

Zur Neurovirulenzprüfung von Poliomyelitis-Lebendimpfstoffen

V. Mitteilung

Ein statistisches Verfahren zur Auswertung des Versuches zum Vergleich der Neurovirulenz für den Affen

D. SEINSCHE und O. BONIN

Staatliche Anstalt für experimentelle Therapie „Paul Ehrlich-Institut“
Frankfurt/Main (K. Direktor: Prof. Dr. G. HEYMANN)

F. UNTERHARNSCHEIDT

Deutsche Forschungsanstalt für Psychiatrie (Max-Planck-Institut) München
(Direktor: Prof. Dr. G. PETERS) sowie
Division of Neuropathology, Experimental Neurosurgery and Experimental
Neurology (Prof. Dr. F. UNTERHARNSCHEIDT)
The University of Texas, USA, Medical Branch, Galveston/Tex. 77550

K. SCHMIDT und I. SCHMIDT

Staatliche Anstalt für experimentelle Therapie
„Paul Ehrlich-Institut“ Frankfurt/Main

Eingegangen am 23. Juni 1966

In den vorangegangenen Mitteilungen dieser Reihe wurde gezeigt, daß die einzelnen Typen des Impfstammsatzes von SABIN sowohl nach intrathalamischer als auch nach intralumbaler Infektion beim Affen histopathologische Läsionen eines bestimmten Ausbreitungsmusters und typischer Qualität hervorrufen. Bei der Beurteilung der Neurovirulenz von einzelnen Impfstoffchargen kommt es nun darauf an festzustellen, ob sich die mit dem Impfstoffvirus erzeugten Läsionen von den in einem Parallelversuch mit dem Bezugsvirus erzeugten Läsionen so unterscheiden, daß man einen echten Unterschied in der Neurovirulenz beider Viren annehmen muß.

Dabei ist es wesentlich, die Unterschiede im Ausbreitungsmuster der Läsionen nicht durch ein zu stark vereinfachendes numerisches Auswertungsverfahren zu verwischen.

Ein Versuch zu einer statistisch-numerischen Bewertung von Neurovirulenzversuchen, bei denen aus praktischen Gründen nur eine begrenzte Tierzahl eingesetzt werden kann, wird im folgenden geschildert:

Beim Neurovirulenztest wird für verschiedene Dosen nach der gleichen Technik jeweils eine Gruppe Affen mit dem Bezugsvirus und eine weitere Gruppe Affen mit dem zu prüfenden Impfstoffvirus injiziert. Nach der Injektionstechnik unterscheidet man den intrathalamischen

und den intralumbalen Teilversuch. Die einer statistischen Auswertung zugrundeliegenden Ergebnisse eines solchen Tests sind die Befunde von histologischen Untersuchungen an Schnitten aus den verschiedenen Regionen des Zentralnervensystems der injizierten Affen.

Bei jedem Versuchstier finden wir ein bestimmtes pathomorphologisches Bild, das für die statistische Auswertung als eine Klassifikation sämtlicher untersuchter Schnitte des Affen nach dem darin gefundenen Läsionsgrad und der Region des Zentralnervensystems, aus der die Schnitte stammen, beschrieben wird. Dieses pathomorphologische Bild beschreibt die Spuren der letzten Phase eines noch weitgehend unbekannten Verbreitungsprozesses der Viren im Zentralnervensystem der Affen. Da die morphologischen Alterationen bei der Poliomyelitis nicht gleichmäßig, sondern diskontinuierlich verbreitet sind, ist das pathomorphologische Bild eines Affen, wenn man das in den Schnitten untersuchte Gewebe mengenmäßig dem gesamten Gewebe des Zentralnervensystems gegenüberstellt, selbst bei großer Schnitzahl noch ein im höchsten Maße zufälliges Bild der tatsächlichen Verhältnisse. Ein systematisches Suchen nach Gesetzmäßigkeiten des Ausbreitungsprozesses in den pathomorphologischen Bildern würde hierdurch natürlich erschwert. Ja es dürfte sogar schon deshalb von vornherein aussichtslos sein, weil selbst bei den unfangreichsten Versuchen das Verhältnis der Anzahl von im Versuch gefundenen pathomorphologischen Bildern zur Anzahl der insgesamt möglichen verschiedenen Bilder von ähnlicher Größenordnung wäre, wie das Verhältnis der Menge des untersuchten Gewebes zur Menge des gesamten Gewebes im Zentralnervensystem. Es ist daher nicht verwunderlich, daß solche Gesetzmäßigkeiten bei den pathomorphologischen Bildern bis heute noch nicht bekannt sind. Schon eine so plausible Vorstellung, daß der Anteil der Schnitte mit Nervenzellenuntergang an denjenigen Schnitten, die überhaupt eine Läsion zeigen, in allen Regionen gleich sein könnte, wird durch unsere Versuche klar widerlegt. (Vgl. Abb. 1, 3 und 7 der II. Mitteilung.)

Ein Auswertungsverfahren, das — wie es bei vielen Impfstoffen angewandt wird — auf der Dosisabhängigkeit beruht, etwa durch Schätzung und Vergleich von Dosen mit einer bestimmten Wirkung (eine solche Dosis wäre z. B. diejenige, bei der 50% aller Schnitte eine Läsion zeigen) entbehrt daher bis jetzt jeder statistischen Grundlage. Darüber hinaus müßte es konsequenterweise zu Aussagen von der Form „ $a \text{ cm}^3$ von Viruspräparat A wirken wie $b \text{ cm}^3$ von Viruspräparat B“ führen. Solche Aussagen dürfen jedoch bei Polioviren unter anderem schon deshalb ausscheiden, weil sich hier verschiedene Typen durch eine unterschiedliche Qualität und Ausbreitung in die einzelnen Regionen des Zentralnervensystems charakterisieren lassen.

Ein auf den ersten Blick bestechender Lösungsansatz beruht auf der Vorstellung, daß das Gesamtausmaß der morphologischen Alterationen, die im Zentralnervensystem eines Affen gefunden werden, sich aus den Läsionen in den einzelnen untersuchten Schnitten zusammensetzt und daher messen läßt als Summe von Bewertungsziffern für die einzelnen Schnitte. Die einzelne Bewertungsziffer hängt dann jeweils ab: 1. Von dem Läsionsgrad, den der Schnitt zeigt, 2. von der Region, aus der er stammt, 3. von der Konzentration des Präparates, das die Läsionen hervorgerufen hat und 4. von der Injektionstechnik, die angewandt wurde. Aber schon die abstrakte Formulierung dieses Konzeptes: „A viele Schnitte aus der Region B und mit dem Läsionsgrad C hervorgerufen durch die Konzentration D wiegen genau so schwer wie E viele Schnitte aus der Region F und mit dem Läsionsgrad G hervorgerufen durch die Konzentration H“ zeigt die ganze Problematik dieses Vorgehens.

Darüber hinaus würde sich hier die Frage erheben, wie die diesem Verfahren zugrunde liegenden Bewertungsziffern zu bestimmen sind. Man könnte sie zunächst auf Grund bisheriger Erfahrungen mehr oder weniger per definitionem festlegen, wobei sich wohl schwerlich eine Übereinstimmung erreichen ließe. Man könnte sie auf Grund der Ergebnisse eines Testes selbst festlegen — aber dann würden sich die Bewertungsziffern bei jedem Versuch ändern. Man könnte sie auf Grund der Ergebnisse aller bisher mit dem gleichen Virus durchgeföhrten Versuche festlegen; aber dies könnte zur Vernachlässigung von Unterschieden in den Ergebnissen von Tests zu verschiedenen Zeiten und mit unterschiedlichem Tiermaterial föhren. Außerdem wäre mit der Wahl einer dieser Möglichkeiten eine endgültige Entscheidung über die Größe der Bewertungsziffern ja noch lange nicht getroffen.

Dies ist das Kernproblem einer jeden nicht mehr zu verbessernden statistischen Auswertungsmethode. Man muß ein Maß für die geweblichen Alterationen im Zentralnervensystem der Affen finden, das dann mit einem Maß für die Qualität von Impfstoffen gleichbedeutend ist. Jede mehr oder weniger schematische oder willkürliche Lösung dieses Problems muß jedoch zu Auswertungsmethoden föhren, die nicht mehr erkennen lassen, ob signifikante Unterschiede in den Testergebnissen auch signifikante Unterschiede in den Impfstoffqualitäten bedeuten bzw. ob nicht-signifikante Unterschiede in den Testergebnissen auch nicht-signifikante Unterschiede in den Impfstoffqualitäten bedeuten.

Solange ein solches plausibles Maß für die geweblichen Veränderungen noch nicht existiert, kann man nicht die Ergebnisse von mit verschiedenen Dosen injizierten Affen oder die Ergebnisse aus verschiedenen Regionen des Zentralnervensystems zusammenfassen. Eine gewissenhafte Auswertungsmethode muß sich daher heute noch auf das Testen von Einzelfragen von der Form beschränken: „Waren unter den mit der Dosis A behandelten Affen in der Region B Schnitte mit dem Läsionsgrad C häufiger bei den mit dem Bezugsvirus oder bei den mit dem Impfstoffvirus behandelten Tieren zu erwarten?“ Die Beantwortung

dieser Einzelfragen durch einen einwandfreien statistischen Test stellt dann allerdings für den Verantwortlichen eine exakte und vollständige Grundlage für die schwierige Entscheidung über Zulassung oder Zurückweisung eines Impfstoffes dar.

Die Frage, ob Schnitte mit einem bestimmten Läsionsgrad in einer bestimmten Region bei einer der beiden mit Bezugsvirus und Impfstoffvirus behandelten Tiergruppen signifikant häufiger aufgetreten sind, ließe sich ohne weiteres mit Hilfe eines statistischen Tests zum Vergleich der beiden entsprechenden gefundenen Häufigkeiten untersuchen. Leider sind die üblichen Methoden jedoch Näherungsmethoden, die bei kleineren und größeren Häufigkeiten, wie sie beim Affen-Neurovirulenz-Test meist zu erwarten sind, versagen. Wir haben uns daher für die Anwendung eines sogenannten Randomisierungstests entschlossen, eines statistischen Tests, der zwar mit großem Rechenaufwand verbunden ist und den Einsatz eines Elektronenrechners erfordert, dafür aber auf einer höchst plausiblen Überlegung beruht.

Statistische Methode

Nach der verteilungsfreien Methode der Randomisierungstests glauben wir die Frage, ob unter zwei Affengruppen A und B in einer Region C des Zentralnervensystems Schnitte mit einem Läsionsgrad D bei der Gruppe A signifikant häufiger auftreten als bei der Gruppe B, folgendermaßen beantworten zu können.

Man berechnet zunächst für die Gruppe A die durchschnittliche Häufigkeit der Schnitte in der Region C und mit dem Läsionsgrad D. Dann beginnt man damit, Affen zwischen den beiden Gruppen systematisch auszutauschen. Für jede Austauschmöglichkeit (einschließlich dem Austausch überhaupt keines Affen) berechnet man die gleiche durchschnittliche Häufigkeit für die nach dem Austausch jeweils neu arrangierte Gruppe A. Als Endergebnis berechnet man, für wie viele pro Tausend dieser Austauschmöglichkeiten die berechnete Häufigkeit kleiner war als für die ursprüngliche Gruppe A, in der noch keine Affen ausgetauscht wurden. Wenn diese Zahl extrem groß ist, die Affen sich also auf die beiden Gruppen so verteilen, daß die durchschnittliche Häufigkeit der Schnitte in der Region C und mit dem Läsionsgrad D in der Gruppe A extrem größer ist als in der der Gruppe B, dann fühlen wir uns berechtigt, von einem signifikanten Unterschied zu sprechen. Im folgenden sprechen wir bei Signifikanzzahlen von 950–989 von schwacher Signifikanz und bei Zahlen von 990–1000 von starker Signifikanz. Das heißt mit anderen Worten, bei schwacher Signifikanz hat die ursprüngliche Gruppe A eine höhere Häufigkeit von Läsionen des betreffenden Grades in der betreffenden Region des Zentralnervensystems, als es bei 950–989 von tausend Möglichkeiten für den Austausch von Tieren der Fall wäre. Starke Signifikanz wird angenommen, wenn die Läsionshäufigkeit in der ursprünglichen Gruppe A größer ist als die bei mindestens 990 von 1000 Austauschmöglichkeiten.

Bei Anwendung dieses Verfahrens benötigt man einen Elektronenrechner, weil sich schon bei zwei verhältnismäßig kleinen Affengruppen die Zahl der Austauschmöglichkeiten von Affen zwischen den beiden Gruppen einer fast astronomischen Größenordnung nähert. Wir sind daher dazu übergegangen, von jedem Affen, der

Abb. 1. Lochkarte mit den histopathologischen Befunden eines Versuchsaffen

je in einem Neurovirulenztest unserer Institute benutzt wurde, eine Lochkarte nach dem Muster von Abb. 1 anzulegen. Auf einer solchen Lochkarte sind neben dem pathomorphologischen Bild, das durch die Anzahl der pro Region und Läsionsgrad gefundenen Schnitte gegeben wird, in Ziffern ausgedrückte Angaben gemacht über die Versuchsnummer, das benützte Viruspräparat und eine Konzentration, die Injektionsart und die Species und laufende Nummer des Affen. Die Lochkarten dienen somit nicht nur der rechnerischen Auswertung der Neurovirulenzversuche, sondern sind zugleich auch eine nützliche Dokumentation der bei jedem Tier erhobenen Befunde.

Ergebnisse

Ergebnisse mit dieser Methode liegen bisher nur für den Test mit intralumbaler Injektion vor. Beim intrathalamischen Test ergaben sich wegen der größeren Affengruppen zu lange Rechenzeiten. Ein verbessertes Rechner-Programm, das das häufige Auftreten von Affen ohne jede Läsion beim intrathalamischen Test ausnutzt, wird zur Zeit erprobt.

Die Eignung dieser Methode zur Unterscheidung verschiedener Viren soll zunächst an einem Vergleich zwischen dem attenuierten Virus des Typs I (Bezugsvirus OP 2) und einem Wildvirus des Typs III (Stamm Saukett) demonstriert werden — ein Beispiel, bei dem Unterschiede mit Sicherheit zu erwarten sind. Tab. 1 enthält für jede Verdünnungsstufe und jede Region drei mit dem Elektronenrechner ermittelte Zahlenpaare. Diese unterscheiden sich durch die Art der bei ihrer Berechnung berücksichtigten Läsionen. Läsionsgrad 1—2 (obere Zeile) berücksichtigt nur Schnitte mit mesodermal-gliösen Reaktionen ohne Nervenzelluntergang, Läsionsgrad 3—4 (mittlere Zeile) nur Schnitte mit Nervenzelluntergang und Läsionsgrad 1—4 (untere Zeile) alle Schnitte mit Läsionen gleichgültig welcher Qualität mit Ausnahme von Artefakten und solchen Läsionen, die auf das mechanische Trauma bei der Injektion zurückgeführt werden müssen. In jedem Zahlenpaar spricht eine signifikant hohe Zahl links für eine höhere Läsionshäufigkeit durch das mit dem Bezugspräparat verglichene Virus (zweite Zeile des Tabellenkopfes) — hier das Wildvirus des Typs III. Eine signifikant hohe Zahl rechts würde für eine höhere Läsionshäufigkeit durch das Virus sprechen, das für den betreffenden Versuch als Bezugsvirus gilt (erste Zeile des Tabellenkopfes) — hier das attenuierte Virus OP 2 des Typs I.

In der Tab. 1 finden wir signifikant höhere Läsionshäufigkeiten nur für das Typ-III-Wildvirus.

Die Tabelle lässt außerdem zwei interessante Phänomene erkennen:

1. Signifikant hohe Zahlen finden sich prinzipiell häufiger in den Tiergruppen, die mit großen Virusdosen injiziert sind.
2. Signifikant hohe Zahlen finden sich prinzipiell häufiger in den Regionen, aus denen größere Schnitzzahlen untersucht wurden — also in abnehmender Reihenfolge in der Lumbarregion (18 Schnitte), der Cervicalregion (12 Schnitte), der Medulla oblongata und dem Mesencephalon (je 4 Schnitte) und der motorischen Großhirnrinde (2 Schnitte).

Tabelle 1. Ergebnisse des Randomisierungstests für den Vergleich zwischen dem attenuierten Bezugsvirus, Typ I (OP 2) und dem neuropathogenen Wildvirus, Typ III (Saukett)

| Virus: | Typ: | Versuch Nr.: | Affenart: | Injektion: |
|------------|------|--------------|-----------|-------------|
| Bezug OP 2 | 1 | 464 | Rhesus | intraspinal |
| Saukett | 3 | 964 | Rhesus | intraspinal |

| Verd. | Qual. | Lumbarregion | Cervicalregion | Medulla oblong. | Mesencephalon | Motor. Region | |
|-----------|-------|--------------|----------------|-----------------|---------------|---------------|-----|
| 10^0 | 1-2 | 44 935 | 783 135 | 191 580 | 44 774 | 945 | 0 |
| | 3-4 | 999 1 | 994 0 | 961 15 | 997 0 | 985 | 0 |
| | 1-4 | 1000 0 | 997 1 | 850 88 | 954 8 | 997 | 0 |
| 10^{-1} | 1-2 | 71 871 | 871 14 | 500 314 | 500 186 | 786 | 0 |
| | 3-4 | 971 14 | 986 0 | 986 0 | 786 0 | 214 | 214 |
| | 1-4 | 986 0 | 986 0 | 929 0 | 886 29 | 643 | 71 |
| 10^{-2} | 1-2 | 0 841 | 972 1 | 742 152 | 318 152 | 0 | 417 |
| | 3-4 | 999 0 | 999 0 | 990 0 | 990 0 | 583 | 0 |
| | 1-4 | 999 0 | 999 0 | 999 0 | 981 1 | 318 | 152 |
| 10^{-3} | 1-2 | 848 0 | 583 0 | 583 0 | 0 0 | 583 | 0 |
| | 3-4 | 955 0 | 848 0 | 583 0 | 848 0 | 583 | 0 |
| | 1-4 | 990 0 | 848 0 | 583 0 | 848 0 | 583 | 0 |
| 10^{-4} | 1-2 | 0 417 | 0 0 | 0 0 | 0 0 | 0 | 0 |
| | 3-4 | 583 0 | 583 0 | 583 0 | 583 0 | 583 | 0 |
| | 1-4 | 583 152 | 583 0 | 583 0 | 583 0 | 583 | 0 |

In der betreffenden Region, Verdünnung und Läsionsart spricht jeweils eine signifikant hohe Zahl links für eine höhere Läsionshäufigkeit beim Virus Saukett und eine signifikant hohe Zahl rechts für eine höhere Läsionshäufigkeit beim Virus OP 2. Zahlen von 950-989 (kursiv) bedeuten schwache Signifikanz und Zahlen von 990-1000 (Fettdruck) bedeuten starke Signifikanz.

Die Erklärung für diese Phänomene ist in der statistischen Methode zu suchen: Je mehr Schnitte man untersucht und je häufiger man in den untersuchten Schnitten Läsionen erwarten kann, um so größer wird die Chance, mit diesem Verfahren bestehende Unterschiede der mittleren Läsionshäufigkeit als signifikant bestätigt zu bekommen.

Das Auffinden von signifikanten Unterschieden zwischen einem Wildvirus und einem attenuierten Virus besagt natürlich noch nichts über die Brauchbarkeit der Methode für die Prüfung von Poliomyelitis-Lebendimpfstoffen, bei der ja zwei attenuierte Viren verglichen werden. Es ergibt sich daher die Frage, ob diese Methode wirksam genug ist, die in den früheren Mitteilungen dieser Reihe zusammenfassend beschriebenen Unterschiede zwischen den attenuierten Viren der verschiedenen Typen auch im Einzelversuch zu erkennen.

Beim Vergleich mit dem attenuierten Bezugsvirus des Typs I sind Unterschiede am ehesten für die attenuierten Viren des Typs III zu erwarten. Tab. 2 faßt daher die in Versuchen mit 7 verschiedenen

Tabelle 2. Signifikante Ergebnisse aus 7 Versuchen mit verschiedenen Pools von attenuierten Typ III-Viren im Vergleich zum attenuierten Bezugsvirus, Typ I (OP 2)

| Viruspool | Virusverd. | Region | Qualität | Signifikanzzahlen für höhere Läsionshäufigkeiten beim | |
|-----------|------------|---------|----------|---|-------|
| | | | | Pool | Bezug |
| 01 | 10^0 | Med. | 3—4 | 951 | — |
| 02 | 10^{-1} | Lumb. | 1—4 | — | 979 |
| | 10^{-2} | Lumb. | 1—2 | — | 965 |
| 03 | 10^0 | Mesenc. | 1—4 | 951 | — |
| | 10^{-1} | Med. | 1—2 | 973 | — |
| | 10^{-1} | Mesenc. | 1—4 | 955 | — |
| 04 | 10^0 | Lumb. | 1—2 | 983 | — |
| 05 | 10^0 | Lumb. | 1—2 | 955 | — |
| | 10^0 | Mesenc. | 1—2 | 972 | — |
| | 10^0 | Mesenc. | 1—4 | 972 | — |
| | 10^{-2} | Lumb. | 3—4 | 969 | — |
| | 10^{-3} | Lumb. | 1—2 | 973 | — |
| 06 | 10^{-3} | Lumb. | 1—4 | 973 | — |
| | 10^0 | Lumb. | 3—4 | 957 | — |
| | 10^0 | Cerv. | 1—4 | 951 | — |
| | 10^0 | Mesenc. | 1—4 | 955 | — |
| | 10^{-1} | Lumb. | 1—2 | 999 | — |
| | 10^{-1} | Lumb. | 3—4 | 983 | — |
| | 10^{-1} | Lumb. | 1—4 | 999 | — |
| | 10^{-1} | Cerv. | 1—2 | 962 | — |
| | 10^{-1} | Cerv. | 1—4 | 962 | — |
| | 10^{-1} | Med. | 1—2 | 962 | — |
| 07 | 10^{-1} | Med. | 1—4 | 962 | — |
| | 10^{-1} | Mesenc. | 3—4 | 962 | — |
| | 10^{-1} | Mesenc. | 1—4 | 962 | — |

Zahlen von 950—989 (kursiv) bedeuten schwache Signifikanz und Zahlen von 990—1000 (Fettdruck) bedeuten starke Signifikanz.

Pools¹ von attenuierten Typ-III-Viren gefundenen signifikanten Unterschiede zum Typ-I-Bezugsvirus OP 2 zusammen.

Stark signifikante Unterschiede im Sinne dieses Verfahrens ergaben sich dabei nur für das Pool 06 in der Lumbalregion und der Verdünnungsstufe 10^{-1} für die Läsionsqualitäten 1—2 (mesodermal-gliöse Reaktionen) und 1—4 (alle Läsionen zusammen) — nicht aber für die Nervenzellschädigung der Läsionsqualitäten 3—4. Überhaupt zeigen die Tiere, die mit dieser Verdünnungsstufe des Pool 06 injiziert waren, auch in den anderen Regionen des Zentralnervensystems stärkere mesodermal-gliöse Reaktionen, was in Tab. 2 in den zahlreichen schwach signifikanten Unterschieden bei Pool 06, Verdünnungsstufe 10^{-1} zum Ausdruck kommt. Ob dies

¹ Bei den „Pools“ in dieser Studie handelt es sich zum Teil um experimentelle Vermehrungsansätze und zum Teil um Impfstoffkonzentrate.

Tabelle 3. Signifikante Ergebnisse aus 5 Versuchen mit verschiedenen Pools von attenuierten Typ II-Viren im Vergleich zum attenuierten Bezugsvirus, Typ I (OP 2)

| Viruspool | Virusverd. | Region | Qualität | Signifikanzzahlen für höhere Läsionshäufigkeiten beim | |
|-----------|------------|---------|----------|---|------------|
| | | | | Pool | Bezug |
| 01 | 10^0 | Lumb. | 1-2 | 972 | — |
| | 10^{-2} | Lumb. | 1-2 | — | 994 |
| | 10^{-2} | Lumb. | 1-4 | — | 994 |
| 02 | 10^{-2} | Lumb. | 1-2 | — | 965 |
| 03 | 10^{-1} | Lumb. | 1-2 | — | 963 |
| | 10^{-1} | Lumb. | 1-4 | — | 966 |
| 04 | 10^{-1} | Lumb. | 1-2 | — | 968 |
| | 10^{-1} | Lumb. | 1-4 | — | 988 |
| 05 | 10^0 | Mesenc. | 1-2 | 993 | — |
| | 10^0 | Mesenc. | 1-4 | 993 | — |
| | 10^{-2} | Lumb. | 1-2 | — | 991 |
| | 10^{-2} | Lumb. | 1-4 | — | 991 |
| | 10^{-2} | Med. | 1-2 | — | 987 |
| | 10^{-2} | Med. | 1-4 | — | 987 |

Zahlen von 950—989 (kursiv) bedeuten schwache Signifikanz und Zahlen von 990—1000 (Fettdruck) bedeuten starke Signifikanz.

mit einem möglicherweise stärkeren Injektionstrauma bei den Tieren dieser Gruppe zusammenhängen könnte, wird zur Zeit noch genauer untersucht.

Bei den meisten anderen Typ-III-Pools finden sich dagegen nur Unterschiede schwacher Signifikanz, und zwar mit nur einer Ausnahme zuungunsten der Typ-III-Viren. Man kann auch nicht sagen, daß bestimmte Regionen des Zentralnervensystems oder bestimmte Virusverdünnungen bevorzugt solche Unterschiede zeigen würden, wie es bei Pool 06 der Fall ist.

Die höhere Neurovirulenz der attenuierten Typ-III-Viren im Vergleich zu denen des Typs I kommt also auch bei diesem Bewertungsverfahren im Einzelversuch zum Ausdruck. Bei Bewertung unseres begrenzten Materials hat es den Anschein, als ob die Neurovirulenzunterschiede zwischen den Typen III und I im Sinne dieses Verfahrens nur schwach signifikant wären.

Auch die Ergebnisse mit 5 Pools des attenuierten Typ-II-Virus (Tab.3) scheinen mit der „Wertreihe der Neurovirulenz“ in Übereinstimmung zu stehen.

In Verdünnungsstufe 10^{-2} hat dieses Virus bei 2 unter 5 Pools im Lendenmark stark signifikant weniger Läsionen (insbesondere von Grad 1—2) hervorgerufen als das Bezugsvirus von Typ I im Parallelversuch. Schwach signifikante Unterschiede zugunsten des Typ-II-Virus bestehen in der gleichen Verdünnungsstufe bei einem weiteren Pool (02) und in Verdünnungsstufe 10^{-1} bei zwei Pools (03 und 04). Die stärkere Tendenz des Typ-II-Virus, im Mesencephalon mesodermal-gliöse Reaktionen hervorzurufen, kommt mit dieser Bewertung der Einzelversuche nur bei einem Pool (05) als stark signifikanter Unterschied zum Ausdruck. Die zweifellos

Tabelle 4. Signifikante Ergebnisse aus 11 Versuchen mit verschiedenen Pools von attenuiertem Typ I-Viren im Vergleich zum attenuierten Bezugsvirus, Typ I (OP 2)

| Viruspool | Virusverd. | Region | Qualität | Signifikanzzahlen für höhere Läsionshäufigkeiten beim | |
|-------------|------------|---------|----------|---|------------|
| | | | | Pool | Bezug |
| 01 | 10^{-1} | Cerv. | 1-2 | — | 951 |
| 02 | — | — | — | — | — |
| 03 | — | — | — | — | — |
| 04 | 10^0 | Lumb. | 1-2 | 998 | — |
| | 10^0 | Lumb. | 1-4 | 998 | — |
| | 10^0 | Cerv. | 1-2 | 959 | — |
| | 10^0 | Cerv. | 1-4 | 959 | — |
| 04, Wdh. | 10^{-1} | Lumb. | 1-2 | 990 | — |
| | 10^{-2} | Lumb. | 1-2 | 963 | — |
| | 10^{-2} | Lumb. | 3-4 | 981 | — |
| | 10^{-2} | Lumb. | 1-4 | 978 | — |
| 05 | 10^0 | Lumb. | 1-2 | 992 | — |
| | 10^0 | Lumb. | 1-4 | 992 | — |
| | 10^{-1} | Mesenc. | 1-2 | 950 | — |
| | 10^{-1} | Mesenc. | 1-4 | 950 | — |
| 06 | 10^0 | Lumb. | 1-2 | 958 | — |
| | 10^0 | Lumb. | 1-4 | 970 | — |
| 07 | — | — | — | — | — |
| 08 | — | — | — | — | — |
| 09 | 10^{-1} | Cerv. | 1-2 | — | 974 |
| | 10^{-1} | Cerv. | 1-4 | — | 974 |
| 10 | — | — | — | — | — |
| 11 | — | — | — | — | — |

Zahlen von 950—989 (kursiv) bedeuten schwache Signifikanz und Zahlen von 990—1000 (Fettdruck) bedeuten starke Signifikanz.

vorhandenen Unterschiede im pathomorphologischen Bild der Affenpoliomyelitis durch die attenuierten Viren der Typen II und I sind also mit diesem Verfahren auch bei einer Bewertung des Einzelversuchs zu erkennen. Sie scheinen so stark zu sein, daß sie meist zu schwach signifikanten — teilweise aber auch zu stark signifikanten Unterschieden der Läsionshäufigkeiten in vergleichbaren Bewertungsgruppen führen.

Am interessantesten für die Beurteilung dieses Bewertungsverfahrens sind jedoch die Ergebnisse aus Vergleichsversuchen mit Viren des gleichen Typs. Tab. 4 bringt zusammenfassend alle signifikanten Unterschiede aus Vergleichen von 11 Pools mit dem homologen Bezugsvirus.

Mit Ausnahme von zwei dieser Viruspools (04 und 05) wurden nur in drei Bewertungsgruppen (eine Region des Nervensystems von Affen, die mit gleichen Konzentrationen der zu vergleichenden Viren injiziert

Tabelle 5. Ergebnisse des Randomisierungstests für den Vergleich zwischen dem attenuierten Bezugsvirus, Typ I (OP 2) und dem attenuierten Virus, Typ I, Pool 04 (1. Versuch)

| Virus: | Typ: | Versuch Nr.: | | Affenart: | Injektion: | | | | | | |
|------------------|------|--------------|----------|--------------|----------------|-----------------|---------------|---------------|-----|-----|---|
| Bezug OP 2 | 1 | 862 | | Rhesus | intraspinal | | | | | | |
| Pool 04 | 1 | 862 | | Rhesus | intraspinal | | | | | | |
| | | Verd. | Qual. | Lumbarregion | Cervicalregion | Medulla oblong. | Mesencephalon | Motor. Region | | | |
| 10 ⁰ | 1—2 | 998 | 1 | 959 | 25 | 914 | 21 | 685 | 70 | 0 | 0 |
| | 3—4 | 583 | 0 | 0 | 0 | 0 | 462 | 0 | 0 | 0 | 0 |
| | 1—4 | 998 | 1 | 959 | 25 | 865 | 37 | 685 | 70 | 0 | 0 |
| 10 ⁻¹ | 1—2 | 946 | 50 | 463 | 502 | 269 | 437 | 563 | 168 | 0 | 0 |
| | 3—4 | 583 | 0 | 0 | 462 | 0 | 462 | 0 | 0 | 0 | 0 |
| | 1—4 | 946 | 40 | 432 | 523 | 269 | 584 | 563 | 168 | 0 | 0 |
| 10 ⁻² | 1—2 | 630 | 359 | 424 | 455 | 597 | 208 | 900 | 41 | 545 | 0 |
| | 3—4 | 0 | 0 | 0 | 455 | 0 | 455 | 0 | 0 | 0 | 0 |
| | 1—4 | 630 | 359 | 424 | 455 | 511 | 359 | 900 | 41 | 545 | 0 |
| 10 ⁻³ | 1—2 | 0 | 0 | 0 | 417 | 583 | 0 | 0 | 0 | 0 | 0 |
| | 3—4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 1—4 | 0 | 0 | 0 | 417 | 583 | 0 | 0 | 0 | 0 | 0 |
| 10 ⁻⁴ | 1—2 | 930 | 0 | 808 | 0 | 685 | 70 | 538 | 0 | 0 | 0 |
| | 3—4 | 0 | 0 | 0 | 0 | 538 | 0 | 0 | 0 | 0 | 0 |
| | 1—4 | 930 | 0 | 808 | 0 | 685 | 70 | 538 | 0 | 0 | 0 |

In der betreffenden Region, Verdünnung und Läsionsart spricht jeweils eine signifikant hohe Zahl links für eine höhere Läsionshäufigkeit beim Virus 04 und eine signifikant hohe Zahl rechts für eine höhere Läsionshäufigkeit beim Virus OP 2. Zahlen von 950—989 (kursiv) bedeuten schwache Signifikanz und Zahlen von 990—1000 (Fettdruck) bedeuten starke Signifikanz.

waren) schwach signifikante Unterschiede festgestellt, davon zweimal zuungunsten des Bezugsvirus (01 und 09) und einmal zuungunsten des Pools (06). Bei der großen Anzahl von insgesamt 225 Bewertungsgruppen für diese 9 Pools ist dies sicher kein verwunderliches Ergebnis.

Auffällig ist dagegen die Häufung von stark und schwach signifikant höheren Läsionshäufigkeiten bei den Pools 04 und 05. Mit Pool 04 wurde deswegen ein Wiederholungsversuch durchgeführt, dessen signifikante Ergebnisse in Tab. 4 ebenfalls enthalten sind.

Da dieses Pool besonders interessant erscheint, werden die Signifikanzberechnungen für die beiden damit angesetzten Neurovirulenztests in Tab. 5 und 6 im Detail mitgeteilt. Der erste Versuch mit Pool 04 (Tab. 5) brachte zwei stark signifikante Unterschiede im Lendenmark der mit unverdünntem Virus injizierten Affen und zwei schwach signifikante Unterschiede für die Cervicalregion der gleichen Tiergruppe. Beide betrafen die mesodermal-gliösen Reaktionen und waren so stark, daß sie sich auch noch bei der Gesamtbewertung aller Läsionsqualitäten niederschlugen, obwohl Nervenzelluntergang nicht wesentlich häufiger war. Im Wiederholungsversuch (Tab. 6) ergab sich ein stark signifikanter Unterschied in der gleichen Region, der aber die mit 10⁻¹ verdünntem Virus injizierten Tiere betraf. Weitere

Tabelle 6. Ergebnisse des Randomisierungstests für den Vergleich zwischen dem attenuierten Bezugsvirus, Typ I (OP 2) und dem attenuierten Virus, Typ I, Pool 04 (2. Versuch)

| Virus: | Typ: | | Versuch Nr.: | Affenart: | Injektion: |
|------------|------|--|--------------|-----------|-------------|
| Bezug OP 2 | 1 | | 464 | Rhesus | intraspinal |
| Pool 04 | 1 | | 464 | Rhesus | intraspinal |

| Verd. | Qual. | Lumbarregion | Cervicalregion | Medulla oblong. | Mesencephalon | Motor. Region |
|-----------|-------|--------------|----------------|-----------------|---------------|---------------|
| 10^0 | 1—2 | 678 316 | 633 295 | 226 591 | 359 500 | 500 0 |
| | 3—4 | 500 484 | 233 500 | 233 500 | 500 0 | 0 0 |
| | 1—4 | 662 330 | 367 562 | 229 647 | 500 359 | 500 0 |
| 10^{-1} | 1—2 | 999 4 | 677 192 | 180 697 | 424 345 | 0 0 |
| | 3—4 | 81 907 | 745 0 | 567 0 | 0 0 | 0 667 |
| | 1—4 | 743 234 | 737 111 | 315 582 | 424 345 | 0 667 |
| 10^{-2} | 1—2 | 963 27 | 805 51 | 590 195 | 200 267 | 0 533 |
| | 3—4 | 981 0 | 467 0 | 0 0 | 467 0 | 0 0 |
| | 1—4 | 978 18 | 805 51 | 590 195 | 446 123 | 0 533 |
| 10^{-3} | 1—2 | 877 0 | 467 0 | 467 0 | 467 0 | 0 0 |
| | 3—4 | 467 0 | 467 0 | 0 0 | 0 0 | 0 0 |
| | 1—4 | 877 0 | 467 0 | 467 0 | 467 0 | 0 0 |
| 10^{-4} | 1—2 | 0 533 | 0 0 | 0 0 | 0 0 | 0 0 |
| | 3—4 | 0 0 | 0 0 | 0 0 | 0 0 | 0 0 |
| | 1—4 | 0 533 | 0 0 | 0 0 | 0 0 | 0 0 |

In der betreffenden Region, Verdünnung und Läsionsart spricht jeweils eine signifikant hohe Zahl links für eine höhere Läsionsfähigkeit beim Virus 04 und eine signifikant hohe Zahl rechts für eine höhere Läsionshäufigkeit beim Virus OP 2. Zahlen von 950—989 (kursiv) bedeuten schwache Signifikanz und Zahlen von 990—1000 (Fettdruck) bedeuten starke Signifikanz.

schwach signifikant höhere Läsionshäufigkeiten sowohl für die mesodermal-gliösen Reaktionen als auch für die Nervenzellschädigungen — wurden im Lendenmark der mit der nächstfolgenden Virusverdünnung injizierten Tiere gefunden.

Nach den Ergebnissen von 11 Versuchen kommen beim Vergleich von Viren des gleichen Typs normalerweise anscheinend keine stark signifikanten Unterschiede vor. Bei einem Viruspool, das solche stark signifikanten Unterschiede zum homologen Bezugsvirus zeigte, wurde im Wiederholungsversuch ein im wesentlichen gleichartiger Unterschied gefunden. Die Annahme, daß dieses Pool tatsächlich eine etwas höhere Neurovirulenz besitzt, liegt daher nahe.

Diskussion

Das statistische Verfahren des Randomisierungstests ist auf Grund seiner theoretischen Grundlage von vornherein geeignet, nur stärkere Unterschiede in den aktuellen Ergebnissen von Versuchen als signifikant zu bestätigen. Man konnte daher erwarten, daß es signifikante Unter-

schiede nur beim Vergleich solcher Viren aufzeigen würde, deren Neurovirulenz sich tatsächlich stärker unterscheidet.

Die begrenzte praktische Erfahrung mit diesem Verfahren hat folgendes gezeigt:

1. Die vorhandenen Neurovirulenzunterschiede zwischen den attenuierten Viren der Typen III und I führen fast immer nur zu schwach signifikant höheren Läsionshäufigkeiten bei den mit Typ-III-Virus injizierten Tieren.

2. Die vorhandenen Neurovirulenzunterschiede zwischen den attenuierten Viren der Typen II und I führen teilweise zu schwach signifikanten, teilweise aber auch zu stark signifikanten Unterschieden, wobei das attenuierte Virus des Typs II im Lendenmark die schwächeren und im Mesencephalon die stärkeren Läsionen hervorzurufen scheint.

3. Beim Vergleich von homologen Viren des Typs I sind signifikante Unterschiede im Sinne dieses Verfahrens sehr selten und — wenn überhaupt vorhanden — nur schwach signifikant. In einem Falle, bei dem stark signifikante Unterschiede gefunden wurden, war dieses Ergebnis im Wiederholungsversuch reproduzierbar.

Für die Praxis der Impfstoffprüfung erscheint es daher vertretbar, im Falle des Vergleichs homologer Viren einen Impfstoff zuzulassen, wenn sich bei Anwendung dieses Verfahrens keine stark signifikanten Unterschiede zeigen. Wenn solche auftreten, scheint es ratsam, die Entscheidung über Zulassung oder Zurückweisung vom Ergebnis eines Wiederholungsversuchs abhängig zu machen.

Für den Vergleich heterologer Viren lassen sich auch mit diesem Verfahren keine verlässlichen Grenzen für die Zulassung oder Zurückweisung von Impfstoffen angeben, da die Unterschiede zwischen den Virustypen etwa so stark sind, daß sie der Empfindlichkeitsschwelle des Verfahrens entsprechen.

Zusammenfassung

Ein statistisches Verfahren, das auf Grund eines Neurovirulenztests ohne weiteres die Entscheidung erlaubt, ob ein geprüfter Poliomyelitis-Lebendimpfstoff von ausreichend geringer Neurovirulenz für den Affen ist, kann nicht angegeben werden. Hierzu fehlen die Kriterien, die die Impfstoffqualität auf Grund von pathomorphologischen Bildern bei injizierten Affen definieren könnten. Es wird ein statistisches Verfahren auf der Grundlage der Randomisierungstests beschrieben, mit dem sich zahlreiche Einzelfragen untersuchen lassen, deren Beantwortung eine exakte und ausreichende Grundlage für die schwierige Entscheidung über die Zulassung oder Zurückweisung von Einzelimpfstoffen darstellt.

Die Ergebnisse mit dieser Methode beim Vergleich von Viren verschiedener Typen im Einzelversuch stehen ohne Ausnahme im Einklang

mit den in der II. Mitteilung bei der Charakterisierung der Typen beschriebenen Unterschieden in Qualität und Ausbreitung der Läsionen im Zentralnervensystem.

Beim Vergleich homologer Viren des Typs I wurden nur in 2 von 11 Versuchen mit verschiedenen Virus-Pools und demselben Bezugsvirus stark signifikante Unterschiede gefunden. Bei der Wiederholung des Versuchs mit einem dieser beiden Pools war dieses Ergebnis reproduzierbar.

Literatur

Bezüglich der Arbeiten zur Neuropathologie und den Bewertungsverfahren siehe bei BONIN, O., u. F. UNTERHARNSCHEIDT: Zur Neurovirulenzprüfung von Poliomyelitis-Lebendimpfstoffen. I. Mitteilung: Grundlagen, Technik und Bewertung des Versuchs zur Messung der Neurovirulenz für den Affen. Arch. Psychiat. Nervenkr. **206**, 260–271 (1964).

UNTERHARNSCHEIDT, F., u. O. BONIN: Zur Neurovirulenzprüfung von Poliomyelitis-Lebendimpfstoffen. II. Mitteilung: Pathomorphologie der Affenpoliomyelitis durch attenuierte Polioviren nach intralumbaler Injektion, unter besonderer Berücksichtigung der Unterschiede zwischen den Virustypen. Arch. Psychiat. Nervenkr. **206**, 454–473 (1965).

Bezüglich der mathematischen Methodik siehe bei

PFANZAGL, J.: Allgemeine Methodenlehre der Statistik, Bd. II, Sammlung Göschen, Bd. 747. Berlin: W. de Gruyter & Co. 1962.

Diplommathematiker D. SEINSCHE und
Professor Dr. O. BONIN
6 Frankfurt am Main 10
Paul Ehrlich-Straße 42–44