

# EMPIRISCHE GRENZEN DER ERKENNBARKEIT VON KAUSALZUSAMMENHÄNGEN DURCH EPIDEMIOLOGISCHE UNTERSUCHUNGEN

K. Überla

Institut für Med. Informationsverarbeitung, Biometrie und Epidemiologie der Ludwig-Maximilians-  
Universität München

Man erwartet vom Epidemiologen in vielen Bereich Kausalaussagen über sehr kleine Risiken. Überzeugende Paradigmen für den Nachweis von Kausalbeziehungen fehlen aber. Wir nähern uns in weiten Bereichen der Risikoabschätzung - Arzneimittel, Müllverbrennung, Passivrauchen, Umweltschadstoffe - einer zunehmenden Konfusion.

Die Epidemiologie hat sich seit jeher um Kausalaussagen bemüht. Bei großen Risiken ist dies einfach. Bei sehr kleinen Risiken werden die grundsätzlichen Grenzen des Wißbaren erreicht oder überschritten. Wie in allen Wissenschaften gibt es in der Epidemiologie solche Grenzen der Erkennbarkeit und des sicher Wißbaren. Sie aufzufinden und zu formulieren, ist eine wichtige Aufgabe.

Der Begriff der Kausalität ist eine Kürzel für klare und überschaubare Verhältnisse. In der heilen Welt der Wissenschaft reduziert er die Dinge auf ein einfaches Schema: Wenn - Dann. Dort, wo die Verhältnisse einfach sind, wo man den Rest der Welt bis auf eine Ursache vernachlässigen kann, ist der Kausalbegriff enorm nützlich. Kausalität bedeutet im Kern Monokausalität. Alle anderen Ursachen sind zu vernachlässigen, mehr oder weniger, bis auf diese eine. In der realen biologischen Welt ist diese Monokausalität eine Seltenheit. Viele Faktoren sind die Voraussetzungen für Krankheit und Tod. Die Genetik, die Exposition, die Vorgeschichte und die biologische Erfahrung eines Individuums sind jedes für sich notwendige, aber allein nicht hinreichende Bedingungen. Entfernt man einen Stein aus dem Mosaik, bricht die ganze Kausalkette zusammen. Multikausalität löst den Begriff der Kausalität im Kern bereits auf. Welcher Tropfen war es eigentlich, der das Glas der Verursachung zum Überlaufen brachte? Unterscheidet er sich in der Wirkung von anderen, die vorher waren, ist er wichtiger als der Tropfen vorher?

Der Wissenschaftler hat ein starkes Kausalbedürfnis. Die Suche nach der Ursache ist das übergreifende Leitmotiv. Selbst dort, wo keine Ursache mehr vorhanden ist, wo das Rauschen des Zufalls alles erklärt, wird eine Ursache gedacht, ein Kausalmosaik als Erklärung herangezogen, eine Statistik mißbraucht. In der Epidemiologie ist die Abgrenzung vom Rauschen des Zufalls schwerer als in anderen Wissenschaften. Kausalität kann in der Epidemiologie und der Statistik nur über das Experiment oder über den Ausschluß aller anderen Faktoren etabliert werden. Letzteres ist ein Sisyphusweg.

Verfahren zur Abgrenzung sicheren Kausalwissens sind:

- 1) Randomisierung verbunden mit dem Signifikanztest in kontrollierten klinischen Studien. Sie führt am nächsten an eine Kausalaussage heran. In vielen Situationen ist dies aber unmöglich, aus mancherlei Gründen.
- 2) Statistische Assoziation und Signifikanz ohne Randomisierung. Dies hat nichts mit einer Kausalaussage zu tun. Es ist ein theoretisches Konzept.
- 3) Prüfung allgemeiner Kriterien als Voraussetzung für die Annahme von Kausalität. Diese Kriterien stammen ursprünglich von Bradford-Hill (7) und wurden später erweitert. Sie werden neuerdings von einigen Epidemiologen kritisiert (2,3), die vorschlagen, daß Kausalschlüsse kein Gegenstand der Wissenschaft sein sollen.
- 4) Metaanalysen. Sie sind ein fraglicher Weg aus folgenden Gründen: a) Ein einheitliches Risikomaß ist umso unsinniger, je heterogener das Material ist. Man kann nicht Äpfel und Birnen zusammenzählen. Bei kontrollierten klinischen Studien mit denselben Therapien und Zielkriterien sind sie indiziert, für epidemiologische Risikoschätzungen aus heterogenen Studien fallen sie aber oft aus. b) Wenn ein systematischer Bias in allen Studien vorhanden ist, hat auch das Ergebnis der Metaanalyse diesen Bias. c) Wenn die Studien minimalen Qualitätsstandards nicht genügen, können Metaanalysen nur irreführende technisch signifikante Ergebnisse liefern. Falsch plus falsch ist nicht gleich richtig. d) Metaanalysen liefern kein eindeutiges Ergebnis. Wenn man nach der Qualität der einbezogenen Studien variiert, nach Einschlusskriterien oder anderen Gesichtspunkten, erhält man ein ganzes Spektrum möglicher Ergebnisse. Man kann Sensitivitätsanalysen durchführen, die nicht immer leicht zu interpretieren sind. e) Metaanalysen fügen keine einzige neue Beobachtung hinzu, sie eröffnen aber eine neue Dimension der Manipulation. Die Grenzen ihres Einsatzes sind also durchaus fraglich.
- 5) Konsensuskonferenzen. Sie lassen sich wohl kaum so formalisieren, daß sie als sicherer Weg zur Abgrenzung von Kausalwissen dienen können. Ihre Ergebnisse hängen z. B. von der Auswahl der Teilnehmer ab.
- 6) Methodisch orientierte Review-Artikel. Einer sorgfältigen methodischen Autopsie jeder einzelnen Studie folgt die Auswahl der aussagekräftigsten Studie im Sinn eines "best evidence synthesis" -Ansatzes (4). Ein fachliches Urteil des Biometrikers ist dabei nötig, nicht nur sein methodisches Wissen. Ein solches Vorgehen kann durchaus sinnvoll sein.

Insgesamt: Es gibt keine formale Methode, um in der Epidemiologie sicheres Kausalwissen vom Zufall, Bias und Confounding eindeutig abzugrenzen. Die Grenzen der Erkennbarkeit sind unzureichend spezifiziert. Der gesunde Menschenverstand bleibt als bescheidener Ausweg. Wie könnte man die Situation verbessern? Dazu möchte ich 4 Bausteine vorschlagen:

1. Die Konstruktion von Kausalindizes
2. Die standardisierte Angabe von Risiken in einem Risk-Pictogramm.
3. Eine Basisinzidenz von  $10^{-5}$  bis  $10^{-6}$  als empirische Grenze der sicheren Erkennbarkeit.
4. Odds-ratios größer als 2-3 als empirische Grenze der sicheren Erkennbarkeit.

## 1. Die Konstruktion von Kausalindizes

In der Tabelle 1 sind 10 bekannte Kriterien für Kausalität wiedergegeben. Sie wurden so weit als möglich quantifiziert: jede Frage wird bei positiver Antwort mit 1 bewertet, bei negativer mit 0. Der maximale Kausalindex ist 10, der minimale 0. Bei Aktivrauchen und Lungenkrebs erhält man z.B. 10, bei Passivrauchen und Lungenkrebs 2 oder 3.

**Tab. 1: Konstruktion von Kausalindizes**

Kriterium	Kausalindex	Teilindex A	Teilindex B
Sichere Konsistenz und Wiederholbarkeit	1	1	-
Expositionsmaß reliabel und valide (r > 0.7)	1	-	1
Outcome reliabel und valide (r < 0.7)	1	-	1
Wichtigste Faktoren für Bias hinreichend ausgeschlossen	1	-	1
Wichtigste Confounding-Faktoren ausgeschlossen	1	-	1
Statistische Signifikanz (p < 0.05 doppels.)	1	1	-
Stärke der Assoziation (RR > 2 - 3)	1	1	-
Dosis - Wirkungsbeziehung (p < 0.05 doppels.)	1	1	-
Erfolg einer Intervention nachgewiesen	1	-	-
Biologische Plausibilität gegeben	1	-	-
Maximaler Index	10	4	4

### Bewertung des Kausalindex:

- ≤ 5: Kein Hinweis auf kausalen Zusammenhang;      6: Schwacher Hinweis auf kausalen Zusammenhang  
 7: Begrenzter Hinweis auf kausalen Zusammenhang    8: Wahrscheinlicher Kausalzusammenhang  
 9, 10: Kausalzusammenhang so gut wie sicher

In den beiden rechten Spalten wurden die ersten 8 Fragen zu Teilindizes zusammengefaßt. Wenn ein Interventionserfolg nachgewiesen ist, wird man immer einen Kausalzusammenhang annehmen können. Die biologische Plausibilität läßt sich schwer quantifizieren. Beide Fragen wurden in den Teilindizes weggelassen. Der Teilindex A enthält die wichtigen Dinge, gewissermaßen die Basis: die sichere Wiederholbarkeit, Signifikanz, Stärke der Assoziation und Dosis-Wirkungsbeziehung. Der Teilindex B spezifiziert die eher technischen Aspekte: Reliabilität und Validität von Expositionsmaßen und Outcome, sowie Ausschluß von Bias und Confounding. Beide Scores lassen sich zu einer einheitlichen Bewertung zusammenfassen, wie die Tabelle 2 zeigt. Die Bewertung wurde so gewählt, daß die gleichen Gesamtaussagen wie in Tab. 1 resultieren.

Tab.2: Mögliche Bewertung von Teilindex A und Teilindex B

		Teilindex A				
		0	1	2	3	4
Teilindex B	0	-	-	-	-	-
	1	-	-	-	-	+
	2	-	-	-	+	++
	3	-	-	+	++	+++
	4	-	+	++	+++	++++

-	Kein Hinweis auf kausalen Zusammenhang
+	Schwacher Hinweis auf kausalen Zusammenhang
++	Begrenzter Hinweis auf kausaler Zusammenhang
+++	Wahrscheinlicher Kausalzusammenhang
++++	Kausalzusammenhang so gut wie sicher

Natürlich ist die Beurteilung der 10 Kriterien nicht immer eindeutig möglich. Man könnte auch andere Fragen, Gewichte oder Kombinationen erproben. Entscheidend ist, daß man einen konsistenten Bewertungsalgorithmus verfolgt, der über verschiedene Fragestellungen hinweg konstant ist. In der Intelligenzmessung und als prognostische Scores haben sich derartige Ansätze bewährt. Es sind verfeinerte Konzepte denkbar, deren Ergebnisse miteinander korrelieren werden. Ich schlage vor, solche Kausal-Scores empirisch zu erproben und weiterzuentwickeln.

## 2. Die standardisierte Angabe von Risiken in einem Risk-Pictogramm

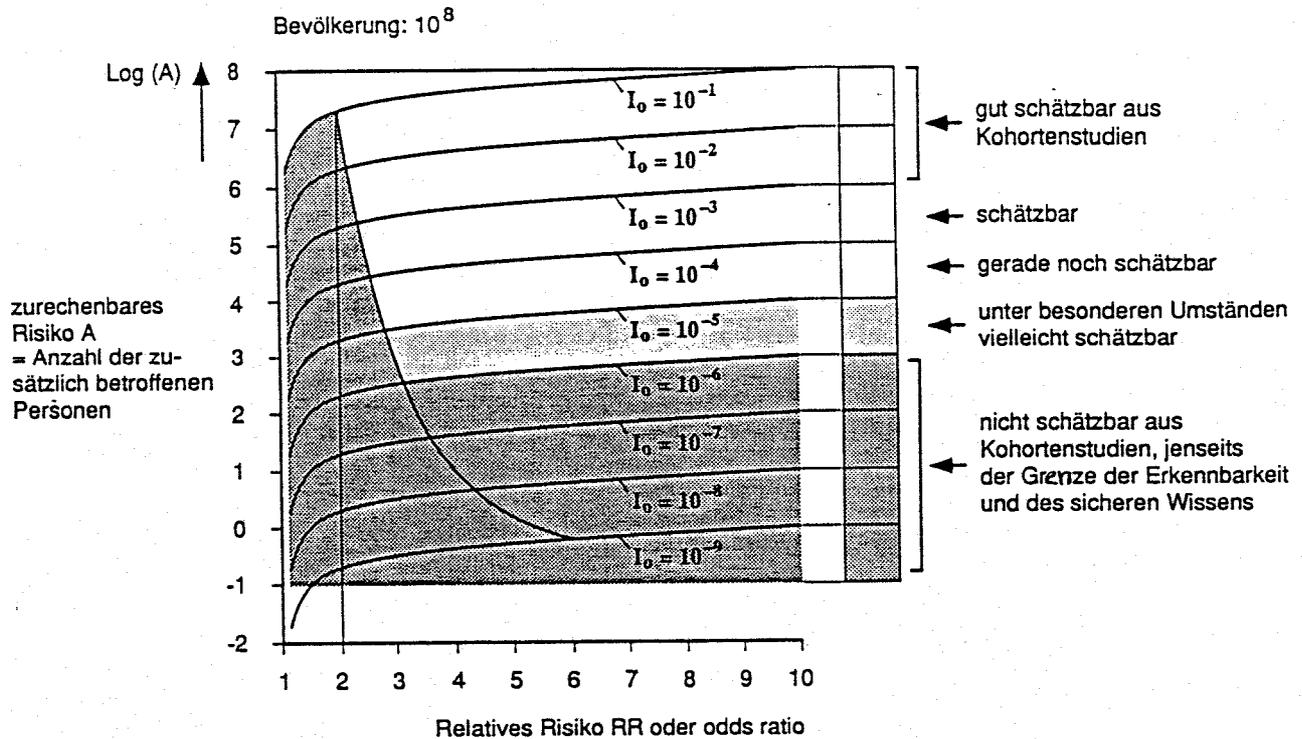
In der Epidemiologie werden 4 Risikomaße verwendet: die Basisinzidenz  $I_0$  in der Vergleichsgruppe oder Population, die Inzidenz  $I_1$  in der exponierten Gruppe, das relative Risiko als Verhältnis der beiden Inzidenzen und das attributive Risikomaß als Differenz der beiden Inzidenzen. Das entscheidende Risikomaß ist das attributive Risiko, d.h. die Zahl der zusätzlich durch eine bestimmte Exposition betroffenen Personen. Das relative Risiko allein ist irreführend, es enthält nur einen kleinen Teil der Information. Die 4 Risikomaße sind durch ihre Definition verknüpft. Alle Risiko-Informationen lassen sich in einer Darstellung einheitlich und standardisiert zusammenfassen, die ich Risk-Pictogramm genannt habe (5,6).

Nach rechts ist in der Abb. 1 das relative Risiko oder die odds ratio abgetragen, nach oben das zurechenbare Risiko als Logarithmus der Zahl der zusätzlich betroffenen Personen. Als Bezug ist eine Population von 100 Millionen Personen gewählt, die sich leicht umrechnen läßt. Die Kurven geben die rechnerische Beziehung zwischen Basisinzidenz, relativem Risiko und attributivem Risiko wieder. Bei einem relativen Risiko von 3 und einer Basisinzidenz von  $10^{-5}$  wären ungefähr  $10^3$ , d.h. 1 000 Personen in einer Population von 100 Millionen Exponierten zusätzlich betroffen, bei 1 Million Exponierten wären es noch 10.

Wenn man Basisinzidenz und relatives Risiko kennt, kann man das attributive Risiko unmittelbar ablesen. Punkte für individuelle Risiken kann man einzeichnen. Die Kurven sind rechnerische Konsequenzen der Definitionen. Man kann um sie Konfidenzintervalle bilden. Man erkennt unmittelbar, daß die Änderung des relativen Risikos für das attributive Risiko weniger wichtig ist als die Änderung der Basisinzidenz.

## Abb. 1: Risk - Pictogramm

Relatives Risiko RR, Inzidenz  $I_0$  und ihre Beziehung zum zurechenbaren Risiko A



### 3. Einen Basisinzidenz von $10^{-5}$ bis $10^{-6}$ als empirische Grenze der sicheren Erkennbarkeit

Auf der rechten Seite der Abb 1 ist angegeben, in welchen Bereichen man die Basisinzidenzen aus Kohortenstudien abschätzen kann. Bis zu  $10^{-3}$  können wir Inzidenzen gut schätzen, bis  $10^{-5}$  können wir sie gerade noch schätzen, bis  $10^{-6}$  vielleicht unter besonderen Umständen. Niedrigere Inzidenzen aus Kohortenstudien sind beim heutigen Stand der Technik nicht mehr reproduzierbar. Das wird lange so bleiben. Wir können einige Millionen Leute für Jahrzehnte nicht ohne Verluste lückenlos beobachten. Mir ist keine Studie bekannt, bei der Angaben zu so kleinen Inzidenzen nicht zu unlösbaren Kontroversen geführt hätten. Dies ist deswegen so, weil die wesentlichen Störeinflüsse in ihrer Richtung und Größe empirisch unbekannt bleiben, die Modelle können sie also nicht berücksichtigen. Inzidenzen kleiner als  $10^{-5}$  bis  $10^{-6}$  - also der graue Bereich - liegen unterhalb der Grenze der Erkennbarkeit, erst recht Inzidenzunterschiede. Es sind aber gerade solche kleinen Inzidenzunterschiede, die uns alle erregen. Wir können sie schlicht nicht erkennen. Dies ist nicht so, weil unsere Modelle nicht genau genug wären, sondern weil unsere Daten nicht mehr reproduzierbar beobachtet werden können. Wir können 2-5 Fälle in Großstudien als Basiszahlen nicht einfach so hinnehmen. Fehlende Werte, differentielle Mißklassifikationen, Bias und Confounding führen zu erheblichen Unsicherheiten, die durch die Modelle nicht eskompiert werden können, weil ihre Richtung unbekannt ist und sie im Zufallsfehler nicht enthalten sind. Unsere Meßinstrumente sind empirisch ungenau. Andere Statistiken, z.B. Mortalitätsdaten, können wir nicht verwenden, da sie auch die exponierten Gruppen in einem nicht bekannten Umfang enthalten, der sich im zeitlichen Verlauf ändert. Für die Angabe des Konfidenzbereichs des zurechenbaren Risikos reichen relative Risiken und ihr Fehler nicht, der Fehler wird durch den Fehler der Basisinzidenz mit bestimmt.

#### 4. Odds-ratios größer als 2-3 als empirische Grenze der sicheren Erkennbarkeit

Eine weitere Grenze der sicheren Erkennbarkeit für Kausalaussagen sind odds-ratios kleiner als 2-3. Die Erfahrung zeigt, daß kleine odds-ratios stark variieren. Ein Faktor 2-5 ist nicht ungewöhnlich zwischen Studien in Abhängigkeit von Vergleichsgruppen, Design, Bias und Confounding. Je kleiner die odds-ratio, desto unsicherer wird sie. Man erhält bei kleinen odds-ratios auch leicht technisch signifikante Ergebnisse, z. B. bei Metaanalysen durch die künstliche Aufblähung der Fallzahl. Da Bias und Confounding meist schwer zu erkennen sind, sagen kleine odds-ratios nicht viel bezüglich der Kausalität. Odds-ratios von 5-10 sind dagegen durchaus ernst zu nehmen.

Die Grenze der empirischen Erkennbarkeit für odds-ratios von etwa 2 wird mit der Basisinzidenz zusammenhängen. Wenn diese kleiner ist, wird die Grenze höher liegen. Das ist durch die eingezeichnete Kurve in Abb. 1 angedeutet. Sie ist willkürlich und ihr genauer Verlauf ist unbekannt.

Unsere empirischen Erhebungstechniken sind sehr viel ungenauer als unsere Formeln. Die Grenze der Aussagefähigkeit und Erkennbarkeit wird natürlich durch die Qualität unserer Erhebungstechniken bestimmt. Dadurch sind die beiden genannten empirischen Grenzen der Erkennbarkeit festgelegt.

Risiken aufgrund epidemiologischer Untersuchungen sind dann relevant, wenn

1. der Kausalindex hoch ist
2. die Basisinzidenz größer als  $10^{-5}$  ist
3. das relative Risiko größer als 2-3 ist.

Wir sollten uns auf diese relevanten Risiken konzentrieren. Es gibt ihrer genug. Im Nebel des Zufalls unter der Schwelle der Erkennbarkeit herumzustochern kann nur zu fruchtlosen Kontroversen führen.

#### 5. Toxikologie und Epidemiologie als gleichwertige Partner

Bei kleinen Risiken wird heute eine Kausalbeziehung meist aufgrund toxikologischer Überlegungen angenommen. Auch hier gibt es Grenzen der Erkennbarkeit. Die Kritik des toxikologischen Paradigmas muß hier aus Zeitgründen entfallen, sie ist gravierend. Der Mensch ist keine große Ratte. Die Toxikologie kann nur bestimmen, ob es ein Risiko am Menschen geben kann, aber nicht, daß ein solches Risiko tatsächlich existiert und wie hoch es ist. Die Epidemiologie bestimmt, ob ein Risiko am Menschen vorhanden ist und wie hoch es ist.

Epidemiologie und Toxikologie sind gleichwertige Partner in der Einschätzung von Risiken. Die Evidenz aus beiden Bereichen stimmt nicht immer überein, wie aus der Tabelle 3 hervorgeht.

Die toxikologische Evidenz wird für sich allein bewertet und die epidemiologische Evidenz auch. Beide werden nach 4 Stufen klassifiziert, die teilweise an die IARC (9) angelehnt sind. In den mit A bezeichneten Fällen kann ein Kausalzusammenhang angenommen werden, in den mit B bezeichneten ist er möglich, in den mit D bezeichneten ist er nicht gegeben. Die beiden mit C bezeichneten Felder enthalten widersprüchliche Evidenz. Das Feld in der 3. Zeile ist häufig: begrenzte Evidenz aus der Toxikologie und inadäquate oder keine Evidenz aus der Epidemiologie. In einem solchen Fall sollte man mit dem Urteil zurückhaltend sein und weitere empirische Evidenz abwarten.

Tab. 3: Die Verbindung von toxikologischer und epidemiologischer Evidenz in der Einschätzung von Risiken

		Toxikologische Evidenz			
		++	+	0	-
Epid. Evidenz	++	A	A	A	A
	+	A	A	B	B
	0	B	C	D	D
	-	C	D	D	D

++ sufficient evidence

+ limited evidence

0 unzureichende Evidenz oder keine Daten

- Evidenz für protektiven Effekt

A Kausalzusammenhang kann angenommen werden

B Kausalzusammenhang ist möglich

C Kausalzusammenhang ist widersprüchlich/offen

D Kausalzusammenhang nicht gegeben

### 6. Ansätze für eine ganzheitliche Theorie der Risikoabschätzung

Beim risk assessment kommt es auf einen ganzheitlichen Ansatz an, der in sich konsistent und insgesamt überzeugend ist. Bausteine für eine solche Theorie des risk assessment sind: hohe Kausalindices als Voraussetzung einer Angabe des zurechenbaren Risikos, standardisierte Vergleiche mit anderen Risiken in einem Risk-Pictogramm, Inzidenzen größer als  $10^{-6}$  und relative Risiken größer als 2, verbunden mit dem systematischen Vergleich mit der toxikologischen Evidenz. In ihrer Gesamtheit grenzen diese Instrumente sicheres Wissen über Kausalzusammenhänge von unsicherer Spekulation ab.

Angesichts der empirischen Grenzen der Erkennbarkeit wird sich der Epidemiologe als Wissenschaftler begrenzen auf das, was Sache ist. Er wird vermeiden, unbegründete Hoffnungen oder Befürchtungen zu wecken. In der heutigen Situation könnte es für den gesellschaftlich engagierten Epidemiologen befriedigender sein, bekannte und relevante Kausalzusammenhänge in präventives Handeln umzusetzen, als zu versuchen, neue Risikoaussagen jenseits des empirisch sicher Erkennbaren aufzuspüren.

#### Literatur:

- (1): IARC Working Group: IARC Monographs on the evaluation of carcinogenic risk of chemicals to humans. 1985 Vol 36 Lyon 18-19.
- (2): Rothmann KJ: Modern epidemiology 1986 Boston/Toronto
- (3): Lanes S: Causal inference is not a matter of science. AM J Epidemiology 1985, 122, 550
- (4): Slavin RE: Best evidence synthesis: An alternative to meta-analytic and traditional reviews. Educational Researcher 1986, Nov. 5-11
- (5): Überla, K.: Epidemiologie und Minimalrisiken - Möglichkeiten und Grenzen einer handlungsrelevanten Risikobeurteilung. Öff. Gesundheits-Wesen 52 (1990) 23-28
- (6): Überla, K.: Boundaries of perception and knowledge for risk assessment in epidemiology. In: Rylander et al (Eds). Assessing low risk agents for lung cancer: methodological aspects. Am J Epid. (in print).
- (7): US Department of Health, Education and Welfare. Smoking and Health: Report of the Advisory Comittee to the Surgeon General of the Public Health Service. PHS Publ. No 1103 Washington DC Printing Office 1964