnosestudien muss die Untersuchung des diagnostischen Tests in einem angemessenen Spektrum von Patienten erfolgen. Unterscheidet sich das Untersuchungskollektiv systematisch von der Patientenpopulation, in der der Test zur Anwendung kommen soll, so kann dies zum *Spectrum Bias* führen (s.a. Abschnitt 4.2). Um diesen Bias zu vermeiden, muss der diagnostische Test in einem repräsentativen Patientenkollektiv untersucht werden.

Bei der Untersuchung von Screening-Programmen besteht die Gefahr, dass durch Vorverlagerung der Diagnose eine nur scheinbare Verlängerung der Überlebenszeit beobachtet wird, die jedoch in Wahrheit auf nicht vergleichbare Anfangszeitpunkte zurückzuführen ist (*Lead Time Bias*). Eine verlängerte Überlebenszeit kann ebenso bei Krankheiten mit langer präklinischer Phase vorgetäuscht werden (*Length Bias*). Einen Schutz vor diesen Verzerrungsmechanismen bietet die Durchführung einer randomisierten Studie zur Beurteilung der Wirksamkeit einer Screening-Maßnahme (s.a. Abschnitt 4.3).

Ein allgemeines Problem bei der Schätzung von Effekten ist eine Verzerrung durch *Messfehler* und *Missklassifikationen* in den erhobenen Studiendaten [3,4]. In der Praxis sind Messfehler kaum zu vermeiden und es ist bekannt, dass auch nicht-differentielle Messfehler zu einem Bias in der Effektschätzung führen können. Im Fall eines einfachen linearen Regressionsmodells mit zufälligem klassischen Messfehler in der erklärenden Variable kommt es zum *Dilution Bias*, d.h. zu einer Verzerrung des Schätzers in Richtung des Nulleffekts. In anderen Modellen und komplexeren Situationen sind jedoch Verzerrungen in alle Richtungen möglich. Daher sollte die Stärke möglicher Messfehler in allen Studien diskutiert werden, und bei Bedarf sollten Methoden zur Adjustierung von Messfehlerverzerrungen angewendet werden.

Ein ähnliches Problem stellen fehlende Werte dar. Fehlende Werte, die nicht durch einen Zufallsmechanismus zustande kommen, können ebenso einen Bias im Ergebnis verursachen [5]. Die möglichen Ursachen und Auswirkungen fehlender Werte sind daher im Einzelfall zu diskutieren und bei Bedarf sind statistische Verfahren erforderlich, die einen Bias berücksichtigen bzw. ausgleichen.

Im Rahmen von systematischen Übersichten spielt noch der *Publication Bias* eine wichtige Rolle [6]. Da häufig signifikante Ergebnisse eher publiziert werden als nicht signifikante, kommt es bei der Zusammenfassung der publizierten Ergebnisse zu einer systematischen Verzerrung des gemeinsamen Effektschätzers. Zur Entdeckung und Berücksichtigung von Publication Bias können grafische Methoden wie der Funnel Plot und/oder statistische Methoden wie die Meta-Regression eingesetzt werden [7].

Literatur

- [1] Sackett DL. Bias in analytic research. J Chron Dis 1979; 32: 51-63.
- [2] Feinstein AR. Clinical Epidemiology. The Architecture of Clinical Research. Philadelphia: WB Saunders Co., 1985.
- [3] Carroll RJ, Ruppert D, Stefanski LA. Measurement Error in Nonlinear Models. London: Chapman & Hall, 1995.
- [4] Cheng C-L, van Ness JW. Statistical Regression with Measurement Error. London: Arnold, 1999.
- [5] Little RJA, Rubin DB. Statistical Analysis with Missing Data. New York: Wiley, 1987.
- [6] Begg CB, Berlin JA. Publication bias: A problem in interpreting medical data (with discussion). J R Stat Soc A 1988; 151: 419-463.
- [7] Sterne JAC, Egger M, Davey Smith G. Investigating and dealing with publication and other biases in meta-analysis. BMJ 2001; 323: 101-105.

3.20 *Nachweis der Verschiedenheit*

Beim empirischen Nachweis, dass sich gewisse Gruppen bezüglich eines bestimmten Merkmals unterscheiden, sind verschiedene Aspekte zu beachten. Zunächst ist festzuhalten, dass der Ausdruck „Nachweis“ nicht als „Beweis“ im mathematischen Sinne zu verstehen ist. Mit Hilfe empirischer Studiendaten lassen sich Aussagen nur unter Zulassung gewisser Irrtumswahrscheinlichkeiten treffen. Durch Anwendung statistischer Methoden lassen sich diese Irrtumswahrscheinlichkeiten jedoch gezielt kontrollieren und minimieren, um auf diese Weise eine Aussage „statistisch nachzuweisen“. Die übliche Methode eines solchen statistischen Nachweises in der medizinischen Forschung ist die Anwendung von Signifikanztests (s.a. Abschnitt 3.2). Diese Argumentationsebene ist zu trennen von der Beurteilung der klinischen Relevanz eines Unterschieds (s.a. Abschnitt 3.3). Die Kombination beider Argumente liefert in der Praxis eine adäquate Beschreibung eines Unterschieds aufgrund empirischer Daten.

Bei der Anwendung eines Signifikanztests zum Nachweis eines Unterschieds sind a priori die wissenschaftliche Fragestellung und darauf basierend die Zielvariable, das Effektmaß und die statistische Hypothesenformulierung festzulegen. Damit eine Studie groß genug ist, um einen Unterschied entdecken zu können, ist es notwendig den benötigten Stichprobenumfang vor Beginn der Studie zu berechnen. Dazu sind in einfachen Situationen neben obigen Informationen zusätzlich die Angabe des klinisch relevanten Unterschieds sowie ein Schätzwert der Variabilität der Zielgröße erforderlich. Bei komplexeren Designs oder Fragestellungen sind zusätzlich weitere Angaben wie Korrelationsstruktur, Rekrutierungsschema, Drop-Out-Schema usw. erforderlich [1,2].

Zur abschließenden Darstellung der Ergebnisse gehört neben der Signifikanzangabe einer Aussage ein adäquates Konfidenzintervall für das gewählte Effektmaß (s.a. Abschnitt 3.1), die deskriptive Angabe weiterer Effektmaße um unterschiedliche Aspekte des Ergebnisses darzulegen und eine Diskussion der klinischen Relevanz der Ergebnisse, die auf der Ermittlung patienten-relevanter Zielkriterien aufbauen sollte (s.a. die Abschnitte 3.3 und 3.5).

Literatur

- [1] Desu MM, Raghavarao D. Sample Size Methodology. Boston: Academic Press, 1990.
- [2] Bock J, Toutenburg H. Sample size determination in clinical research. In: Rao CR, Chakraborty R, editors. Handbook of Statistics Vol. 8. Amsterdam: Elsevier, 1991: 515-538.

3.21 Nachweis der Gleichheit

Einer der häufigen schwerwiegenden Fehler bei der Interpretation medizinischer Daten ist es, ein nicht signifikantes Ergebnis eines traditionellen Signifikanztests als Evidenz für die Richtigkeit der Null-Hypothese zu werten [1]. Zum Nachweis einer „Gleichheit“ ist die Anwendung von Methoden zur Untersuchung von Äquivalenzhypothesen erforderlich [2]. Wichtig ist hierbei zunächst das Verständnis, dass eine exakte „Gleichheit“, also z.B. dass die Differenz der Mittelwerte zweier Gruppen exakt Null beträgt, mit Hilfe statistischer Methoden nicht möglich ist. In der Praxis ist aber auch nicht der Nachweis einer exakten Gleichheit gefordert, sondern der Nachweis eines höchstens irrelevanten Unterschieds zwischen zwei Gruppen. Um dieses Ziel zu erreichen muss logischerweise zunächst definiert werden, was ein irrelevanter Unterschied ist, d.h. es ist die Festlegung eines Äquivalenzbereichs erforderlich.

Um sinnvoll auf Äquivalenz schließen zu können, muss – ähnlich wie beim Nachweis eines Unterschieds (s.a. Abschnitt 3.20) – a priori die wissenschaftliche Fragestellung und darauf basierend die Zielvariable, das Effektmaß und die statistische Hypothesenformulierung festgelegt werden. Zusätzlich ist bei Äquivalenzstudien der Äquivalenzbereich klar zu definieren. Dies kann entweder zweiseitig, was zu einem Äquivalenzintervall führt, oder einseitig im Sinne einer „höchstens irrelevanten Unterlegenheit“ erfolgen. Im letzteren Fall spricht man von einer „Nicht-Unterlegenheits-Hypothese“ [3].

Wie bei Überlegenheitsstudien ist es auch in Äquivalenzstudien erforderlich, den benötigten Stichprobenumfang vor Beginn der Studie zu berechnen. Welche Methode hierfür in Frage kommt, hängt von der genauen Hypothese und von der gewählten Analyse-methode ab [4].

Zur Datenanalyse in Äquivalenzstudien müssen speziell für diesen Zweck entwickelte Methoden verwendet werden. Ein häufig angewendetes Verfahren ist die Konfidenzintervall-Methode. Liegt das berechnete Konfidenzintervall völlig im definierten Konfidenzintervall, so gilt dies als Nachweis der Äquivalenz. Um das Niveau $\alpha=0,05$ einzuhalten, genügt hierbei die Berechnung eines 90%-Konfidenzintervalls [5].

Im Vergleich zu Überlegenheitsstudien besitzen Äquivalenzstudien spezielle methodische Probleme. Zum einen ist es häufig schwierig, Äquivalenzbereiche sinnvoll zu definieren [6]. Zum anderen schützen z.B. die üblichen Designkriterien Randomisierung und Doppelblindheit nicht mehr hinreichend sicher vor Verzerrungen [7]. Auch ohne Kenntnis der Therapiegruppe ist es z.B. möglich, die Differenz der Behandlungsunterschiede zur Null und damit zur gewünschten Alternativhypothese hin zu verschieben. Somit ist bei der Beurteilung von Äquivalenzstudien besondere Vorsicht geboten.

Literatur

- [1] Altman DG, Bland JM. Absence of evidence is not evidence of absence. BMJ 1995; 311: 485.
- [2] Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. BMJ 1996; 313: 36-39.
- [3] D'Agostino RB, Massaro JM, Sullivan KM. Non-inferiority trials: Design concepts and issues - the encounters of academic consultants in statistics. Stat Med 2003; 22: 169-186.
- [4] Roebruck P, Elze M, Hauschke D, Leverkus F, Kieser M. Literaturübersicht zur Fallzahlplanung für Äquivalenzprobleme. Inform Biom Epidemiol Med Biol 1997; 28: 51-63.
- [5] ICH E9 Expert Working Group. ICH Harmonised Tripartite Guideline: Statistical Principles for Clinical Trials. Stat Med 1999; 18: 1905-1942.
- [6] Lange S, Freitag G. Choice of delta: requirements and reality - results of a systematic review. Biometrical J 2004; 46 (in press).
- [7] Senn S. Inherent difficulties with active control equivalence studies. Stat Med 1993; 12: 2367-2375.

3.22 Meta-Analysen

Unter Meta-Analyse wird die statistische Zusammenfassung von Ergebnissen mehrerer Studien im Rahmen einer systematischen Übersicht verstanden (s.a. Abschnitt 4.6). Bevor eine Meta-Analyse durchgeführt wird, muss zunächst überlegt werden, ob die Zusammenfassung der betrachteten Studien überhaupt sinnvoll ist. Zum einen muss die Vergleichbarkeit der Studien bezüglich der Fragestellung gegeben sein, zum anderen ist die Heterogenität der Studien bezüglich der

Ergebnisse zu untersuchen [1]. Für diesen Zweck gibt es spezielle neuere statistische Methoden, wie z.B. das I^2 -Maß [2,3]. Ist die Heterogenität der Studien zu groß, so ist eine statistische Zusammenfassung der Studienergebnisse unter Umständen nicht sinnvoll [4]. In diesem Zusammenhang spielt auch die Wahl des Effektmaßes eine Rolle. Es kann sein, dass die Wahl eines bestimmten Maßes zu großer Studienheterogenität führt, ein anderes Maß jedoch nicht. Bei binären Daten sind relative Effektmaße häufig stabiler als absolute, da sie nicht so stark vom Basis-Risiko abhängen [5]. In solchen Fällen, sollte die Datenanalyse über ein relatives Effektmaß erfolgen, zur deskriptiven Darstellung aber absolute Maße hieraus abgeleitet werden.

Literatur

- [1] Glasziou PP, Sanders SL. Investigating causes of heterogeneity in sytematic reviews. Stat Med 2002; 21: 1503-1511.
- [2] Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002; 21: 1539-1558.
- [3] Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003; 327: 557-560.
- [4] Deeks JJ, Higgins JPT, Altman DG. Analysing and presenting results. In: Alderson P, Green A, Higgins JPT, editors. Cochrane Reviewers' Handbook 4.2.2 [updated March 2004]; Section 8. In: The Cochrane Library, Issue 1, 2004. Chichester: Wiley, 2004.
- [5] Furakawa TA, Guyatt GH, Griffith LE. Can we individualize the 'number needed to treat'? An empirical study of summary effect measures in meta-analyses. Int J Epidemiol 2002; 31: 72-76.

3.23 Adjustierung und multifaktorielle Verfahren

Vor allem in nicht randomisierten Studien spielen multifaktorielle Verfahren, die es ermöglichen, den Effekt von Confoundern auszugleichen, eine zentrale Rolle (s.a. Abschnitt 3.19) [1]. Ein weiteres wichtiges Anwendungsgebiet sind Studien mit mehreren Behandlungen [2]. Die Darstellung von Ergebnissen multifaktorieller Methoden ist in der medizinischen Fachliteratur leider oft unzureichend [3,4]. Zu fordern ist eine Darstellung der wesentlichen Aspekte im Rahmen der statistischen Modellbildung [5,6] sowie Angaben zur Güte des gewählten Modells (Goodness-of-Fit) [7]. Es gibt verschiedene Übersichten und Guidelines über die Anforderungen zum Umgang mit multifaktoriellen Verfahren in der Literatur [1,8,9]

Literatur

- [1] Katz MH. Multivariable analysis: A primer for readers of medical research. N Engl J Med 2003; 138: 644-650.

- [2] McAlister A, Straus SE, Sackett DL, Altman DG. Analysis and reporting of factorial trials: A systematic review. JAMA 2003; 289: 2545-2553.
- [3] Bender R, Grouven U. Logistic regression models used in medical research are poorly presented (Letter). BMJ 1996; 313: 628.
- [4] Müllner M, Matthews H, Altman DG. Reporting on statistical methods to adjust for confounding: A cross-sectional survey. Ann Intern Med 2002; 136: 122-126.
- [5] Harrell FEJr, Lee KL, Mark DB. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996; 15: 361-387.
- [6] Royston P. A strategy for modelling the effect of a continuous covariate in medicine and epidemiology. Stat Med 2000; 19: 1831-1847.
- [7] Hosmer DW, Lemeshow S. The importance of assessing the fit of logistic regression models: A case study. Am J Public Health 1991; 81: 1630-1635.
- [8] Bender R, Grouven U. Ordinal logistic regression in medical research. J R Coll Physic London 1997; 31: 546-551.
- [9] Bagley SC, White H, Golomb BA. Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. J Clin Epidemiol 2001; 54: 979-985.

3.24 Social science and qualitative research

3.24.1 Review of qualitative studies

Qualitative research methods are used to explore and understand behaviour, personal experiences and the social world [1-4]. In most cases, qualitative research will be used for hypothesis generation and to assist the interpretation of data. Data collection, analysis and description are in words. Quantitative research primarily works with numbers, but often also includes analysis of personal and social experiences and responses. The key questions for the Institute will be whether a study's methodology was appropriate for the questions being addressed, the quality of the study's conduct and the reliability of its results.

The Institute will apply the same standards to all the types of research it uses. Every type of study will be assessed according to the scientific rigour appropriate for that type of study. There is not as much agreement on the validity of criteria for the conduct, assessment and synthesis of qualitative research studies as there is for other types of research [1-5]. The Institute will use explicit criteria for its assessment of qualitative research studies, adapting these as the evidence base for the use of qualitative research in systematic reviews and health technology strengthens [1].

3.24.2 Use of qualitative research methods and consultation techniques

The Institute will use a variety of qualitative research and consultation techniques in its work. However these will of necessity usually be limited to rapid appraisal methods (such as focus groups). This will be preceded by a search for relevant qualitative data. The results of rapid appraisal methods will be interpreted in the context of results of relevant and more rigorous qualitative research whenever it is available.

The Institute will use rapid appraisal methods appropriate to the question and purpose being addressed. The Institute's guidelines will draw from existing guidelines for the choice, conduct and analysis of these techniques. This will include ethical guidelines where techniques are used that have the potential to cause harm to participants, such as focus groups.

Rapid appraisal methods have varying strengths and weaknesses, and vary greatly in their representativeness and validity. Those that may be used by the Institute include:

- Key informant interviews [6]
- Focus groups [1,7,8]
- Group interviews, group meetings and consultations [9-11]
- Surveys and polling (including online polling and feedback mechanisms).

The Institute may also at times use consensus techniques such as Delphi techniques [12], and participatory evaluation [13].

The Institute may also develop Health Impact Assessments [14]. This technique uses both qualitative and quantitative methods and is based on explicit and transparent values, including the potential impact of decisions on equity and social justice.

As the techniques for gaining people's views vary greatly in their representativeness, the Institute will need to take care to ensure that the views of disadvantaged groups are adequately represented in its processes.

3.24.3 Peer review methods

Gaining expert feedback (including from consumers and patients) will be a key strategy for improving the quality of the Institute's reports and other information. Evidence on the standard practices of peer review in medical journals [15,16], including review by consumers and patients [17], shows little strong evidence of benefit. There are also potential drawbacks of peer review [18]. However, the evidence base is relatively small, although growing. The Institute will need to monitor its practices and new evidence about improving its methods. Departments may use a

variety of methods of rapid appraisal techniques (section 3.24.2), alongside the method traditionally used by journals (individual assessment of manuscripts).

There is not enough evidence to guide decisions about specific processes for formal peer review by individuals. This includes how many peer reviews to seek, and whether or not peer review is open or closed (identity of the peer reviewer not disclosed). Therefore it will be up to each Department to consider the number of peer reviewers it will have for each report. In the absence of evidence about the effects of open or closed peer review, the Institute's peer review process will be open in keeping with its overall commitment to open-ness and accountability.

The Institute will develop a form on key items for reviewers to complete (including such issues as their interests and how many reports they are able to review), and maintain a database of these data [19]. Any publisher with a high level of use of peer reviewers needs to develop a system for editorial rating of the quality of peer reviewers' contributions [20]. The Institute will develop a scoring system based on validated systems [20].

All peer reviewers will be sent the copies of the other peer review reports as well as the final published report.

3.24.4 Community and population level

As part of its approach to helping build scientific literacy in the German community, the Department on Patient Information and Research may also use and adapt some of the techniques [20] of consultation, community decision making and consensus development to serve a dual purpose of education and open community participation. These projects will be conducted jointly between the Departments of Patient Information and Public Relations, involving the Institute leadership and the Steering Committee members. These will use the Institute's web site, and may involve additional multi-media strategies. Key goals will be to increase popular understanding of the concepts of science in health and evidence-based health care and engage community interest in the work of the Institute.

Population level techniques have most widely been used in resource allocation decisions in health [20,21], but several techniques could be adapted by the Institute for the purpose of education and to open its processes to the wider community. These techniques could include:

- Online surveys and polling;
- Citizens' juries [20];
- Public hypothetical debates.

Citizens' juries have been found to be particularly effective when complex subject matter needs to be considered.

References

- [1] Dixon-Woods M et al. Integrative approaches to qualitative and quantitative evidence. London: NHS Health Development Agency; 2004.
- [2] Murphy E et al. Qualitative research methods in health technology assessment: a review of the literature. NHS R&D HTA Programme. Health Technology Assessment, 1998; Vol 2: No 16.
- [3] McClelland S. Qualitative research methods: Their role in health services research. NHS Management Briefing, National Electronic Library for Health. 00&04, February 2001. <http://libraries.nelh.nhs.uk/healthmanagement>. Zugriff am 22. Oktober 2004.
- [4] Harden A et al. Applying systematic review methods to studies of people's views: an example from public health research. J Epidemiol Community Health 2004; 58: 794-800.
- [5] Thomas J et al. Integrating qualitative research with trials in systematic reviews. BMJ 2004; 328: 1010-1012.
- [6] USAID Center for Development Information and Evaluation. Conducting key informant interviews. Washington: USAID Center for Development Information and Evaluation; 1996 No 2. http://www.usaid.gov/pubs/usaid_eval/pdf_docs/pnabs541.pdf. Zugriff am 22. Oktober 2004.
- [7] Aylward P. Conducting research with focus groups. Staff Development Session. Flinders University of South Australia. <http://www.flinders.edu.au/staffdev/courses/research/resources/Focusgroups.pdf>. Zugriff am 22. Oktober 2004.
- [8] USAID Center for Development Information and Evaluation. Conducting focus group interviews. Washington: USAID Center for Development Information and Evaluation; 1996 No 10. http://www.usaid.gov/pubs/usaid_eval/pdf_docs/pnaby233.pdf. Zugriff am 22. Oktober 2004.
- [9] National Resource Centre for Consumer Participation in Health. Feedback, participation and diversity: A literature review. Canberra: Commonwealth of Australia, 2000. <http://www.participateinhealth.org.au>. Zugriff am 22. Oktober 2004.
- [10] National Health and Medical Research Council, Consumers' Health Forum of Australia. Statement on consumer and community participation in research. Canberra: Commonwealth of Australia, 2002. <http://www.nhmrc.gov.au/publications/pdf/r22.pdf>. Zugriff am 22. Oktober 2004.
- [11] National Resource Centre for Consumer Participation in Health. Methods and models of consumer participation. Melbourne: National Resource Centre for Consumer Participation in Health, 2004. <http://www.participateinhealth.org.au>. Zugriff am 22. Oktober 2004.
- [12] Liberati A, Sheldon TA, Banta HD et al. Eur-Assess project subgroup report on methodology: Methodological guidance for the conduct of health technology assessment. Int J Tech Assess Health Care, 1997; 13: 186-219.
- [13] USAID Center for Development Information and Evaluation. Conducting a participatory evaluation. Washington: USAID Center for Development Information and Evaluation; 1996 No 1. http://www.usaid.gov/pubs/usaid_eval/pdf_docs/pnabs539.pdf. Zugriff am 22. Oktober 2004.
- [14] Barnes R, Scott-Samuel A. Health Impact Assessment: A ten minute guide. Liverpool: International Health Impact Assessment Consortium, 2000. <http://www.ihia.org.uk/hiaguide.html>. Zugriff am 25. Oktober 2004.

- [15] Fletcher RH, Fletcher SW. The effectiveness of journal peer review. In Godlee F, Jefferson T (eds). Peer Review in Health Sciences (2nd ed). London: BMJ Books, 2003; 248-262.
- [16] Jefferson TO, Alderson P, Davidoff F, Wager E. Editorial peer-review for improving the quality of reports of biomedical studies. The Cochrane Database of Methodology Reviews 2001, Issue 3.
- [17] Bastian H. Non-peer review: consumer involvement in research review. In Godlee F, Jefferson T (eds). Peer Review in Health Sciences (2nd ed). London: BMJ Books, 2003; 248-262.
- [18] Rennie D. Editorial peer review: Its development and rationale. In Godlee F, Jefferson T (eds). Peer Review in Health Sciences (2nd ed). London: BMJ Books, 2003; 248-262.
- [19] Smith J. How to set up a peer review system. In Godlee F, Jefferson T (eds). Peer Review in Health Sciences (2nd ed). London: BMJ Books, 2003; 151-163.
- [20] Ryan M et al. Eliciting public preferences for healthcare: A systematic review of techniques. Health Technology Assessment, 2001; Vol 5 No 5.
- [21] Hicks N, Harford J. Summary report on consumer participation in resource allocation. Melbourne: National Resource Centre for Consumer Participation in Health, 2000.

4. Spezielle Beurteilung medizinischer und versorgungsrelevanter Inhalte

Die Extraktion von Daten aus den bei der Literaturrecherche als relevant angesehenen Publikationen (s.a. Abschnitt 2.10) und deren Bewertung wird strukturiert erfolgen und dokumentiert werden. Zur Dokumentation können noch zu entwickelnde Extraktionsbögen Verwendung finden. Falls diese Extraktionsbögen nicht verwendet werden sollen, muss dies begründet und mit dem Projektleiter des Instituts abgestimmt sowie ein alternativer Extraktionsbogen entwickelt, abgestimmt und angewendet werden.

Extraktionsbögen u.a. für

- HTA-Berichte
- Systematische Übersichten
- Interventionsstudien
- Diagnosestudien zur Evaluierung von Testgütekriterien
- Prognosestudien

zur Verfügung gestellt.

4.1 *Medikamentöse und nicht-medikamentöse Interventionen*

Die Beurteilung einer Studie zu einer medikamentösen oder nicht-medikamentösen Intervention hat zum Ziel, darzustellen, mit welcher Sicherheit sich ein Effekt oder das Fehlen eines Effekts aus dieser Studie ableiten lässt (Ergebnissicherheit). Darüber hinaus ist es notwendig zu beschreiben, ob und in welchem Umfang sich das Studienergebnis auf lokale Gegebenheiten (z.B. Population, Versorgungsbereich etc.) übertragen lässt bzw. welche lokalen Besonderheiten Auswirkung auf die Ergebnisse selbst bzw. deren Interpretation haben oder haben könnten.

Die Ergebnissicherheit wird im Wesentlichen durch drei Komponenten beeinflusst

- die interne Validität der Studie,
- das Studiendesign,
- die zu Grunde liegende Erkrankung, auf die die medizinische Intervention therapeutisch oder präventiv ausgerichtet ist.

Die Kriterien zur Beurteilung der internen Validität sind im Kapitel 3 ausführlich beschrieben. Sie kommen bei der Bewertung von Studien zu medikamentösen und nicht-medikamentösen Interventionen entsprechend zur Anwendung.

Das Studiendesign hat insofern erheblichen Einfluss auf die Ergebnissicherheit, als dass mit epidemiologischen Studien, prospektiv oder retrospektiv, ein kausaler Zusammenhang zwischen Intervention und Effekt nicht dargestellt werden kann, während die experimentelle Forschung hierzu grundsätzlich geeignet ist [1]. Dies gilt zumindest dann, wenn andere das Ergebnis beeinflussende Faktoren ganz oder weitgehend ausgeschaltet werden, weshalb die randomisierte, kontrollierte Studie (RCT) den Goldstandard bei der Beurteilung medikamentöser und nicht-medikamentöser Interventionen darstellt [2].

Das Institut wird daher nur im Ausnahmefall auf nicht-randomisierte Interventionsstudien oder epidemiologische Untersuchungen bei der Effektivitätsbeurteilung zurückgreifen. Diese Ausnahmen müssen begründet sein. Die Begründung kann einerseits in der Nicht-Durchführbarkeit einer RCT liegen, andererseits darin, dass für die vorliegende Fragestellung auch andere Studienarten eine hinreichende Ergebnissicherheit liefern könnten. Diesbezüglich kommt insbesondere der dritte oben genannte Punkt zum Tragen. Bei Erkrankungen, die ohne Intervention nach kurzer Zeit sicher zum Tod führen, liefern beispielsweise mehrere, konsistente Fallbeobachtungen darüber, ob eine bestimmte Intervention diesen zwangsläufigen Verlauf verhindert, hinreichende Ergebnissicherheit [3].

Das Institut wird im Rahmen der Erstellung des Berichtsplans (s.a. die Abschnitte 2.7 und 2.8) daher vorab festlegen, welche Studienarten auf Grund der Fragestellung als theoretisch hinreichend ergebnissicher (bei hoher interner Validität) angesehen werden. Studien, die diesem Qualitätsanspruch nicht mindestens entsprechen (s.a. die Abschnitte 3.9 und 3.10), werden in den Bewertungsprozess nicht einfließen.

Die Übertragbarkeit des Studienergebnisses selbst muss schließlich in einem gesonderten Prozess überprüft werden, der von Design und Qualität der Studie zunächst unabhängig ist.

Literatur

- [1] Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. Lancet 2002; 359: 57-61.
- [2] Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. The Cochrane Database of Methodology Reviews 2002, Issue 4. Art. No.: MR000012. DOI: 10.1002/14651858.MR000012.
- [3] Liberati A, Sheldon TA, Banta HD et al. Eur-Assess project subgroup report on methodology: Methodological guidance for the conduct of health technology assessment. Int J Tech Assess Health Care 1997; 13: 186-219.

4.2 Diagnostische Verfahren

Die Methodik zur Evaluierung diagnostischer Verfahren hat bis heutzutage bei weitem nicht die gleiche Durchdringung erfahren wie diejenige zur Prüfung von therapeutischen Verfahren [1]. Somit wird damit zu rechnen sein, dass nur in Ausnahmefällen die im Folgenden aufgelisteten Informationen zur Bewertung eines diagnostischen Verfahrens komplett vorliegen. Wesentliche Basis für die Bewertung ist eine präzise Fragestellung, da Studien zu diagnostischen Verfahren mit unterschiedlichen Zielen durchgeführt werden, und je nach Zielsetzung nicht alle Informationen relevant sind. Die Formulierung der Fragestellung bzw. der Ablauf, der zu einer endgültigen Fragestellung führt, folgt den im Abschnitt 2.7 niedergelegten Grundsätzen.

4.2.1 Allgemeine Aspekte

- Klare Definition der zu diagnostizierenden / entdeckenden Erkrankung(en) bzw., allgemeiner, des zu entdeckenden Gesundheitsstatus (z.B. Fitness).
- Angaben zur Prävalenz der zu diagnostizierenden / entdeckenden Erkrankung(en) in der zu untersuchenden Population und in Sub-Populationen.
- Eindeutige Definition des „Goldstandard“, d.h. derjenigen Methode, mit der die zu entdeckende Erkrankung (der zu entdeckende Gesundheitsstatus) allgemein akzeptiert eindeutig erfasst werden kann.
- Genaue Beschreibung des diagnostischen Verfahrens mit Darstellung des materiellen und personellen Aufwands für das Erlernen, die Durchführung und die Auswertung.
- Darstellung der Risiken, die mit der Anwendung des diagnostischen Verfahrens verbunden sind, und der Akzeptanz sowie Zumutbarkeit sowohl für die Patienten als auch für das medizinische Personal als auch für die Allgemeinheit (z.B. Umweltrisiken).
- Angaben zu den weiteren Konsequenzen, die sich aus den jeweiligen Befunden ergeben (z.B. weiterführende Diagnostik, Therapie, andersartige nicht-therapeutische Intervention, Monitoring, Lifestyle, Basis für informierte Entscheidungen), mit Darstellung inwieweit sich aus diesen Konsequenzen ein Nutzen für den Patienten ergibt.
- Angaben zu alternativen diagnostischen Verfahren und ggf. Darstellung der Vorteile des neuen gegenüber den alten Verfahren.

4.2.2 Testgütekriterien / Testeigenschaften

I. Technische Voraussetzungen

- Angaben zur analytischen Richtigkeit
- Angaben zur analytischen Sensitivität und Spezifität
- Angaben zur Reproduzierbarkeit (Wiederholbarkeit)
 - Intra-Test-/Beurteiler-Variabilität
 - Inter-Test-/Beurteiler-Variabilität
 - Intra-Patient-Variabilität (Kurz- und Langzeit)
- Ggf. Angaben zur Übereinstimmung (von Messwerten) mit bereits etablierten Standardverfahren
- Angaben zu möglichen Störfaktoren, vor allem systematisch verzerrende Einflüsse

II. Diskriminationsfähigkeit

- Angaben zur diagnostischen Sensitivität und Spezifität, alternativ Angaben zu Likelihood-Quotienten; bei quantitativen Verfahren ROC-Kurven mit Angabe eines geeigneten Trennpunkts (bzw. von geeigneten Trennpunkten) und Begründung für dessen (deren) Wahl (Gewichtung von Sensitivität und Spezifität)

III. Vorhersage

- Angaben zu prädiktiven Werten (bei quantitativen Verfahren Angaben zu prädiktiven Werten bei gewähltem Trennpunkt [gewählten Trennpunkten])

4.2.3 Wirksamkeitsnachweis

- Nachweis, dass sich durch die Anwendung des diagnostischen Verfahrens eine Verbesserung im Outcome für die Patienten ergibt.

Bei Studien zu den technischen Voraussetzungen (4.2.2.I) ist darauf zu achten, dass die entsprechenden Kenngrößen (auch) unter klinischen Alltagsbedingungen und in der Anwendungssituation erhoben werden. Beispielsweise reichen in aller Regel Angaben zur Reproduzierbarkeit des diagnostischen Verfahrens allein bei Gesunden nicht aus.

Bei Studien zur Diskriminationsfähigkeit (4.2.2.II) lassen sich prinzipiell zwei Vorgehensweisen unterscheiden: Zum einen die Untersuchung bei ausgewählten Personen mit bekanntem

Krankheitsstatus, zum anderen die Prüfung in der Anwendungssituation bei unselektierten Personen und unbekanntem Krankheitsstatus [2-4]. Die erste Vorgehensweise ist in aller Regel (bei entsprechenden Ergebnissen) die Voraussetzung für die Durchführung einer (zumeist aufwändigeren) Studie gemäß der zweiten Vorgehensweise. Studien der ersten Vorgehensweise liefern typischerweise zu optimistische Schätzungen der diskriminatorischen Fähigkeit [5,6], was bei einer Bewertung berücksichtigt werden sollte.

Folgende methodische Grundprinzipien sollten darüber hinaus bei der Bewertung von Studien zur Evaluierung der Diskriminationsfähigkeit (4.2.2.II) und Vorhersagefähigkeit (4.2.2.III) diagnostischer Verfahren beachtet werden:

- Klare Formulierung einer Fragestellung und damit verbunden der Studienplanung; dies schließt eine Fallzahlplanung ein, die sich beispielsweise an der gewünschten Präzision der Schätzung (Breite des Konfidenzintervalls) und/oder dem Nachweis des Überschreitens einer Mindestschwelle orientieren kann.
- Durchführung in einer Population von Patienten bzw. Personen, bei der das Verfahren auch zur Anwendung kommen soll (angemessenes Patientenspektrum bzw. Vermeidung eines Spektrum-Bias)
- Von einander unabhängige und verblindete Beurteilung von zu evaluierendem diagnostischen Verfahren und Goldstandard (wechselseitige Blindheit).
- Die Diagnosesicherung (Goldstandard) bzw. die Art der Diagnosesicherung (unterschiedlicher Goldstandard) sollte nicht vom Ergebnis des zu evaluierenden diagnostischen Verfahrens abhängig gemacht werden (Gefahr des Verification-Bias). Falls eine Diagnosesicherung nicht bei allen Patienten durchgeführt werden kann, sollte die Auswahl zufällig erfolgen^c.
- Angemessene Berücksichtigung von Patienten mit unklaren oder nicht interpretierbaren Testergebnissen (kein simpler Ausschluss!)
- Falls das zu evaluierende diagnostische Verfahren in eine Diagnosestrategie eingebettet ist bzw. werden soll, ist eine isolierte Betrachtung dieses Verfahrens häufig nicht sinnvoll (Problem der Abhängigkeit der Testgütekriterien von der Testkombination).

^c Dies löst das Problem allerdings auch nicht vollständig, insbesondere in Situationen mit niedrigen a priori Wahrscheinlichkeiten [7].

- Falls das zu evaluierende diagnostische Verfahren Bestandteil des Goldstandard ist, treten ebenfalls besondere methodische Probleme auf, die einer genauen Erörterung und Berücksichtigung bedürfen.

Im Hinblick auf eine weitestgehend einheitliche und erschöpfende Darstellung (Publikation) wurde unlängst in Analogie zum CONSORT-Statement für Therapiestudien eine ähnliche Empfehlung für Studien zu diagnostischen Verfahren publiziert [8]. Whiting et al. haben eine Checkliste zur Qualitätsbeurteilung von Diagnosestudien im Rahmen von systematischen Übersichten zusammengestellt [9].

Studien zum Wirksamkeitsnachweis (4.2.3.) von diagnostischen Tests können z.B. als Vergleich von Patienten, bei denen das diagnostische Verfahren durchgeführt wird, mit solchen, bei denen das Verfahren nicht durchgeführt wird, konzipiert sein; für die Bewertung solcher Studien gelten im wesentlichen die gleichen Anforderungen, wie in Abschnitt 4.3.4. formuliert. Ein Nachteil derartiger Studien ist, dass der Wert der diagnostischen Information nicht von der sich daraus ableitenden Konsequenz getrennt werden kann, d.h., es kann z.B. bei einem negativen Ausgang nicht differenziert werden, ob die diagnostische Information unzureichend ist, oder (beispielsweise) die Therapie (bei denjenigen mit pathologischem Testergebnis) unwirksam.

Alternativ zur Durchführung des Tests kann auch die Offenlegung der Testergebnisse geprüft werden, d.h. es werden Personen miteinander verglichen, bei denen das Testergebnis bekannt ist, mit solchen, bei denen das Ergebnis verblindet bleibt [4]. Ein solches Vorgehen hat den Vorteil, den natürlichen Verlauf der Personen mit positivem Testergebnis beurteilen zu können.

In einem weiteren Designvorschlag wird in einer Therapieprüfung das zu evaluierende diagnostische Verfahren bei allen Patienten (unabhängig von der Therapiegruppe) angewendet, das Ergebnis aber für die Dauer der Prüfung für alle Patienten verblindet belassen. In einer solchen Studie lässt sich die Wechselwirkung zwischen diagnostischer Information und therapeutischer Wirksamkeit prüfen, ob also Patienten in Abhängigkeit vom Ergebnis des diagnostischen Tests einen unterschiedlichen therapeutischen Nutzen erfahren [3].

Literatur

- [1] Knottnerus JA, van Weel C, Muris JWM. Evidence base of clinical diagnosis. Evaluation of diagnostic procedures. BMJ 2002; 324: 477-480.
- [2] Köbberling J, Trampisch HJ, Windeler J. Memorandum for the Evaluation of Diagnostic Measures. J Clin Chem Clin Biochem 1990; 20: 873-879.
- [3] Richter K, Lange S. Methoden der Diagnoseevaluierung. Internist 1997; 38: 325-336.
- [4] Sackett DL, Haynes RB. Evidence base of clinical diagnosis. The architecture of diagnostic research. BMJ 2002; 324: 539-541.

- [5] Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JHP, Bossuyt PMM. Empirical Evidence of design-related bias in studies of diagnostic tests. JAMA 1999; 282: 1061-1066.
- [6] Whiting P, Rutjes AW, Dinnes J, Reitsma J, Bossuyt PM, Kleijnen J. Development and validation of methods for assessing the quality of diagnostic accuracy studies. Health Technol Assess 2004; 8: iii, 1-234.
- [7] Rückmann A, Windeler J. Selektionsbias bei der Schätzung der Sensitivität von Screeningmaßnahmen. In: Trampisch HJ, Lange S (Hrsg.). Medizinische Forschung – Ärztliches Handeln. München: MMV Medizin Verlag; 1995. p. 227-231.
- [8] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HC; Standards for Reporting of Diagnostic Accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy. The STARD initiative. Radiology 2003; 226: 24-28.
- [9] Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003; 3: 25.

4.3 Screening

Wesentliche Basis für die Bewertung ist eine präzise Fragestellung, da Screening-Programme mit unterschiedlichen Zielen durchgeführt werden, und sie sich aus unterschiedlichen „Bausteinen“ zusammensetzen, die entweder als Ganzes oder in Teilen untersucht werden (können). Die Formulierung der Fragestellung bzw. der Ablauf, der zu einer endgültigen Fragestellung führt, folgt den im Abschnitt 2.7 niedergelegten Grundsätzen.

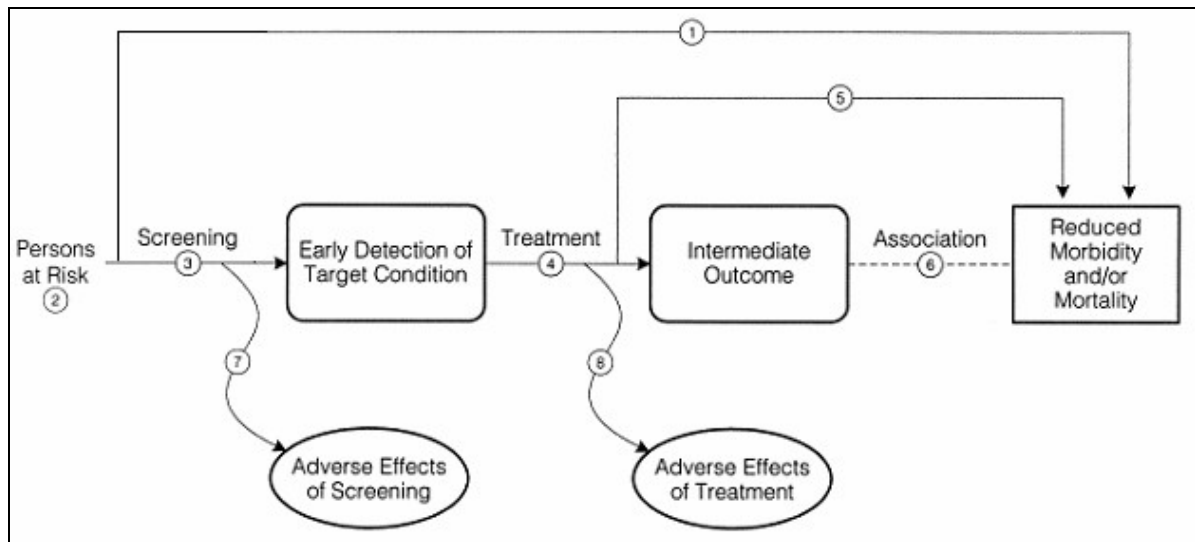
Bei der Bewertung erfolgt eine Orientierung an bereits etablierten und publizierten Kriterien, wie z.B. des UK National Screening Committee (UK NSC [1]) und der US Preventive Services Task Force (US PSTF [2]).

Diese Bewertungskriterien umfassen (1.) die Erkrankung, die es zu entdecken gilt, (2.) den einzusetzenden (diagnostischen) Screening-Test selbst, (3.) die Therapie im Falle eines positiven (pathologischen) Ergebnisses des Screening-Tests bzw. eine andersartige Konsequenz, die aus einem positiven Ergebnis gezogen wird, sowie (4.) das Screening-Programm als Ganzes.

Bei der Bewertung sollte differenziert werden in Programme, bei denen aus dem positiven Ergebnis des Screening-Tests eine therapeutische oder eine anderweitige, nicht-therapeutische Konsequenz folgt. Weiterhin sollte unterschieden werden in Situationen, bei denen direkte Evidenz für die Wirksamkeit des Screening-Programms vorliegt (Vergleich von Personen mit und ohne Screening mit einem Patienten-relevanten Zielkriterium im Rahmen einer Studie, Pfeil 1 in Abbildung), und in Situationen, bei denen die Evidenz über Analogieschlüsse indirekt erhoben wird (mehrere

„Screening-Bausteine“ werden in unterschiedlichen Studien untersucht, z.B. Pfeile 3, 4 und 6 in Abbildung).

Abb.: Screeningkette



modifiziert nach [2].

4.3.1 Erkrankung

Es soll bewertet werden, ob es sich um ein bedeutsames Gesundheitsproblem handelt, wobei sich die Beurteilung dieses Aspekts auf verschiedene Indikatoren beziehen kann, z.B. auf die Häufigkeit, die Schwere oder die Kosten einer Erkrankung, bzw. auf verschiedene Ebenen, z.B. auf die Bevölkerungs- oder eine individuelle Ebene.

Damit verbunden ist die Notwendigkeit, die Epidemiologie und den natürlichen Verlauf der Erkrankung genau zu kennen. Datenquellen dafür entstammen typischerweise epidemiologischen Querschnitt-, Register-, und Kohortenstudien, in Ausnahmefällen auch Fallserien, sowie ökonomischen Studien.

4.3.2 Screening-Test

Es gelten die allgemeinen Anforderungen für die Beurteilung eines diagnostischen Testverfahrens, wie in den Abschnitten 4.2.1 und 4.2.2 formuliert. Es werden jedoch aufgrund der besonderen ethischen Implikationen in aller Regel höhere Anforderungen an die Testgütekriterien und die Qualität der zugrundeliegenden Studien zu stellen sein [3]. Darüber hinaus sollte der Test einfach handhabbar sein, und es ist zu prüfen, ob für den Fall eines positiven Testergebnisses eine

allgemein akzeptierte Strategie zur weiteren diagnostischen Abklärung sowie zu den verfügbaren Alternativen vorliegt.

4.3.3 Therapie

Es ist zu prüfen, ob für Patienten mit positivem Testergebnis eine wirksame Behandlung oder Intervention existiert (für eine diesbezügliche Bewertung siehe Abschnitt 4.1), und ob es Evidenz dafür gibt, dass eine frühzeitige Behandlung zu besseren Ergebnissen führt als eine späte Behandlung. Bei Programmen, bei denen aus dem positiven Ergebnis des Screening-Tests keine unmittelbare therapeutische Konsequenz folgt, muss beurteilt werden, ob die Information, die aus dem positiven Testergebnis resultiert, mit einem anderweitigen (nicht-therapeutischen) Nutzen verbunden ist, z.B. dergestalt, dass die Betroffenen in die Lage versetzt werden, besser informiert persönliche (Lebens-) Entscheidungen zu treffen (z.B. pränatales Screening auf Down-Syndrom, Screening auf Genträger für unheilbare / nicht beeinflussbare Erkrankungen). Für eine solche Beurteilung kann es ggf. sinnvoll sein, Methoden der Entscheidungsanalyse zu verwenden.

4.3.4 Screening-Programm

Idealerweise liegt Evidenz dafür vor, dass das Screening-Programm (als Ganzes) wirksam ist im Hinblick auf eine Reduktion von Morbidität und / oder Mortalität. Zur Beurteilung entsprechender Studien werden die im Abschnitt 4.1 formulierten Kriterien herangezogen. Hierbei ist zu beachten, dass Evidenz aus nicht-randomisierten Studien zur Beurteilung von Screening-Programmen besonders kritisch zu werten ist, da hier spezielle Verzerrungsmechanismen greifen können, wie der Lead-Time-Bias oder der Length-Bias (s.a. Abschnitt 3.19).

Liegt keine direkte Evidenz für die Wirksamkeit des Screening-Programms vor, sondern lediglich Evidenz für die einzelnen Screening-Bausteine, muss neben der Beurteilung der einzelnen Bausteine noch zusätzlich eine Beurteilung der Kohärenz und Konsistenz erfolgen. Kohärenz meint in diesem Zusammenhang, dass sich die Bausteine in einem verständlichen Modell zusammenfügen, Konsistenz bedeutet, dass verschiedene Studien bzw. deren Ergebnisse unter unterschiedlichen Bedingungen zur Kohärenz beitragen [2].

Das Screening-Programm muss einen Netto-Nutzen erzielen, das heißt der Nutzen, der durch ein Screening-Programm erreicht werden kann, muss den möglichen physischen oder psychischen Schaden (hervorgerufen durch den Screening-Test, nachfolgende diagnostische Maßnahmen und / oder die nachfolgende Therapie) übersteigen (zur Bewertung von „unerwünschten Wirkungen“ einer Intervention s.a. Abschnitt 3.7).

Falls das Screening-Programm bzw. seine Bausteine nicht in dem Setting geprüft wurden, in dem das Programm zur Anwendung kommen soll, muss geprüft werden, ob Evidenz dafür vorliegt, dass die Ergebnisse verallgemeinerbar bzw. übertragbar sind.

Literatur

- [1] UK National Screening Committee. Criteria for appraising the viability, effectiveness and appropriateness of a screening programme. <http://www.nsc.nhs.uk/pdfs/criteria.pdf>. Zugriff am 28. Oktober 2004.
- [2] Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, Atkins D; Methods Work Group, Third US Preventive Services Task Force. Current methods of the US Preventive Services Task Force: a review of the process. *Am J Prev Med* 2001; 20(Suppl3): 21-35.
- [3] Ewart RM. Primum non nocere and the quality of evidence: rethinking the ethics of screening. *J Am Board Fam Pract* 2000; 13: 188-196.

4.4 Gesundheitsökonomie

4.4.1 Inhalte

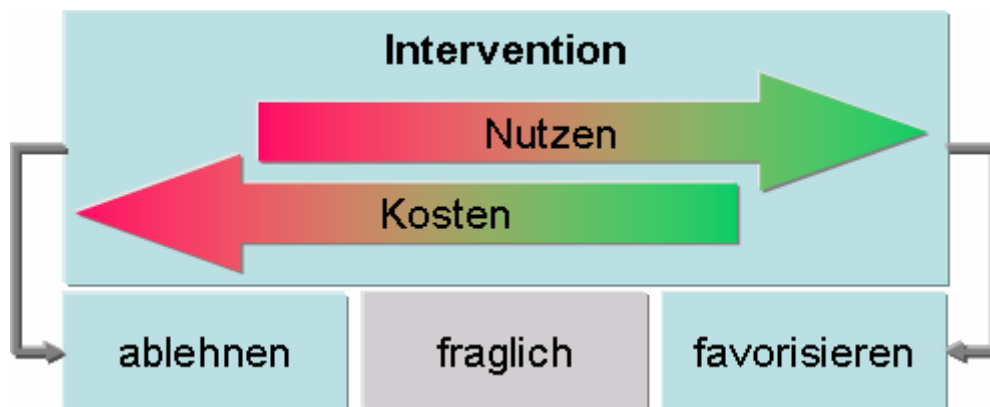
Das Ressort Gesundheitsökonomie erfüllt mittel- und unmittelbare Aufgaben innerhalb des Instituts. Mittelbar im Zusammenhang mit wissenschaftlichen Fragestellungen aus den Bereichen der übrigen Ressorts, bei denen auch ökonomische Aspekte eine Rolle spielen, unmittelbar in Bezug auf direkte Beauftragung durch den Gemeinsamen Bundesausschuss oder das BMGS. Darüber hinaus können auch eigene Fragestellungen für eine wissenschaftliche Aufarbeitung innerhalb des Institutes oder des Ressorts nach Absprache mit dem Steuergremium entwickelt werden.

Eine unter anderem durch den technischen Fortschritt bedingte stete Verteuerung des Gesundheitswesens innerhalb der Industriestaaten kann in den kommenden Jahren zur Folge haben, dass eine Gesundheitsversorgung auf höchstem technischem Niveau in der Routineversorgung weniger durch technische als vielmehr durch ökonomische Hindernisse gehemmt wird.

Wie in allen anderen Lebens- und Wirtschaftsbereichen auch, ist das Gesundheitswesen geprägt von der Zuteilung und Rationierung von Ressourcen wie Material, Mittel, Personal und Zeit. Während eine Rationierung durch die Budgetierung von Leistungen bereits Einzug in das Gesundheitswesen gehalten hat, wird für Teilbereiche eine Effizienzreserve diskutiert, deren Ausschöpfung möglicherweise herangezogen werden kann, um vorhandene Rationierungen teilweise zu entschärfen. Dies bedeutet im weiten Sinne als Zielsetzung der Gesundheitsökonomie

die Vermeidung einer Fehl-, Unter- oder Überversorgung als wichtige Aspekte von Qualitätsdefiziten im Sinne einer unausgewogenen Kosten-Nutzen-Allokation. Kosten-Nutzen-Bewertungen können sich dabei auf verschiedene Aspekte bei der Bewertung vorhandener und zukünftiger medizinischer und gesundheitspolitischer Interventionen beziehen.

Abb.: Kosten-Nutzen-Relation und Entscheidungsspektrum



Während Konstellationen mit geringem Nutzen und hohen Kosten oder solche mit großem Nutzen und geringen Kosten relativ eindeutig beurteilt werden können, ergeben sich für die sicher häufiger anzutreffenden übrigen Konstellationen komplexere Entscheidungsprozesse [1], in die insbesondere auch weitere Rahmenbedingungen mit einfließen müssen.

Im Unterschied zu anderen Wirtschaftsbereichen ergeben sich dabei insbesondere auch ethische Fragestellungen, etwa im Zusammenhang einer Güterabwägung zwischen dem Prinzip der Solidargemeinschaft und des einzelnen Individuums. Dass die Ökonomie letztendlich nicht Vorrang vor der medizinischen Indikation oder der Wirksamkeit einer medizinischen Maßnahme haben darf, gilt dabei als international akzeptiert [2-6].

Jedoch wird es zukünftig weniger um die Fragestellung gehen, ob eine Rationierung unethisch ist, sondern vielmehr darum, welche Rationierung unethisch ist [7]. Dabei wird man sich letztlich aus ökonomischer Perspektive weniger mit den Ausgabengrößen beschäftigen als mit den Regeln, nach denen die Mittelverteilung im Gesundheitswesen erfolgt [8].

4.4.2 Zielgrößen

Die Beurteilung von Kosten-Nutzen-Relationen ist aber auch abhängig von der verfolgten gesellschaftlichen Zielsetzung, wie etwa der Maximierung der durchschnittlichen Lebenserwartung der Gesamtbevölkerung versus einer Maximierung der durchschnittlichen Lebenserwartung von sozialen Einzelgruppen bei oder mit vermuteter geringer Lebenserwartung. Gleiches gilt etwa bei der Verfolgung von Strategien zur Steigerung der Lebensqualität oder dem Vergleich Lebenserwartung versus Lebensqualität. Hier muss zwangsläufig zunächst in Bezug auf eine einzelne Fragestellung die jeweilige Zielgröße festgelegt werden, auf die dann weitere Betrachtungen wie etwa eine Kosten-Nutzen-Analyse abzielen können. Bei der Definition mehrerer Zielgrößen müssen die Auswirkungen auf die jeweiligen Zielgrößen beschrieben werden, um letztlich eine Abwägung der Zielgrößen, insbesondere auch in Hinsicht auf ethische Aspekte, ermöglichen zu können.

Dabei ist aber auch zu beachten, dass verschiedene in den neunziger Jahren veröffentlichte Vorschläge zu einer adäquaten Kosten-Nutzen-Relation gemessen an quality-adjusted-live-years (QALY) sich bislang nicht durchsetzen konnten in Hinblick auf daraus ableitbare standardisierbare Empfehlungen. Dennoch werden häufig 50.000 € pro gewonnenem QALY als noch akzeptabel bzw. kosteneffektiv bezeichnet. Die UK Health Authorities haben in Großbritannien ein Matrixmodell vorgestellt, das eine Kosten-Nutzen-Bewertung auf unterschiedlichen Evidenzklassen abbildet. Damit wird zumindest der Aspekt unterstrichen, dass eine reine Kosten-Nutzen-Betrachtung nur eine Dimension der Versorgungsqualität darstellt. Eine kombinierte Betrachtung aus EBM und KNA zur Ressourcenallokation und Priorisierung medizinischer Maßnahmen erscheint somit sinnvoll, wenn sie weitere Dimensionen wie Solidarität und Subsidiarität angemessen und ausgewogen berücksichtigt.

4.4.3 Methodik

Für die Heranziehung und valide Interpretation von Studienergebnissen und Publikationen im Bereich Gesundheitsökonomie müssen diese definierten Standards genügen, die sich allgemein an den Begriffen Transparenz, Vergleichbarkeit und Qualität orientieren [9]. Dabei haben sich in den letzten Jahren unterschiedliche Bewertungssysteme für die Planung und die Bewertung von Studien etabliert. Hierunter fallen unter anderem die des Commonwealth of Australia (1995), des Ontario Ministry of Health (1994), des Canadian Co-ordinating Office for Health Technology Assessment (1994), der Task Force on Principles of Economic Analysis of Health Care Technology (1996), des US Public Health Service Panel on Cost-Effectiveness in Health and Medicine [10] sowie die Empfehlungen zur gesundheitsökonomischen Evaluation [11], die sich nah an den Empfehlungen der BMJ-Checkliste von Drummond orientieren [12].

Es wird eine der wesentlichen ersten Aufgabe des Ressorts sein, die unterschiedlichen Instrumente ihrerseits zu evaluieren und zu bewerten, um einen Institutseigenen Standard zu erarbeiten, zu publizieren und verbindlich anzuwenden.

Bei Kostenanalysen sind die unterschiedlichen Arten veranschlagter Interventionskosten transparent darzustellen. Hier sind direkte und indirekte Kosten, fixe und variable Kosten, Gemeinkosten und Grenzkosten abzugrenzen. Die Berechnung und der Vergleich von Opportunitätskosten ist dagegen in der Praxis schwer oder nur aufwendig umsetz- und darstellbar [13]. Die ebenfalls kritisch zu betrachtende Berechnung indirekter Kosten muss transparent nachvollziehbar sein und kann sich am Humankapitalansatz oder der Friktionskostenmethode orientieren. Ebenfalls den indirekten Kosten zuzurechnen sind so genannte intangible Kosten, die sich nicht oder nur indirekt monetär bewerten lassen.

Auf Seiten der Nutzenanalyse wird dagegen die Gesundheitsbezogene Lebensqualität zu betrachten sein. Diese zeigt sich darin, wie zufrieden Menschen mit ihrem physischen, psychischen und sozialen Gesundheitszustand sind bzw. wie sie diese drei Aspekte ihrer Gesundheit bewerten [14]. Dazu stehen so genannte nutzentheoretische Messverfahren psychometrischen Verfahren gegenüber. Während erstere eher eine allgemeine Herangehensweise darstellen, sind letztere konkreter in Bezug auf Vorstellungen, was Lebensqualität beinhaltet. Insgesamt ist die Diskussion darüber, welches Verfahren am besten geeignet ist, um Lebensqualität messbar zu machen, nicht abgeschlossen. Dies gilt auch für die Bewertung von Validität und Reliabilität aus methodologischer Sicht. Hier gilt es, die weitere empirische Forschung kritisch zu beobachten und zu begleiten.

Bei der Betrachtung von Kosten und Nutzen gelten als wichtigste Formen die Kostenminimierungs-Analyse (CMA), die Kosten-Effektivitäts- oder –Wirksamkeits-Analyse (CEA), Kosten-Nutzwert-Analyse (CUA) sowie die Kosten-Nutzen-Analyse im engeren Sinn (CBA). Die Kostenanalyse erfolgt bei allen Formen in ähnlicher Form, unterschiedlich geht dagegen die Nutzenbewertung ein. Die CMA geht von einer unterschiedlichen Kostensituation bei unterstelltem gleichem Nutzen aus und ist daher naturgemäß kritisch einzuschätzen. Bei der CEA wird der Nutzen in natürlichen Einheiten gemessen, so dass sich ein abschließender Kosten-Nutzen-Quotient ergibt, der eine Vergleichbarkeit mit anderen Interventionen nur dann ermöglicht soweit einheitliche Nutz- bzw. Effektivitätsparameter verwendet werden. Die Wahl geeigneter Effektivitätsparameter ist häufig schwierig und sollte sich an definierten patienten-relevanten Endpunkten orientieren. Hierzu sind in der Regel große randomisierte und kontrollierte epidemiologische Langzeitstudien notwendig. Da es häufig gilt, Zusatzkosten und –nutzen unterschiedlicher Interventionen zu vergleichen, bedient man sich hier der inkrementellen Analyse, die das Verhältnis von Zusatzkosten zu Zusatznutzen bestimmt. Bei der CBA wird auch der Nutzen

in monetären Einheiten gemessen, z. B. als Einspareffekt. Die dabei häufig angewandte Umrechnung von Lebensqualität und –dauer in monetäre Einheiten, etwa über die Zahlungsbereitschaft („willingness to pay“ - WTP) ist dabei grundsätzlich als problematisch einzustufen. Bei der CUA [15] werden die Kosten zumeist (gewonnenen) qualitätsadjustierten Lebensjahren (QALYs) gegenübergestellt. Sie ermöglicht den Vergleich unterschiedlicher Interventionsstrategien und eignet sich insbesondere für Fragestellungen bezüglich der Lebensqualität, Mortalität und Morbidität. Nach den Empfehlungen des US Panel on Cost-Effectiveness in Health and Medicine soll eine gesundheitsökonomische Studie grundsätzlich als Kosten-Nutzwert-Analyse durchgeführt werden.

Literatur

- [1] Haycox A, Bagust A, Walley T. Clinical guidelines – the hidden costs. BMJ 1999; 318: 391-393.
- [2] Pellegrino ED. The commodification of medical and health care: the moral consequences of a paradigm shift from a professional to a market ethic. J Med Phil 1999; 3: 243-266.
- [3] Schwartz P. Medical ethics under managed care. Int J Fert 1996; 2: 124-128.
- [4] Ulsenheimer K. Qualitätssicherung und Risk-Management im Spannungsverhältnis zwischen Kostendruck und medizinischem Standard. MedR 1995; 11: 438-442.
- [5] Weltärztebund. Erklärung des Weltärztebundes zur Verantwortung des Arztes hinsichtlich der Qualität der medizinischen Versorgung. Oktober 1996.
- [6] Westhofen M. Operative Hochleistungsmedizin. Handlungszwang zwischen ärztlicher Ethik, wirtschaftlichem Erfolg und Qualitätskontrolle. In: Brudermüller G (Hrsg). Angewandte Ethik und Medizin. Würzburg: Schriften des Instituts für Angewandte Ethik e.V. 1999, Bd. 1; 171-184.
- [7] Lauterbach KW. Utilitarismus und Kant. In: Lauterbach, Schrappe (Hrsg). Gesundheitsökonomie, Qualitätsmanagement und Evidence-based Medicine. 2 Aufl. Stuttgart, New York: Schattauer 2004; 3-10.
- [8] Breyer F, Zweifel P. Gesundheitsökonomie. 2 Aufl. Berlin, Heidelberg, New York: Springer 1997.
- [9] Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. BMJ Economic Evaluation Working Party. BMJ 1996; 313: 275-283.
- [10] Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendation of the Panel on Cost-Effectiveness in Health and Medicine. JAMA 1996; 276: 1253-1258.
- [11] Hannoveraner Konsensus Gruppe. Deutsche Empfehlungen zur gesundheitsökonomischen Evaluation – revidierte Fassung des Hannoveraner Konsens. Gesundheitsökonomie u Qualitätsmanagement 1999; 4: A 62-65.
- [12] Kurscheid T, Schrappe M, Lauterbach KW. Kritische Bewertung gesundheitsökonomischer Studien. In: Lauterbach, Schrappe (Hrsg). Gesundheitsökonomie, Qualitätsmanagement und Evidence-based Medicine. 2 Aufl. Stuttgart, New York: Schattauer 2004; 114-126.
- [13] Schumann J. Mikroökonomie. 6 Aufl. Berlin Heidelberg New York: Springer 1992.

- [14] Konerding U. Gesundheitsbezogene Lebensqualität In: Lauterbach, Schrappe (Hrsg). Gesundheitsökonomie, Qualitätsmanagement und Evidence-based Medicine. 2 Aufl. Stuttgart, New York: Schattauer 2004; 160-182.
- [15] Sinclair C, Frankel M. the effect of quality assurance activities on the quality of mental health services. Qual Rev Bull 1982; 8: 7-15.

4.5 Leitlinien und Disease Management Programme

4.5.1 Hintergrund der Bewertung von Leitlinien

Medizinische Leitlinien gelten als Schlüsselinstrumente zur Verbesserung und Sicherung der medizinischen Qualität in der Patientenversorgung [1]. Ihr Sinn liegt darin, durch die Formulierung von konkreten Handlungsempfehlungen unangemessene Unterschiede in der Patientenversorgung zu reduzieren und die Versorgung der Patienten zu verbessern. Darüber hinaus werden sie in Deutschland auch als Basis für Entscheidungen zu Steuerungszwecken im Gesundheitswesen wie z.B. im Zusammenhang mit der Formulierung von Anforderungen von Disease-Management-Programmen (nach §137f SGB V) genutzt und beeinflussen so auch auf der strukturellen Ebene zunehmend Entscheidungen im Gesundheitswesen.

Vor diesem Hintergrund muss sichergestellt werden, dass medizinische Leitlinien auf der besten verfügbaren und aktuellen wissenschaftlichen Evidenz basieren und unter Berücksichtigung klinischer Erfahrung formuliert werden.

In vielen Fällen fehlt jedoch der Bezug zur aktuellen wissenschaftlichen Evidenz [2,3] und Leitlinien zu gleichen Themen weisen teilweise erhebliche inhaltliche Unterschiede in ihren Empfehlungen auf [4,5].

Ein wesentlicher Grund hierfür ist, dass die international geforderten Qualitätsstandards für die Entwicklung von Leitlinien nicht durchgängig berücksichtigt werden [6-8].

4.5.2 Ziele der Bewertung von Leitlinien

Die spezielle Beurteilung medizinischer Leitlinien soll zu einer Verbesserung der Versorgung durch mehr Transparenz im Gesundheitswesen führen. Hierzu ist es insbesondere notwendig

- methodisch und inhaltlich gute von schlechten Leitlinien zu diskriminieren,
- klare fachliche Aussagen über die Sinnhaftigkeit und Wirksamkeit der Umsetzung verschiedener medizinischer Empfehlungen zu geben,

- dem Gemeinsamen Bundesausschuss oder seinen Gremien eine Entscheidungsgrundlage für die Beratungen über strukturierte Behandlungsprogramme zu bieten,
- sicherzustellen, dass nur geprüfte (qualitätsgesicherte) Leitlinien in die Versorgung eingeführt werden, bei denen Hinweise auf eine Verbesserung des Outcomes bestehen,
- Forschungsbedarf zu benennen und sinnvolle Projekte zur Entwicklung und Implementierung von evidenzbasierten Empfehlungen anzustoßen sowie
- die Einbindung in umfassende Qualitätsmanagement - Prozesse (UQM) zu fördern.

Darüber hinaus ermöglichen die Ergebnisse dieser Arbeit dem Nutzer von Leitlinien (Ärzte, Gesundheitseinrichtungen, gesundheitspolitische Ausschüsse- Entscheider im Gesundheitswesen und Patienten) eine Orientierung über sinnvolle und adäquate Empfehlungen zu prioritären Versorgungsproblemen.

Für die spezielle inhaltliche Bewertung medizinischer Leitlinien soll die vorhandene methodische Kompetenz und inhaltliche Expertise von externen Instituten / Einrichtungen / Organisationen soweit wie möglich genutzt und einbezogen werden.

4.5.3 Methoden der Bewertung von Leitlinien

Wesentliche Aspekte bei der Bewertung und Prüfung von medizinischen Leitlinien sind

- Interne Validität,
- Relevanz, Angemessenheit und Praktikabilität der Empfehlungen in der Versorgung sowie
- Externe Validität.

Zentrales Qualitätskriterium einer medizinischen Leitlinie ist die Validität. Eine Leitlinie gilt dann als valide, wenn durch ihre Anwendung tatsächlich der erwartete Nutzen (Outcome) (medizinisch/ökonomisch) erzielt werden kann [1,9,10]. Streng genommen lässt sich die Validität einer Leitlinie nur durch eine rigorose Effektevaluation prüfen [2]. Aufgrund des finanziellen und personellen Aufwands ist dies jedoch nicht bei jeder bestehenden Leitlinien zu realisieren. Auch Pilotstudien wurden bislang vor der Veröffentlichung nur für wenige Leitlinien durchgeführt [11,12].

Zur abschließenden Bewertung wird deshalb ein mehrstufiger Prozess, der verschiedene Aspekte der formalen und inhaltlichen Prüfung kombiniert und in einem Gesamtbericht zusammenfasst, durchgeführt. Die dabei vom Institut verwendete Methodik wird unter Berücksichtigung aktueller wissenschaftlicher Publikationen sowie nationaler und internationaler Erfahrungen regelmäßig überprüft und ggf. aktualisiert.

1. Formale Prüfung

Eine Annäherung an die Frage der Gültigkeit von Leitlinien gelingt durch eine formale Prüfung mit methodischen Kriterien, für die ein großer Einfluss auf die Validität von Leitlinien nachgewiesen wurde [5,6,13]. Die formale Leitlinienbewertung erfolgt in Anlehnung an die Methoden des Leitlinien-Clearingverfahrens und unter Bezug auf die Beurteilungskriterien für Leitlinien von Bundesärztekammer und Kassenärztlicher Vereinigung [14,15] in strukturierter Form. Sie wird durch zwei unabhängige Bewerter durchgeführt. Bei unterschiedlichen Einschätzungen werden die Fragen erneut diskutiert und einer erneuten Bewertung unterzogen. Bleibt der Dissens bestehen, werden die unklaren Aspekte gesondert dokumentiert.

Die Bewertung der wissenschaftlichen Grundlagen für die Formulierung der einzelnen Empfehlungen innerhalb der Leitlinien ist von besonderer Bedeutung. Die bisher mit den gängigen Instrumenten (ÄZQ-Checkliste^d, AGREE-Instrument^e) abgefragten Kriterien zur Identifizierung und Interpretation der Evidenz sowie zur Formulierung der Leitlinienempfehlungen sind im Wesentlichen Transparenzkriterien, in denen allein die Darstellung des Prozesses z.B. zur Literaturrecherche positiv bewertet wird, ohne z.B. eine Beurteilung der Vollständigkeit und Aktualität der Recherche abzugeben. Zu den wesentlichen Schlüsselempfehlungen einer Leitlinie muss deshalb die Ableitung der Empfehlungen aus der zugrunde liegenden Evidenz im Einzelnen geprüft werden.

Die Identifizierung wesentlicher Schlüsselempfehlungen erfolgt in Abstimmung mit den betroffenen Ressortleitern und externen Experten.

2. Leitlinienvergleich

Weltweit existieren zu bestimmten medizinischen Fragestellungen mehrere Leitlinien, die sich methodisch und inhaltlich z. T. sehr stark voneinander unterscheiden [4].

Mit einem synoptischen Vergleich der Kernempfehlungen aus relevanten Publikationen wird deutlich, welche Maßnahmen übereinstimmend empfohlen werden und in welchen diagnostischen oder therapeutische Verfahren keine Übereinstimmung besteht. Der synoptische Vergleich stellt jedoch lediglich eine Erleichterung im Bewertungsprozess dar, eine Überprüfung der Evidenzgrundlagen ist auch bei übereinstimmend empfohlenen Verfahren sinnvoll.

Insbesondere ist eine Gegenüberstellung der Leitlinienempfehlungen mit in der Routineversorgung allgemein üblichen Verfahren sinnvoll. Sofern komplexe Veränderungen in den Leitlinien

^d In AWMF 2001, Leitlinien-Manual

^e Appraisal of Guidelines Research and Evaluation (AGREE)

gefordert werden, ist die Implementierung schwieriger und muss von unterstützenden Maßnahmen und Tools begleitet werden [5].

Zur Identifizierung der aktuellen, existierenden Leitlinien für die jeweilige Fragestellung wird unter Berücksichtigung der im Abschnitt „Struktur der Recherche“ genannten Vorgehensweise eine umfassende Leitlinienrecherche in den relevanten Fachdatenbanken (Leitlinien- und Literaturdatenbanken) durchgeführt. Die Recherchestrategie (Suchbegriffe, Auswahl der Datenbanken etc.) sowie Ein- und Ausschlusskriterien werden vorab unter optionaler Hinzuziehung eines fachlichen Experten festgelegt und dokumentiert.

3. Bewertung der Relevanz, Angemessenheit und Praktikabilität von Empfehlungen

Was wissenschaftlich richtig ist, muss nicht notwendiger Weise auch sinnvoll, praktikabel und angemessen sein. Einzelempfehlungen/Kernpunkte von Leitlinien können in Hinblick auf ihre Relevanz, Angemessenheit und Praktikabilität von den Betroffenen sehr unterschiedlich bewertet werden. Dies lässt sich durch eine formale Prüfung nicht erfassen, sondern sollte mittels fachlicher Expertise in Form eines Peer Reviews und durch Patienten bewertet werden. Dazu können beispielsweise unabhängige Fokusgruppen aus Anwendern, unterschiedlicher Versorgungsbereiche [11,16,17] und Patienten eingerichtet werden, die die inhaltliche Relevanz der Empfehlungen prüfen und ggf. wichtige fehlende, versorgungsrelevante Aspekte benennen.

Die Zusammenstellung der Fokusgruppen erfolgt durch das Institut.

Stellt die Fokusgruppe fest, dass wesentliche Fragestellungen für die Verbesserung der Versorgung nicht durch die vorliegenden Leitlinien abgedeckt werden oder dass aktuell vorliegende Evidenz nicht berücksichtigt wurde, so muss zu diesen Schlüsselfragen die fehlende Evidenz recherchiert und aufgearbeitet werden.

4. Verbesserung des Outcomes

Die zentrale Frage bei der Entwicklung und Implementierung von Leitlinien ist, ob die Umsetzung der Empfehlungen zu messbaren Verbesserungen in der Versorgung führt. Liegen Ergebnisse aus Pilotstudien oder Projekten vor, in denen die Leitlinien getestet wurden, sind diese in die Gesamtbewertung einzubeziehen (z.B. durch Darlegung der Methoden, Qualitätsindikatoren, Ergebnisse und Konsequenzen). Bei ausländischen Leitlinien ist hierbei insbesondere zu prüfen, ob die Ergebnisse auf das deutsche Gesundheitssystem übertragbar sind und/oder welche strukturellen Voraussetzungen für die erfolgreiche Einführung der Empfehlungen realisiert werden müssen.

4.5.4 Darlegung der Qualitätsprüfung

Aus den vorliegenden Ergebnissen der Prüfung wird ein strukturierter Bericht erstellt, der dem Gemeinsamen Bundesausschuss als Grundlage weiterer Beratungen zur Verfügung gestellt wird. In der Frage der DMPs werden auf der Grundlage der Bewertungen konkrete Vorschläge für die Strukturierung von DMPs erarbeitet.

Die Berichte können weiterhin auch als Basis für die Erstellung von themenbezogenen Informationen für Ärzte und Patienten dienen.

4.5.5 Update der Berichte

Um die Aktualität der Berichte zu gewährleisten, ist eine regelmäßige Überprüfung Voraussetzung. Die Berichte werden in regelmäßigen Abständen und bei relevanten Änderungen in der Evidenzlage überarbeitet. Ergeben sich daraus für die Strukturierung der Versorgung relevante Änderungen, werden die Ergebnisse an den Gemeinsamen Bundesausschuss zur Beratung weitergeleitet.

4.5.6 Abgabe von Empfehlungen zu Disease-Management-Programmen

Der Gemeinsame Bundesausschuss nach §91 SGB V gibt Empfehlungen für strukturierte Behandlungsprogramme ab. Dies betrifft die Auswahl der zu empfehlenden chronischen Erkrankungen, orientiert an:

1. der Zahl der von der Krankheit betroffenen Versicherten,
2. den Möglichkeiten zur Verbesserung der Qualität der Versorgung,
3. der Verfügbarkeit von evidenzbasierten Leitlinien,
4. dem sektorenübergreifenden Behandlungsbedarf,
5. der Beeinflussbarkeit des Krankheitsverlaufs durch Eigeninitiative des Versicherten und
6. der Höhe des finanziellen Aufwands für die Behandlung.

Für die ausgewählten chronischen Erkrankungen empfiehlt der Gemeinsame Bundesausschuss dem Bundesministerium für Gesundheit und soziale Sicherung Anforderungen an die Ausgestaltung der Programme. Dies betrifft insbesondere:

1. die Behandlung nach dem aktuellen Stande der medizinischen Wissenschaft unter Berücksichtigung von evidenzbasierten Leitlinien oder nach der jeweils besten, verfügbaren Evidenz sowie unter Berücksichtigung des jeweiligen Versorgungssektors,
2. durchzuführende Qualitätssicherungsmaßnahmen,

3. Voraussetzungen und Verfahren für die Einschreibung des Versicherten in ein Programm, einschließlich der Dauer der Teilnahme,
4. Schulungen der Leistungserbringer und der Versicherten,
5. die Dokumentation und
6. die Bewertung der Wirksamkeit und der Kosten (Evaluation) und die zeitlichen Abstände zwischen den Evaluationen eines Programms sowie der Dauer seiner Zulassung nach §137g SGB V.

Ziele und Rolle des Institutes bei der Abgabe von Empfehlungen zu Disease-Management-Programmen (DMPs)

Ziel ist es, für den Gemeinsamen Bundesausschuss wissenschaftliche Erkenntnisse zu einzelnen Aspekten der DMPs aufzubereiten und als Grundlage für seine Entscheidungen zur Verfügung zu stellen. Da es bei den Anforderungen nicht alleine um die Benennung von evidenzbasierten Leitlinien geht, sondern auch um die Konkretisierung der Leistungspflicht für die teilnehmenden Leistungserbringer, können die Anforderungen Bestandteil von aktuellen evidenzbasierten Leitlinien sein, sie können darüber hinaus aber auch Empfehlungen enthalten, die auf der Basis von aktuellen Ergebnissen aus systematischer Literaturrecherche und Bewertung erstellt wurden. Darüber hinaus können auf dieser Grundlage auch Qualitätsindikatoren entwickelt werden, die die Verbesserung der Versorgung abbilden.

Weitere Möglichkeiten der Unterstützung des Gemeinsamen Bundesausschusses bei der Entwicklung von DMPs werden in Abstimmung mit den zuständigen Gremien geprüft.

Literatur

- [1] Europarat. Methodology for drawing up guidelines on best medical practice – recommendation No R(01)13. Europarat, Straßburg, Eigenverlag, <http://www.coe.int>. Deutsche Version: Entwicklung einer Methodik für die Ausarbeitung von Leitlinien für optimale medizinische Praxis. Empfehlung Rec. (2001) 13 des Europarates und erläuterndes Memorandum. Z ärztl Fortb Quallsich 2001; 95 (Suppl. III): 96.
- [2] Helou A, Ollenschläger G. Ziele, Möglichkeiten und Grenzen der Qualitätsbewertung von Leitlinien. Ein Hintergrundbericht zum Nutzermanual der Checkliste „Methodische Qualität von Leitlinien“. Z ärztl Fortbild Quallsich 1998; 92: 361-365.
- [3] Savoie I, Kazanjian A, Bassett K. Do clinical practice guidelines reflect research evidence? J Health Serv Res Policy 2000; 5: 76-82.
- [4] Irani J, Brown CT, van der Meulen J, Emberton M. A review of guidelines on benign prostatic hyperplasia and lower urinary tract symptoms: are all guidelines the same? BJU Int 2003; 92: 937-942.

- [5] Burgers JS, Cluzeau FA, Hanna SE, Hunt C, Grol R, and the AGREE Collaboration. Characteristics of high quality guidelines: Evaluation of 86 clinical guidelines developed in ten European countries and Canada. *Int J Technol Assess Health Care* 2003; 19: 148-157.
- [6] Cluzeau F, Littlejohns P, Grimshaw J, Feder G, Moran S. Development and application of a generic methodology to assess the quality of clinical guidelines. *Int J Qual Health Care* 1999; 11: 21-28.
- [7] Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed literature. *JAMA* 1999; 281: 1900-1905.
- [8] Bandolier Forum, Independent evidence-based health care, On Care Pathways, July 2003, p. 1-12. <http://www.ebandolier.com>; Zugriff am 10. Oktober 2004.
- [9] Grol R, Dalhuijsen J, Thomas S, in t'Veeld C, Rutten G, Mokkink H. Attributes of clinical guidelines in general practice: observational study. *BMJ* 1998; 317: 858-861.
- [10] Worrall G, Chaulk P, Freake D. The effects of clinical practice guidelines on patient outcomes in primary care: a systematic review. *CMAJ* 1997; 156: 1705-1712.
- [11] Ärztliches Zentrum für Qualität in der Medizin. Leitlinien-Clearingbericht „COPD“. ÄZQ-Schriftenreihe Band 14; Verlag videel, Niebüll; 2003.
- [12] Ärztliches Zentrum für Qualität in der Medizin. Leitlinien-Clearingbericht „Depression“. ÄZQ-Schriftenreihe Band 12; Verlag videel, Niebüll; 2003.
- [13] Field MJ, Lohr KN. Clinical practice guidelines. Directions from a new program. Institute of Medicine, Washington D.C. 1990.
- [14] Ärztliche Zentralstelle Qualitätssicherung. Checkliste „Methodische Qualität von Leitlinien“. *Dtsch Arztebl* 1998; 95: A2576-A2578, C1838-C1840.
- [15] Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften, Ärztliche Zentralstelle Qualitätssicherung. Das Leitlinien-Manual. *Z ärztl Fortbild Qualsich* 2001; 95 (Suppl. I): 1-84.
- [16] Kitzinger J. Introducing focus groups. *BMJ* 1995; 311: 299-302.
- [17] Kean S. Focus Group Interviews: Ein qualitativer Forschungsansatz in der Pflege. *Pflege* 2000; 13: 145-151.

4.6 Systematische Übersichten und HTA-Berichte

Systematische Übersichten (Reviews) und HTA-Berichte stellen als zusammenfassende Darstellung der aktuellen Studienlage eine wertvolle Entscheidungsgrundlage für Arzt und Patient, aber auch für Entscheidungsträger im Gesundheitswesen dar. Ebenso wird das Institut bei der Erstellung von Berichten und Patienten-Informationen auf vorliegende systematische Übersichten und HTA-Berichte zurückgreifen. Notwendige Voraussetzung hierfür ist, dass die Methodik dieser Arbeiten den Anforderungen des Instituts entspricht.

Eine systematische Übersicht kann wie jede andere wissenschaftliche Publikation nicht nur zu fehlerhaften Ergebnissen führen, sondern auch durch systematische Verzerrungen in ihrer Aussage

gelenkt werden [1]. Grundsätzlich ist zu berücksichtigen, dass es auch bei systematischen Übersichten erhebliche Qualitätsunterschiede gibt, die eine generelle Einstufung auf eine hohe Evidenzstufe nicht immer rechtfertigen.

In den letzten Jahren ist die Anzahl der systematischen Übersichten stark angestiegen [2]. Mit der Zunahme von publizierten systematischen Übersichten zeigen sich auch Probleme durch systematische Übersichten mit widersprüchlichen Aussagen und Unterschieden zwischen Ergebnissen von systematischen Übersichten [3] und randomisierten kontrollierten Studien zu vergleichbaren Fragestellungen [4].

Notwendige, aber nicht hinreichende Voraussetzung für die Einbeziehung von systematischen Übersichten in die Berichte des Instituts ist eine methodische Prüfung in Anlehnung an das Quorum Statement [1] und die Methoden des Centre for Reviews and Dissemination in York, nach denen systematische Übersichten geprüft werden, bevor sie in die Datenbank DARE aufgenommen werden (s.a. <http://www.york.ac.uk/inst/crd/>). Darüber hinaus werden diesen formalen Prüfungen inhaltliche Prüfungen nachgestellt.

Ein wesentlicher Aspekt in der Bewertung ist neben der Beschreibung der angewandten Methodik zur Erstellung der Übersicht auch die Frage, ob die einbezogenen Studien einer Qualitätsprüfung unterzogen wurden, und was das Ergebnis war. Die systematische Übersicht kann nicht besser sein, als die Einzelstudien, auf deren Basis sie erstellt wurde. Die Einbeziehung von Studien schlechter Qualität kann zu einer deutlichen Verzerrung des Ergebnisses [5] führen.

Kritische Bewertung von HTA-Berichten

Im Gegensatz zu systematischen Übersichten werden HTA-Berichte nicht primär als Arbeitshilfen für den klinischen Arbeitsalltag geschrieben. Ihr Ziel ist es, entscheidungsrelevante Informationen auf verschiedenen Ebenen der Steuerung des Gesundheitswesens zur Verfügung zu stellen [6]. Hierzu zählen beispielsweise Investitionsentscheidungen, Kostenübernahmeentscheidungen und die Gestaltung des Leistungskatalogs. Darüber hinaus werden HTA-Berichte oft als wichtige Grundlage zur Erstellung medizinischer Leitlinien und von Patienteninformationen heran gezogen.

Ein standardisiertes Format für HTA-Berichte, wie es beispielsweise von der EUR-ASSESS-Gruppe vorgeschlagen wurde [7], hat sich bisher nicht durchgesetzt, was deutlich macht, dass der lokale Bezug immer einen wesentlichen Einfluss auf das Berichtsformat hat. Qualitätskriterien, nach denen HTA-Berichte geprüft werden können, beziehen sich deshalb primär auf die ordnungsgemäße, transparente und reproduzierbare Durchführung und Interpretation der Einzelbestandteile (z.B. gesundheitsökonomischer Übersichten oder der gesundheitsökonomischen Analyse) [8].

Das Institut wird sich zur Prüfung von HTA-Berichten an den international akzeptierten Standards (z.B. INAHTA^f oder ECHTA^g) für die Erstellung und Bewertung [9,10] orientieren.

Insbesondere werden folgende Aspekte [6] näher betrachtet:

- die Darstellung des Hintergrunds der Technologiebewertung,
- die Formulierung der spezifischen Fragestellung,
- Angaben zum Status quo der Technologie,
- technische Charakteristika,
- die systematische Evaluation der Sicherheit und klinischen Wirksamkeit,
- die gesundheitsökonomische Evaluation,
- der Zusammenhang von Organisationsstrukturen und Abläufen mit der Technologie,
- die Diskussion der Generalisierbarkeit und Übertragbarkeit, Einschätzung ethischer, sozialer und juristischer Implikationen.

Für das deutsche Gesundheitssystem hat sich in Hinblick auf die Akzeptanz der HTA-Berichte u.a. das Vorgehen des ehemaligen Bundesausschuss Ärzte und Krankenkassen bewährt, Stellungnahmen interessierter Parteien [8] einzubeziehen. Das Institut wird im Rahmen der Bewertung von HTA-Berichten auch berücksichtigen, ob solche Stellungnahmen einbezogen wurden.

Literatur

- [1] Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomized controlled trials: the QUORUM statement. Lancet 1999; 354: 1896-1900.
- [2] Chalmers I, Haynes RB. Reporting, updating, and correcting systematic reviews of the effects of health care. In: Chalmers I, Altman DG, eds. Systematic reviews. London: BMJ Publishing Group 1995: 86-95.
- [3] Jadad AR, Cook DJ, Browman G. A guide to interpreting discordant systematic reviews. Can med Assoc J 1997; 156: 1411-1416.
- [4] LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. N Engl J Med 1997; 337: 536-542.

^f INAHTA – International Network of Agencies for health Technology Assessment

^g ECHTA – European Collaboration on Health Technology Assessment

- [5] Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995; 273: 408-412.
- [6] Perleth M. Kritische Bewertung von HTA-Berichten. In: Kunz R, Ollenschläger G, Raspe H, Jonitz G, Kolkmann FW eds. Lehrbuch Evidenzbasierte Medizin in Klinik und Praxis. Dt Ärzteverl 2000: 147-155.
- [7] Liberati A, Sheldon TA, Banta HD. EUR-Assess project Subgroup report on methodology: methodological guidance for the conduct of health technology assesement. Int J Techn Assess Health Care 1997; 13: 186-219.
- [8] Gibis B, Rheinberger P. Erfahrungen mit Health Technology Assessment im Bundesausschuss der Ärzte und Krankenkassen. Z ärztl Fortbild Quallsich 2002; 96: 82-90
- [9] Busse R, Velasco M, Perleth M, Orvain J. Best practice in undertaking and reporting HTA. ECHTA working group 4 report: 1-104.
- [10] Hailey D. Toward transparency in health technology assessment. A checklist for HTA reports. Int J Techn Assess Health Care. 2003; 19: 1-7.

4.7 *Prognose*

Wesentliche Basis für die Bewertung von Prognosestudien ist eine präzise Fragestellung, da Studien zur Evaluierung prognostischer Merkmale mit unterschiedlichen Zielen (Risikofaktorevaluierung, Scoreentwicklung / -validierung, sog. „Responder-Analysen“) durchgeführt werden, bzw. die Abgrenzung zu Diagnose- und/oder Screeningstudien schwierig sein kann, und je nach Zielsetzung unterschiedliche Bewertungsprinzipien zum Zuge kommen. Die Formulierung der Fragestellung bzw. der Ablauf, der zu einer endgültigen Fragestellung führt, folgt den im Abschnitt 2.7 niedergelegten Grundsätzen.

Ein prognostisches Merkmal liefert eine Information, die letztlich keinen Selbstzweck haben kann, sondern zu einer Konsequenz führen sollte, die ihrerseits einen prüfbaren Nutzen für den Patienten bedeutet. In diesem Zusammenhang sind an ein Prognoseverfahren ähnliche (allgemeine) Anforderungen wie an einen diagnostischen Test zu stellen (s.a. die Abschnitte 4.2.1 und 4.2.2).

Soll ein prognostisches Merkmal im Sinne eines Screenings bzw. eines Präventionsprogramms eingesetzt werden, sind bei der Beurteilung die in diesen Abschnitten (4.3 bzw. 4.12) formulierten Grundsätze zu berücksichtigen.

Speziell bedeutsam bei der Bewertung von Studien zu prognostischen Merkmalen sind folgende Aspekte^h:

- Klare Festlegung auf *ein* zu evaluierendes Merkmal (in epidemiologischen Studien typischerweise als Exposition oder Risiko- [bzw. protektiver] Faktor bezeichnet). Sollen mehrere Merkmale gleichzeitig innerhalb einer Studie untersucht werden, ist prinzipiell das Problem des multiplen Testens zu beachten (s.a. Abschnitt 3.2).
- Klare Definition des zu evaluierenden Merkmalsⁱ einschließlich der statistischen Handhabung (z.B. Dichotomisierung oder Betrachtung von Terzilen, Quartilen, etc. bei einem quantitativen Merkmal) und einer Begründung für das gewählte Vorgehen.
- Klare Festlegung und Definition von möglichen Störgrößen und Effektmodifikatoren, einschließlich deren statistischer Handhabung.
- Klare Definition des Outcome, an dem die prognostische Wertigkeit orientiert werden soll.
- Klare Formulierung einer Fragestellung und damit verbunden der Studienplanung; dies schließt eine Fallzahlplanung ein, die sich beispielsweise an der gewünschten Präzision der Schätzung (Breite des Konfidenzintervalls) orientieren kann und eine Abschätzung sowohl der Prävalenz der Exposition als auch der Inzidenz im Hinblick auf den interessierenden Outcome erfordert.
- Klare Beschreibung der Ziel- und Stichprobenpopulation (z.B. Bevölkerung, Register, Hausarztbezogen) und Begründung für deren Wahl. Klare Beschreibung für die Auswahl und das Vorgehen bei der Rekrutierung der Studienteilnehmer (Zufallsauswahl, Frage der Repräsentativität).
- Homogenität der betrachteten Population; falls es sich um eine heterogene Population handelt, sollte darauf geachtet werden, dass die prognostische Aussage möglichst konstant über die die Heterogenität verursachenden Subgruppen (beispielsweise unterschiedliche Basis-Risiken für den interessierenden Outcome) getroffen werden kann.
- Bei Kohortenstudien Vollständigkeit des follow-up bzw. Maßnahmen zur Erzielung eines möglichst vollständigen follow-up's. Abschätzung möglicher Selektionseffekte bei unvollständigem follow-up.

^h Es sei allerdings an dieser Stelle darauf hingewiesen, dass es für die Beurteilung von Prognose-Studien keine allgemein akzeptierten Qualitätskriterien gibt [2].

ⁱ Auch wenn, wie oben beschrieben, auch mehrere Merkmale gleichzeitig innerhalb einer Studie untersucht werden können, wird hier und im folgenden – soweit nicht anders ausgewiesen – der besseren Lesbarkeit willen nur die Einzahl verwendet.

- Bei der Betrachtung prognostischer Scores ist zu beachten, dass zwischen Scoreentwicklung und Scorevalidierung differenziert wird, z.B. als Entwicklung des Scores innerhalb einer sogenannten „Lernstichprobe“ und Validierung an einer „Teststichprobe“. Idealerweise erfolgen Scoreentwicklung und Scorevalidierung in unterschiedlichen Studien.

Typische Studiendesigns für die Evaluierung prognostischer Merkmale im Sinne von Risikofaktoren umfassen Fall-Kontroll- und Kohortenstudien, in Ausnahmefällen (bei der Betrachtung unveränderlicher Merkmale) auch Querschnittstudien. Die grundlegenden Prinzipien für die Bewertung solcher Studien über die oben genannten Aspekte hinaus sind im Abschnitt 3.6 formuliert.

Die methodische Qualität von Studien (bzw. deren Publikationen) zu prognostischen Merkmalen ist häufig unzureichend [1]. Insofern sind Meta-Analysen (nicht systematische Übersichten!) von Prognose-Studien oft unangemessen und deren Ergebnisse nur unter Vorbehalt zu verwerten [2]. Die Literatursuche bei der Evaluierung von prognostischen Merkmalen (im Rahmen einer systematischen Übersicht) ist schwieriger als beispielsweise bei Therapiestudien, und es existiert (noch) keine allgemein akzeptierte optimale Suchstrategie. Darüber hinaus wird angenommen, dass dieser Bereich besonders anfällig für einen Publikations-Bias ist [2].

Literatur

- [1] Pocock SJ, Collier TJ, Dandreo KJ, de Stavola BL, Goldman MB, Kalish LA, Kasten LE, McCormack VA. Issues in the reporting of epidemiological studies: a survey of recent practice. *BMJ* 2004; 329: 883-887.
- [2] Altman D. Systematic reviews of evaluations of prognostic variables. In: Egger M, Davey Smith G, Altman DG (Hrsg.). *Systematic reviews in health care: meta-analysis in context*, 2nd ed. London: BMJ Books, 2001.

4.8 Individuelle Risikoabschätzung

Neben Ergebnissen aus Studien, die einzeln oder (zumeist) mehrere prognostische Merkmale untersuchen, werden für die individuelle Risikoabschätzung von Patienten (oder allgemeiner Personen), ein unerwünschtes Ereignis zu erleiden, zunehmend sogenannte „risk charts“ (auch als „risk engines“ bezeichnet) eingesetzt, bei denen für das Zusammentreffen mehrerer Risikofaktoren multifaktorielle Schätzwerte angegeben werden (z.B. Sheffield table [1] oder „joint British chart“ [2]). Bei der Bewertung solcher Instrumente ist zu berücksichtigen,

- aus welcher Art von Studien die zugrundeliegenden Daten stammen,

- ob die für die multifaktorielle Betrachtung herangezogenen Merkmale in diesen zugrundeliegenden Studien ebenfalls gemeinsam betrachtet wurden,
- ob, und wenn ja, wie eine multifaktorielle statistische Analyse in diesen zugrundeliegenden Studien erfolgte,
- ob bei einer eventuell durchgeführten multifaktorielle Analyse in den zugrundeliegenden Studien adäquate statistische Modelle (z.B. log-lineares Modell) mit Berücksichtigung von Wechselwirkungstermen verwendet wurden, und
- ob diese Instrumente jemals in nachfolgenden Studien validiert wurden (Teststichproben).

Literatur

- [1] Wallis EJ, Ramsay LE, Ul Haq I, Ghahramani P, Jackson PR, Rowland-Yeo K, Yeo WW. Coronary and cardiovascular risk estimation for primary prevention: validation of a new Sheffield table in the 1995 Scottish health survey population. *BMJ* 2000; 320: 671-676.
- [2] British Cardiac Society, British Hyperlipidaemia Association, British Hypertension Society, endorsed by the British Diabetic Association. Joint British recommendations on prevention of coronary heart disease in clinical practice. *Heart* 1998; 80(Suppl2): S1-S29.

4.9 *Beurteilungsmethoden in Abhängigkeit von der Prävalenz und der Art der Erkrankung*

Jede generalisierende Aussage einer Studie ist mit einem Fehler behaftet und kann daher zu Fehleinschätzungen führen. Die mögliche Größe des Fehlers ist unter anderem abhängig von dem Studiendesign, der Studiendurchführung, der Güte und Vollständigkeit der Evaluation der Daten und der Qualität des Ergebnisberichts. Bei der Interpretation der Studienergebnisse und der Übertragung in die Versorgungsrealität ist daher die mögliche Größe dieses Fehlers zu berücksichtigen. Die potentielle Möglichkeit der Fehleinschätzung ist dabei in Beziehung zu setzen unter anderem zur Art und Schwere der Erkrankung, der Größe und Art des diagnostischen oder des Therapieeffektes und zur Qualität der Evidenz über bereits vorhandene diagnostische und therapeutische Verfahren ähnlicher Zielsetzung.

Bei sehr seltenen Krankheiten kann es mitunter sehr schwer oder gar unmöglich sein, genügend Patienten in eine klinische Studie einzuschließen, um mit ausreichender Power selbst moderate Effekte statistisch aufdecken zu können. Dies begründet allerdings nicht den Verzicht auf die wesentlichen methodischen Instrumente für das Design klinischer Studien wie Randomisierung, Doppelblindheit, Intention-to-treat-Prinzip, etc. Um dem Powerproblem zu begegnen, kann es bei

der Bewertung von Studien in solchen Indikationsgebieten sinnvoll sein, a priori – also in Unkenntnis der Datenlage – ein größeres Irrtumsniveau zu akzeptieren als das in der medizinischen Forschung üblicherweise verwendete Niveau von (zumeist zweiseitig) 5% (s.a. Abschnitt 3.2). Allerdings sollte darauf geachtet werden, dass das Kriterium „sehr seltene Erkrankung“ nicht durch eine inadäquate Subgruppenbildung bei eigentlich nicht sehr seltenen Krankheiten künstlich erzeugt wurde (sogenanntes „disease slicing“).

Auch bei besonders schwerwiegenden Erkrankungen, also z.B. solchen, die neu beschrieben sind und bei einem großen Anteil von Patienten innerhalb kurzer Frist zu irreversibler Morbidität oder gar zum Tod führen (z.B. AIDS in den 80er Jahren des vorigen Jahrhunderts oder SARS), kann eine Anpassung des (statistischen) Irrtumsniveaus (nach oben) in Erwägung gezogen werden.

Bei Krankheiten mit nahezu deterministischem Verlauf (z.B. Glioblastoma multiforme) können für eine positive Nutzenbewertung auch Studien niedriger Evidenzklasse (z.B. Fallserien) ausreichend sein, wenn solche Studien den für die jeweilige Studienart gültigen Qualitätsansprüchen genügen, und wenn die Ergebnisse solcher Studien einen „dramatischen“, dem natürlichen Verlauf quasi entgegengesetzten Effekt zeigen. Dies gilt selbstverständlich auch in umgekehrter Weise für die Beobachtung und Bewertung von schwerwiegenden (unerwarteten) unerwünschten Ereignissen bei Zustandsbildern mit bekanntem benignem Verlauf.

4.10 *Beurteilung von nicht erkrankungsassoziierten Beschwerden*

Unter „nicht erkrankungsassoziierten Beschwerden“ sollen Symptomkomplexe verstanden werden, die sich nicht eindeutig den gängigen, zumeist pathophysiologisch begründeten Krankheitsentitäten zuordnen lassen (z.B. sogenanntes „chronisches Müdigkeitssyndrom“). Für die Bewertung von therapeutischen Interventionen solcher Symptomkomplexe lassen sich prinzipiell die gleichen Methoden anwenden wie bei anderen Erkrankungen auch. Bei der Bewertung von diagnostischen und prognostischen Verfahren zur Erkennung solcher Symptomkomplexe ist speziell darauf zu achten, dass diese Verfahren nicht Bestandteil des Symptomkomplex sind, da dies eine Evaluierung im Hinblick auf Diskriminationsfähigkeit und Vorhersagegenauigkeit in aller Regel ausschließt.

4.11 *Beurteilung von komplementären Verfahren*

Komplementäre Verfahren gleich welcher Art (im Wesentlichen therapeutisch oder diagnostisch) lassen sich prinzipiell anhand der gleichen Methoden evaluieren und bewerten wie (natur-)

wissenschaftlich begründete Verfahren. Dies ist durch zahlreiche Beispiele belegt (z.B. Homöopathie [1], Akupunktur [2]). Ein Verweis auf die Nicht-Prüfbarkeit solcher Verfahren durch wissenschaftliche Methoden aufgrund von Spezifika wie besonderes Arzt- (bzw. Therapeuten-) Patienten-Verhältnis, besondere Krankheits- (bzw. Gesundheits-) Konstrukte, etc. ist nicht akzeptabel. In aller Regel wird es sinnvoll sein, Vertreter solcher Verfahren auf der Grundlage der Methodik des Instituts in die Protokollplanung für die Erstellung von Berichten und in den Reviewprozess von Berichten einzubeziehen.

Literatur

- [1] Jonas WB, Anderson RL, Crawford CC, Lyons JS. A systematic review of the quality of homeopathic clinical trials. BMC Complement Altern Med 2001; 1 :12.
- [2] Haake M, Muller HH, Schade-Brittinger C et al. The German multicenter, randomized, partially blinded, prospective trial of acupuncture for chronic low-back pain: a preliminary report on the rationale and design of the trial. J Altern Complement Med 2003; 9: 763-770.

4.12 Beurteilung bevölkerungsweiter Interventionen / Prävention

Eine methodische Prüfung der vorliegenden Studien zu bevölkerungsweiten Interventionen erfolgt nach den vorab genannten Kriterien (s.a. Abschnitt 4.3) zur Prüfung der einzelnen Studientypen.

Die zentrale Anforderung, die darüber hinaus an präventive Maßnahmen und Maßnahmen der Gesundheitsförderung gestellt wird, ist, dass sie wirksame Strategien beinhalten, die die Gesundheit der Bevölkerung erhalten und damit Lebensqualität, Mobilität und Leistungsfähigkeit der Bevölkerung nachhaltig verbessern.

Die Vielzahl der möglichen präventiven Maßnahmen und die Begrenztheit der Mittel machen eine Priorisierung der Maßnahmen erforderlich. Dabei sind die allgemeinen Grundsätze zur Priorisierung von Gesundheitsthemen (s.a. Abschnitt 2.13) zu berücksichtigen.

Bisherige Präventionsprogramme setzen mit Hinweis auf die Beeinflussbarkeit von Defiziten die Schwerpunkte oft im jüngeren Lebensalter. Für das höhere Lebensalter [1], die Gesundheitsförderung in Betrieben, Schulen, Einrichtungen des Gesundheitswesens oder in sozialen Brennpunkten bestehen jedoch Angebotslücken [2].

Die präventiven Maßnahmen müssen klar an spezifischen Zielgruppen ausgerichtet sein. Dem muss eine genaue Analyse der Bedürfnisse, Erwartungen und Gewohnheiten der Zielgruppe vorausgehen. Unter der Voraussetzung, dass es sich um eine sinnvolle Intervention handelt, ist zu

klären, welche Voraussetzungen zu schaffen sind, damit alle Mitglieder der Zielgruppe einen gleichberechtigten Zugang zu der Intervention haben.

Eine Nutzenbewertung hat sowohl zeitliche (mittelfristige oder langfristige Verbesserung der Gesundheit) als auch perspektivische Dimensionen (aus Sicht des Einzelnen, gesellschaftliche Aspekte, ökonomische Aspekte). Zu Beurteilung des Nutzens müssen klare Wirkungsziele der Intervention angegeben sein.

Programme, die auf langfristige Wirkungen ausgelegt sind, sollten zu einer Senkung der vorzeitigen Mortalität und/oder Morbidität und zur Verbesserung und/oder zum Erhalt der Lebensqualität führen.

Zur Beurteilung mittelfristiger Wirkungen und zur Abschätzung des Erfolgs solcher Programme dienen insbesondere folgende Parameter:

- gesundheitliche Outcome-Parameter
- gesundheitsbezogene Lebensqualität [3]
- Bewertung der Kompetenzentwicklung (z. B. Wissen, Selbstbewusstsein, Einstellung)
- Beurteilung der Zugangswege
- Zufriedenheit der Zielgruppe mit der Struktur und Durchführung des Programms
- Strukturbildung (capacity building) [4].

Um abzuschätzen, ob und unter welchen Umständen der Nutzen stabil ist, sind auch Sensitivitätsanalysen zu fordern.

Zur Bewertung des Netto-Nutzens für den Einzelnen oder die Gesellschaft, sind gesundheitsökonomische Analysen zu fordern (s.a. Abschnitt 4.4), ggf. auch unter Einbezug von Studienergebnissen aus betrieblichen oder gemeindebezogenen Projekten.

Literatur

- [1] World Health Organization. The Jakarta Declaration on Leading Health Promotion into the 21st Century. WHO 1997, Geneva.
- [2] Badura B. Gesundheitsförderung und Prävention – Schritte in eine „gesunde Gesellschaft“. Public Health Forum 2000; 28: 5-7.
- [3] Bullinger M. Erfassung der gesundheitsbezogenen Lebensqualität mit dem SF-36. Health Survey. Rehabilitation 1996; 35: 17-27.
- [4] Hawe P, Noort M, King L, Jordaens C. Multiplying health gains: the critical role of capacity-building within health promotion programs. Health policy 1997; 39: 29-42.

4.13 Qualitätsmanagement in der Versorgung

Zahlreiche Paragraphen des SGB V sowie zahlreiche Initiativen im ambulanten und im stationären Bereich zielen auf die Verbesserung *der Qualität* der gesundheitlichen Versorgung. Ebenso befassen sich viele neue Versorgungsmodelle mit einer Steigerung der Qualität und einer höheren Kosteneffizienz. Das Institut könnte in Abstimmung mit dem G-BA und anderen Gremien sowie den Ressorts „Gesundheitsökonomie“ und „Biometrie“ die Aufgabe einer Erfassung und Bewertung der verschiedenen Qualitätsaspekte und -modelle übernehmen.

4.14 Beschreibung der Art und Größe des Placeboeffektes

Die Planung, Durchführung und Auswertung einer Studie können durch schwer kontrollierbare Einflüsse einseitig beeinflusst werden und dadurch ihre Ergebnisse verändern. Diese Beeinflussung geschieht vor allem, wenn die aktiven und passiven Studienteilnehmer wissen, wer welcher Gruppe zugeordnet ist [1]. Die Verblindung einer Studie und die Verwendung einer Substanz, die keine pharmakologischen Effekte ausweist, haben dabei zum Ziel, diese Einflüsse zu minimieren. Darüber hinaus hat Placebo möglicherweise eigene Einflüsse auf medizinische Parameter, vor allem solche die einer subjektiven Einschätzung unterliegen wie Schmerz [2]. Bei der Interpretation von placebokontrollierten Studien ist es daher notwendig zu überprüfen, ob durch die Verwendung eines Placebo tatsächlich eine Verblindung gewährleistet wurde. Der Effekt von Placebo scheint mit der Größe der Studie abzunehmen [2]. Daher ist eine Fehlinterpretation von nicht-placebokontrollierten Studien vor allem bei relativ kleinen Fallzahlen zu befürchten. Bei einigen Interventionen kann eine Verblindung aber unmöglich sein, zum Beispiel bei therapeutischen Interventionen mit subjektiv eindeutig identifizierbaren erwünschten oder unerwünschten Wirkungen. Dabei ist es in einer Studie sinnvoll, zu ermitteln, wie viele Patienten und Ärzte korrekt die zugeordnete Gruppe erraten haben. Ist bei der Intervention die Größe des Placeboeffektes bekannt, kann daraus ggf. der Grad der Verzerrung der Ergebnisse durch fehlende bzw. unmögliche Verblindung ermittelt werden. Das Fehlen einer Placebogruppe kann dadurch bedingt sein, dass bereits eine wirksame Therapie für eine Störung oder Erkrankung vorliegt, und es mit der Helsinki Konvention unvereinbar wäre, eine nicht-aktive Therapie in der Kontrollgruppe durchzuführen [3]. In einem solchen Fall ist ggf. eine Verblindung der Therapie auch ohne eine Placebogruppe notwendig.

Literatur

- [1] R. Gillespie. Manufacturing knowledge: a history of the Hawthorne experiments. Cambridge, England: Cambridge University Press, 1991.

- [2] A. Hróbjartsson and P. C. Gøtzsche. Is the placebo powerless? N Engl J Med 2001; 344: 1594-1602.
- [3] E. J. Emanuel and F G. Miller. The ethics of placebo-controlled trials. N Engl J Med 2001; 345: 915-919.

5. Patient Information and Research

5.1 Goal

The Institute aims to become an effective, reliable, trusted and popular provider of information and education for the German community.

A key goal of the Institute is the production of scientific health information for the community, including the development of a health information web site.

It is not intended that the Institute becomes a provider of patient advice. As an independent scientific provider of health information, the Institute's intention will be to enhance understanding and informed choice, rather than aim to bring about particular behaviours [1].

5.2 Information products and contents

The Institute's first medium for publication of scientific health information for the public will be a web site. This can include both electronic and downloadable documents and programs for installing on computers offline, CD-ROMs and/or for printing.

There is evidence that internet-based [2,3] and offline computer-based [4] information interventions can potentially affect people's knowledge, choices, health and emotional wellbeing. However, information and education interventions can also be ineffective or harmful, and some techniques are more effective than others [3,5-8]. The Institute will seek to maintain and continually improve its standards in effective communication by:

- Keeping up-to-date with the scientific literature on effective communication;
- Working with partners to further develop scientific knowledge about effective communication in the German context;
- Maintaining high quality assurance and professional development mechanisms;
- Internal and external evaluation.

The Institute's web site for the community will aim to meet a variety of information needs [5,9] by providing:

- Information in layers with simple and short information at the top and deeper levels of more complex information below [10], including a short, free-standing overall summary;

- Written, visual and interactive health education supplementary tools, including interactive online dictionaries of medical and scientific terms used in the web site;
- Educational items to enhance understanding of scientific concepts and information in health.

In addition to summaries of other work produced by the Institute, the Patient Information web site will include:

- A.** Comprehensive health information reports in modular form (free-standing segments suitable for multiple use) produced internally by the Institute's Department of Patient Information, which will ultimately build into a reliable patient library;
- B.** Summaries directly drawn from individual systematic reviews of evidence, or small groups of systematic reviews, which are of high interest and can be adequately communicated without full health education, or relying on linkages to existing health education items within the Institute's web site.

5.3 *Priority-setting and topic selection*

As well as producing easily understandable versions of its products for use by patients, the Institute is able to generate patient information on topics it selects independently. Its comprehensive patient information reports will be modular (in free-standing segments for multiple use), and will ultimately form a reliable patient library.

However, the reports will necessarily be limited in number initially, as each will involve significant resource investment. As this will be a relatively scarce public resource, a rigorous method of priority-setting will be necessary to ensure that the Institute:

- Allocates resources in patient information by fair methods supported by well-researched material to inform decision making [11-13];
- Is transparent about the basis for its selection decisions;
- Produces a demonstrably balanced and interesting selection of topics;
- Maximises the educational, social and health benefits achievable with its patient health information.

A search of priority-setting mechanisms did not identify a specific and suitable model for the selection of evidence-based patient information topics. The system was developed based on the Institute's priority-setting rationale and values.

The priority-setting process for comprehensive patient information reports will involve:

- Background data collection from a variety of resources, including published literature;
- Topic generation and screening according to established criteria, undertaken by the Patient Information Department in conjunction with the Steering Committee (and perhaps the Scientific Board);
- Research on 'short' list of eligible topics, involving interested members within the Institute;
- Selection of topics after a rating process open to everyone in the Institute, with final decision made by the Steering Committee after considering the results of the rating process and advice from the Patient Information Department;
- Monitoring of system use and internal and external feedback, and review of the model and criteria.

5.4 *General editorial and management mechanisms*

The Patient Information Department, working with the Institute's Director and the Departments with a content interest in specific items of information, will ensure that the content of the Institute's Patient Information web site and other Patient Information products are:

- Scientifically based and kept up-to-date as significant developments in scientific knowledge occur;
- Centred on the information and psychological needs of patients as far as is possible;
- Internally consistent, and consistent with other information being published by the Institute.

The Patient Information Department Leader will be editor of the Patient Information web site. Final approval to publish will be the responsibility of the Steering Committee.

Internal and external concerns about contents or quality of the web site will be referred to the editor, who will be responsible for initiating action and reporting to the Steering Committee regularly. Major and urgent problems will be reported immediately to the Institute's Director. An acting editor will be appointed whenever the editor will be unavailable for more than a day.

Liaison between the editor, the Institute's Director and the Public Relations Department, and all affected Steering Committee members will be needed to ensure that non-web site and web site communications to patients and the community are consistent.

5.5 *Evidence basis for the Institute's production processes*

There are many guidelines and quality rating instruments that purport to rate and/or ensure the quality of patient information and/or internet health information for patients. Reviews of such instruments have concluded that there is no reliable evidence of validity for these instruments [14-17]. It is also unclear whether such tools in fact measure what they claim to measure.

No guide or instrument for the production of patient information has been shown to be a reliable indicator of quality of health information [15,17]. None is supported by strong evidence for each of the components recommended, or evidence to show that the recommended methods for producing patient health information are cost-effective. Further, information from research with patients, including patients in Germany [18,19] often indicates that some issues suggested by instruments to be important to patients, may not be so important in practice to patients. Some recommendations common in guidelines and instruments may actually reduce the scientific quality and reliability of information.

Most existing instruments focus on communicating evidence about the effects of treatments, and do not sufficiently address information on the full range of information, including etiology, prognosis and screening and diagnostic tests. Yet, patient information on diagnosis and screening involves more complex decision making and communication issues than information on treatment [6,7,20].

Therefore, the Institute will not rely on any currently available instrument, nor will it develop one. Rather, the Institute will rely on evidence related to specific aspects that may affect people's decision making, and demonstrably affect the quality of the information produced by the Institute. The Patient Information Department will keep up-to-date with the evidence so that we can adapt our methods in accordance with emerging evidence, as well as testing the Institute's Patient Information with representative patients locally.

5.6 *Research and development of comprehensive patient information reports*

5.6.1 Contents

The following items will be considered for each comprehensive Patient Information report:

- Education about the disease or condition, including anatomy, physiology, disease etiology, pathology, prognosis, possible complications, recovery, recurrence, and who may be affected by the problem;
- How to identify symptoms of disease, complications and recurrence;
- Options for screening and prevention (both primary and secondary, and including lifestyle and behaviours affecting health);
- Self-management;
- Psychosocial issues and living with the disease, for both the person with the disease and their family, including both impact on the family and how friends and family may be able to help;
- Interventions for diagnosis and treatment, including risks, benefits and costs, and enough detail to enable people to use (or avoid) the treatment if they wish;
- Horizon scanning of new options potentially available in the near future, and information about participating in trials where particularly relevant.

5.6.2 Establishing the scope of the report

A detailed scoping exercise will determine the range of topics to be covered in the report, as well as patient-relevant outcomes to be considered. A range of methods can be used to help identify questions which the report needs to address. These vary in their cost and practicality, as well as in their validity and the generalisability of results [21,22]. The Patient Information Department will rely on data identified from well-conducted representative studies whenever these are available, supplemented by telephone interviews with key informants. Focus groups may be a suitable method on some occasions [23]. The Department will retain a focus on ensuring that decisions are made with a concern for remaining representative of the general community, with an additional concern for equity and considering the needs of more disadvantaged members of the community.

There will be two major steps involved in determining the questions and scope of the reports:

- (a) Identifying what interventions and advice are being offered to patients in Germany and elsewhere currently, through telephone interviews with key informants (clinical experts, patient groups and health educators), literature search, and review of key internet sites in the disease area;
- (b) Attempting to identify what patients currently know and what they might want to know, through literature searches, web sites such as the database of patients' experiences [24], brief telephone interviews with key informants (representatives of patient groups or educators), and focus groups when more in-depth discussion is needed and more reliable data sources are lacking.

5.6.3 Searching and assessment of evidence

The same methods and standards used for the scientific work throughout the Institute will be applied in the search for, and assessment of, evidence to be used in the Patient Information reports. Literature searches will include well-conducted surveys and social science research conducted with German patients and people from similar communities.

Patient Information reports will usually rely on completed systematic reviews, rather than undertaking analysis of primary studies. So-called 'mega trials' and other landmark trials will sometimes be relied on, in consultation with the content Department responsible for the intervention involved. The Patient Information Department may also at times seek assistance with topics where trials but no adequate systematic reviews currently exist.

Quality of systematic reviews and other intervention studies to be relied on in Patient Information reports will be assessed in consultation with the relevant Departments in the Institute.

5.6.4 Preparing visual and supplementary interactive materials

Visual and graphic materials which meet the Department's communication standards and health education needs will either be purchased, or where necessary commissioned.

Interactive supplementary materials (such as online dietary calculators) may be prepared for reports, and the same standards for quality of data that are used throughout the Institute will be used to ensure the validity of interactive materials.

5.6.5 Drafting and review process

The Patient Information Department will work closely with the content-specific Departments of the Institute to ensure the accuracy and scientific quality of the information it prepares and coordinates for the community. A reliance on internal expertise, the expertise of key informants in each topic,

and the opportunity for the authors of systematic reviews to comment on the Institute's interpretation of their analyses [25] will be key to the research and quality processes of the Department.

The evidence for most commonly used techniques and systems of peer review is weak [26], including consumer or patient review [27]. The processes proposed for pre- and post-publication quality assurance of Patient Information content reflect current international scientific publishing standards. The Department will remain up-to-date with the growing body of scientific knowledge about the effectiveness of different methods and components of peer review.

The production steps for comprehensive Patient Information reports will involve:

- Multiple revisions of draft material, after internal review by the Institute's Director and Department(s) with relevant content interest. Any conflicting feedback or opinions will be addressed by discussion and further research if necessary.
- Peer review by clinical experts and patient/educator experts. Meeting with patient section of the Institute's Board of Trustees [28]. Lead authors of major systematic reviews used in the report will be sent a copy of a draft and offered the opportunity to comment (usually on the English version).
- Where necessary, readability/comprehensibility tests and/or a focus group may be held (in German). Feedback and major changes will be discussed with the relevant Institute Department(s) and the Institute's Director.
- The 'final draft', in both German and English, will be sent to the Institute's Steering Committee. The final draft (including visual and interactive materials) will be marked up for the Internet. Online usability testing (German only) will be done with 3-5 volunteers [29-31], including at least one 'patient' or patient representative. After this testing is complete, the English version will also be marked up for the Internet and any versions being prepared for downloading or alternate publication method will be finalised.
- When multiple products are to be produced from the source material of the report, this will be prepared and sent through a similar process if the content is altered in important ways (for example, content removed, added or altered in potentially ways).
- The final version for publication cleared by the Patient Information Department will be discussed with the Institute's Director, who will have the responsibility of clearing for publication or requesting additional work or consultation.

Plans for publicity and dissemination of Patient Information materials will be discussed initially with the Institute's Director and Public Relations Department. Proposals will be reported to the

Steering Committee. The Institute's Director, Patient Information Department Leader and Institute Spokeswoman will ensure that non-web site communications about the Patient Information are consistent in content and approach.

5.6.6 Updating and review

- Topics, interventions and reviews will be coded, enabling monitoring for changes in Cochrane reviews and the release of major new evidence which may mean that the Institute's information for patients has become out-of-date. There will also be ongoing discussions between key personnel at the Cochrane Collaboration, the Centre for Reviews and Dissemination and *Evidence Based Medicine*.
- The Patient Information web site will include a mechanism for feedback on Institute publications for patients. Feedback from this, as well as internal feedback and external feedback received by other means, may also result in urgent revisions or scheduled revision.
- Major patient information modules which have not been updated within two years will be reviewed, with a new search undertaken, to ensure that no module on the Institute's Patient Information web site is substantially out of date.

5.7 Short summaries of systematic reviews and key 'mega trials'

In addition to comprehensive Patient Information reports, a much larger range of brief summaries will be produced on subjects which do not require intensive health education to accompany discussion of the issue, or which can be linked to existing explanatory items on the Patient Information web site.

The audience for this larger range of information is envisaged, including doctors. These summaries will both enhance the Institute's Patient Information web site, and provide an additional opportunity for the Institute to communicate evidence-based health information to a broad audience in health in Germany. It will broaden access to information currently only available in English. There will be several hundred such summaries in the medium term. The summary system will also be set up with automatic internet syndication, enabling this information to be used by web sites more widely, and offering the Institute an additional means of communicating the results of evidence to the community and health care system.

These summaries will be:

- Single short items summarising a single systematic review or so-called 'mega trial' from the same priority sources as full Patient Information reports, and some single items in turn summarising several such summaries included on web site;
- Based on a set of guidelines to be developed specifically for the production of these summaries;
- Either produced fully by the Patient Information Department in consultation with the Institute's Director and relevant content Department(s), or by content Departments in the Institute, with editorial oversight on content and communication standards from the Patient Information Department;
- Reviewed internally only, with the authors of the reviews and trials offered the opportunity to comment on the Institute's interpretation of their research.

As these summaries will be quite large in number, an abbreviated version of the priority-setting process will apply. The main goal of priority-setting for items to be included will be the need to maintain balance in the overall provision of information by the Institute.

The Patient Information Department will have editorial responsibility for these summaries, with the Institute's Director having final publication approval. The Patient Information will have responsibility for monitoring and acting on the need for updating, in response to changes in the evidence or feedback.

5.8 *Communication standards*

Throughout its communication of consumer and patient information, the Institute's communication goals are to:

- Communicate respectfully and effectively with the German community, and earn trust as a reliable and popular source of information;
- Make its Patient Information easy and enjoyable to read without sacrificing scientific accuracy;
- Maintain as neutral a communication style as possible and minimise ambiguity;
- Demonstrate sensitivity and respect for patient knowledge, values and concerns, patient autonomy, and cultural differences;
- Support patient empowerment and growth of health and scientific literacy;

- Help people relate the evidence to their own personal situation;
- Respect readers' time.

The Institute will demonstrate its values through the language and visual representations it uses, and through the professionalism and quality of presentation of its Patient Information. It will develop a house style guide for each type of information it produces (very easy to read summaries of reports, more detailed full reports, and the short summaries of research information aimed at both patients and doctors). These guidelines will be based as far as possible on evidence of effectiveness of communication techniques, and embodiment of the Institute's values.

There are particular communication challenges in providing health information without exaggerating what is scientifically known and not telling people what they 'should' do. The effects of making people aware of the scientific uncertainty of much health care is largely unknown [6]. Further, this community is accustomed to more directive styles of patient information, often aimed at changing their opinions and/or behaviour. The Institute will need to evaluate its materials to develop a house style and communication techniques that can be effective and popular in Germany.

The Institute will aim to present information in a variety of ways, to enable as many people as possible to gain a clear perspective on information provided [7]. Ordinarily, the aim of remaining neutral will mean avoiding the use of relative risk information. However, on occasion the inclusion of relative risk data may be necessary to enable people to compare treatments. Relative risk information will not be presented on its own, without absolute risk or number needed to treat/harm included for perspective.

There is evidence that personalised or individualised risk estimates are effective forms of communication [7]. The Institute may in some circumstances develop guides or tools to help patients estimate their personal risk, but only where very reliable data exist to inform the development of these instruments.

Decision aids have been shown to be an effective means of communicating health information and enabling patient choice [32]. The Institute may develop some decision aids, and will seek to incorporate effective elements of decision aids in its patient information.

In addition, the Institute will:

- Present patient information and data in particular in consistent formats;
- Explain the degree of uncertainty associated with data;
- Indicate to whom the evidence applies;
- Aim to achieve the highest possible standards in web usability and web navigation;

- Be very careful to make the difference between 'absence of evidence' and 'evidence of no effect' very clear;
- Aim not to bias information towards the products of any particular company, but use generic names for products, supported by brand names of products available where relevant;
- Support the understandability of all health information on its web site by the inclusion of an online dictionary built alongside the information and integrated into it by hypertext.

5.9 *Translation*

The Institute will aim to publish simultaneously in German and English, and keep both versions up-to-date. In recognition of the difficulties of grading translation quality, no specific standard of assessment of quality of translations is possible.

Translation can be literal (precisely word for word), or aiming to capture and convey the intent of the original. The Institute's translations will not necessarily be literal word-for-word translations.

Computer-assisted translation involves using a set of computer technology applications to help improve and standardise human translation. This includes tools which are used to customise and build terminology and phrases. While machine translation is a method where a human assists the machine, computer-assisted translation is a method where the computer assists the human translator.

There is no strong evidence to demonstrate that any particular method of checking translation quality is more effective or cost-effective than another. The Institute will initially rely on in-house translation by bilingual staff with content knowledge in the Patient Information Department, supported by computer-assisted translation tools. Working in both German and English, the Institute may benefit from feedback from bilingual researchers and peer reviewers of its material. There will be sample checks of quality of translations, and the development of in-house language skills of staff of the Patient Information Department will be supported.

References

- [1] Hope T. Evidence-Based Patient Choice. London: King's Fund, 1996.
- [2] Bessell TL et al. Do internet interventions for consumers cause more harm than good? A systematic review. Health Expectations 2002; 5: 28-37.

- [3] National Institute of Clinical Studies. The impact of the internet on consumers' health behaviour. Prepared by the Centre for General Practice and the Centre for Evidence Based Practice, University of Queensland. NICS, Melbourne: 2003.
- [4] Lewis D. Computer-based approaches to patient education: a review of the literature. *J Am Informatics Association* 1999; 6: 272-282.
- [5] Coulter A, Entwistle V and Gilbert D. *Informing Patients: An Assessment of the Quality of Patient Information Materials*. London: King's Fund Publishing, 1998.
- [6] Entwistle VA et al. Developing information materials to present the findings of technology assessments to consumers: The experience of the NHS Centre for Reviews and Dissemination. *Int J Tech Assess Health Care* 1998; 14: 47-70.
- [7] Edwards A, Bastian H. Risk communication – making evidence part of patient choices? In Edwards A, Elwyn G (eds). *Evidence-Based Patient Choice: Inevitable or Impossible?* Oxford: Oxford University Press, 2001; 144-160.
- [8] Eysenbach G, Jadad AR. Consumer health informatics in the Internet age. In Edwards A, Elwyn G (eds). *Evidence-Based Patient Choice: Inevitable or Impossible?* Oxford: Oxford University Press, 2001; 289-307.
- [9] Entwistle V, Watt IS, Herring JE. Information content. *Information About Health Care Effectiveness: An introduction for consumer health information providers*. London: King's Fund. 1996: 49-50.
- [10] Commission of the European Communities. eEurope 2002: Quality criteria for health related websites. *J Med Internet Research* 2002; 4: e15.
- [11] Hensall C et al. Priority setting for health technology assessment. Theoretical considerations and practical approaches. A paper produced for the EUR-ASSESS project. *Int J Technol Assess Health Care* 1997; 13: 144-185.
- [12] Townsend J, Buxton M, Harper G. Prioritisation of health technology assessment. The PATHS model: methods and case studies. *Health Technology Assessment* 2003; Vol 7: No 20. NHS R&D HTA Programme.
- [13] Ghaffar A, de Francisco A, Matlin S. *The Combined Approach Matrix: A priority-setting tool for health research*. Geneva: Global Forum for Health Research, June 2004.
- [14] Jadad AR, Gagliardi A. Rating health information on the internet: navigating to knowledge or to Babel? *JAMA* 1998; 279: 611-614.
- [15] Eysenbach G. Consumer health informatics. *BMJ* 2000; 320: 1713-1716.
- [16] Gagliardi A, Jadad AR. Examination of instruments used to rate quality of health information on the internet: chronicle of a voyage with an unclear destination. *BMJ* 2002; 324: 569-573.
- [17] Moulton B, Franck LS, Brady H. Ensuring quality information for patients: development and preliminary validation of a new instrument to improve the quality of written health information. *Health Expectations* 2004; 7: 165-175.
- [18] van den Brink-Muinen et al. Doctor-patient communication in different European health care systems: Relevance and performance from the patients' perspective. *Patient Education Counseling* 2000; 39: 115-127.
- [19] Eysenbach G, Kohler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *BMJ* 2002; 324: 573-577.

- [20] Edwards A, Unigwe S, Elwyn G, Hood K. Personalised risk communication for informed decision making about entering screening programs (Cochrane Review). In: The Cochrane Library, Issue 3, 2004. Chichester, UK: John Wiley & Sons, Ltd.
- [21] Liberati A, Sheldon TA, Banta HD et al. Eur-Assess project subgroup report on methodology: Methodological guidance for the conduct of health technology assessment. *Int J Tech Assess Health Care* 1997; 13: 186-219.
- [22] Sixma HJ et al. Quality of care from the patients' perspective: from theoretical concept to a new measuring instrument. *Health Expectations* 1998; 1: 82-95.
- [23] DiPEX. The Database of Patients' Experiences. <http://www.dipex.org>. Zugriff am 18. Oktober 2004.
- [24] Centre for Reviews and Dissemination. How is the quality of DARE assured? York: Centre for Reviews and Dissemination. <http://www.york.ac.uk/inst/crd/faq3.htm>. Zugriff am 21. Oktober 2004.
- [25] Godlee F, Jefferson T (eds). *Peer Review in Health Sciences* (2nd ed). London: BMJ Books, 2003; 248-262.
- [26] Bastian H. Non-peer review: consumer involvement in research review. In Godlee F, Jefferson T (eds). *Peer Review in Health Sciences* (2nd ed). London: BMJ Books, 2003; 248-262.
- [27] Satzung der Stiftung für Qualität und Wirtschaftlichkeit im Gesundheitswesen. <http://www.g-ba.de/pdf/Satzung-Institut.pdf>. Zugriff am 25.10.2004.
- [28] Krug S. *Don't make me think: A common sense approach to web usability*. Indiana: New Riders, 2000.
- [29] Nielsen J. *Designing web usability*. Indiana: New Riders, 2000.
- [30] Inan H. *Measuring the success of your website: A customer-centric approach to website management*. Sydney: Pearson Educational Australia, 2002.
- [31] O'Connor AM et al. Decision aids for people facing health treatment or screening decisions (Cochrane Review) (last updated October 2003). In: The Cochrane Library, Issue 3, 2004. Chichester, UK: John Wiley & Sons, Ltd.