Messen faserförmiger Partikel *

Strategie zur Auswertung des ersten VDI–Ringversuchs zur Asbestfaserbestimmung nach Richtlinie VDI 3492 Blatt 1

Joachim Hartung, Guido Knapp

Zusammenfassung: *Die Umsetzung der Richtlinie zur Bestimmung der Anzahl von anorganischen faserförmigen Partikeln auf staubbelegten Filtern wurde in einem Ringversuch von VDI und DIN überprüft. Die Auswertung der Filter wurde zunächst in zwei Referenzlabors und danach in den teilnehmenden Prüflabors vorgenommen. Die Beurteilung der korrekten Umsetzung der Richtlinie durch ein Prüflabor sollte durch den Vergleich der Zählergebnisse aus diesem Prüflabor mit den Ergebnissen aus den beiden Referenzlabors erfolgen.*
Hier wird nun eine Auswertungsmethode für den Vergleich eines Prüflabors mit den beiden Referenzlabors vorgestellt, die auf einer Wurzeltransformation der Zählergebnisse basiert und mögliche Verzerrungen zwischen den Zählergebnissen der Referenzlabors sowie Korrelationen berücksichtigt. Die Umsetzung der Richtlinie wird schließlich mit Hilfe einer meta–analytischen Zusammenfassung von Teststatistiken überprüft.

1. Einleitung

Die VDI–Projekt- und Service GmbH (VDI–GPS) hat 1993/1994 in Zusammenarbeit mit der Kommission Reinhaltung der Luft (KRdL) im VDI und DIN sowie zwei Akkreditiersystemen den ersten Ringversuch zur Bestimmung der Anzahl anorganischer faserförmiger Partikel (z. B. Asbestfasern, sonstige anorganische Fasern) auf staubbelegten Filtern nach Richtlinie VDI 3492 Bl. 1 durchgeführt. Das zu bewertende Meßverfahren ist dabei kein Verfahren, das sich bis dahin im Routinebetrieb etabliert hat, sondern ein Verfahren, das im Rahmen von Forschungsprojekten entwickelt wurde. An dem Versuch haben insgesamt 72 Labors teilgenommen. Die Gründe für die Durchführung eines solchen Ringversuchs sowie die Durchführung selbst und die verwendete Auswertungsmethode mit den entsprechenden Ergebnissen sind ausführlich in Höfert et al. (1996a, 1996b) dargestellt. Die wesentlichen Punkte der Durchführung des Ringversuchs werden im folgenden noch einmal kurz skizziert.

*Die Arbeit erscheint in der Zeitschrift:
ALLGEMEINES STATISTISCHES ARCHIV

Den teilnehmenden Labors wurden drei verschiedene, unterschiedlich mit Staub belegte Filter zur Auswertung gegeben. Da die Auswertung eines Filters ein langsam zerstörender Vorgang ist — in einem Vorversuch stellte sich heraus, daß eine Filterprobe nur fünf- bis sechsmal in verschiedenen Labors ausgewertet werden kann — , konnten nicht alle Prüflabors dieselben drei Filter auswerten. Daher kamen insgesamt 39 Filter zum Einsatz. Entsprechend den Vorgaben der Richtlinie VDI 3492 Bl. 1 unter Berücksichtigung einer geänderten Faserzählregel, die der in der Richtlinie VDI 3492 Bl. 2 angegebenen Faserzählregel a) entspricht, wurde jeweils eine Teilfläche des Filters (maximal 1 mm^2) bei einer Vergrößerung von 2000 bis 2500fach rasterelektronenmikroskopisch ausgewertet. Die gefundenen Fasern wurden mittels energiedispersiver Röntgenanalytik (EDXA–Verfahren) identifiziert. Für jeden Filter wurde differenziert nach Amphibol (lang/kurz), Chrysotil (lang/kurz) und sonstigen anorganischen Fasern (lang/kurz). Dabei wurde deshalb nach zwei Längenklassen unterschieden, weil das Auffinden kurzer Fasern grundsätzlich schwieriger ist als das der langen Fasern. Außerdem wurde für jede dieser Faserklassen getrennt nach kurzen und langen Fasern die Summe über alle drei Filter gebildet. Insgesamt lieferte jedes Prüflabor somit 24 Einzelergebnisse, d. h. Meßwiederholungen an einem Filter fanden nicht statt. Die den Prüflabors jeweils zur Verfügung gestellten Meßfilter wurden zuvor durch zwei Referenzlabors ausgewertet. Diese Referenzlabors lieferten pro Filter und Faserart ebenfalls nur einen Meßwert. Die erfolgreiche Umsetzung der VDI–Richtlinie sollte dann durch den Vergleich der Zählergebnisse des Prüflabors mit den Zählergebnissen der beiden Referenzlabors überprüft werden, d. h. die Zählergebnisse anderer Prüflabors für denselben Meßfilter wurden nicht berücksichtigt. Zur abschließenden Bewertung des Prüflabors sollten die 24 Einzelvergleiche geeignet zu einem Gesamtergebnis zusammengefaßt werden.

Bezüglich der Rolle der Referenzlabors ist in diesem Zusammenhang nach Höfert et al. (1996a) erwähnenswert, daß die Veranstalter des Ringversuchs diese beiden Labors als Referenzlabors verpflichtet haben, weil bei diesen durch die Entwicklungsarbeit zu dieser Meßtechnik, langjährige Praxiserfahrung sowie nationales und internationales Engagement in meßtechnischen Gremien die notwendige Qualifikation und Übersicht vorausgesetzt wurde. Darüber hinaus haben diese Referenzlabors den Satz von Filtern, die bei typischen Meßaufgaben in der Praxis anfielen (Messung zur Bestandsaufnahme, Sanierungsbeurteilungen usw.), für den Ringversuch zusammengestellt und zunächst gemäß der obigen Richtlinie ausgewertet. Durch diese Vorgeschichte und durch die Kenntnis des Umfeldes der jeweiligen Filter kann bzgl. des Verhältnisses zu einem Prüflabor nicht von einer freien Randomisation unter gleichen Labors aus versuchsplanerischer Sicht ausgegangen werden.

In diesem Artikel wird nun eine andere Auswertungsstrategie als in Höfert et al. (1996b) vorgeschlagen. Dazu wird im zweiten Abschnitt zunächst eine geeignete Datentransformation vorgestellt, und mit Hilfe dieser Datentransformation wird unter der Unabhängigkeitsannahme aller Zählergebnisse jeweils eine Teststatistik zum Vergleich der einzelnen Zählergebnisse für die langen bzw. kurzen Fasern sowie der Summe der Zählergebnisse für die langen bzw. kurzen Fasern hergeleitet. Die hergeleiteten Teststatistiken sind approximativ χ^2–verteilt. Der dritte Abschnitt beschäftigt sich dann mit der Problematik, daß systematische Abweichungen in den Zählergebnisse zwischen den Referenzlabors vorliegen. Die dort hergeleiteten Teststatistiken sind approximativ nicht–zentral χ^2–verteilt. Darüber hinaus wird ein modifizierter F–Test vorgestellt, der sowohl systematische Abweichungen als auch mögliche Korrelationsstrukturen zwischen den beiden Referenzlabors berücksichtigt, um dem oben geschilderten versuchsplanerischen Aspekt Rechnung zu tragen. Dieser F–Test ist dabei in dem Sinne stabil, daß er weder systematische Abweichungen noch Korrelationen von vornherein voraussetzt, jedoch in der Lage ist, auf diese adäquat zu reagieren. Der letzte Abschnitt beschäftigt sich schließlich mit der Zusammenfassung der einzelnen Testergebnisse zu einem Gesamtergebnis, wobei dort Ansätze der Meta–Analyse ausgenutzt werden.

2. Datentransformation und unabhängige, identisch verteilte Zählergebnisse

Das Zählergebnis für eine Faserart auf einem Filter läßt sich gemäß der Richtlinie VDI 3492 Bl. 1 als Realisierung einer Poisson–verteilten Zufallsvariablen auffassen. Bezeichne im folgenden X eine solche Poisson–verteilte Zufallsvariable mit Parameter λ, $\lambda > 0$, so erhält man durch die Wurzeltransformation $\sqrt{X}$ eine neue Zufallsvariable, deren Varianz aufgrund des Gaußschen Fehlerfortpflanzungsgesetzes annähernd $\frac{1}{4}$ beträgt. In Anscombe (1948) wird darüber hinaus gezeigt, daß die Varianz der transformierten Zufallsvariablen $\sqrt{X + \frac{3}{8}}$ noch besser durch $\frac{1}{4}$ approximiert wird, falls der Parameter der zugrundeliegenden Poisson–Verteilung klein ist. Weiterhin wird durch die Wurzeltransformation in der Regel eine gute Annäherung an eine Normalverteilung erreicht (vgl. auch Hartung et al. (1995), S. 350). Im weiteren Verlauf dieser Arbeit wird stets die zuletzt erwähnte Wurzeltransformation angewendet.

Mit Hilfe dieser Datentransformation und der daraus resultierenden guten Annäherung an eine normalverteilte Zufallsvariable wird im folgenden eine Teststatistik zur Überprüfung der Hypothese hergeleitet, daß das einzelne Prüflabor genauso mißt wie die Referenzlabors unter der Annahme, daß die einzelnen Zählergebnisse pro Filter und pro Faserart unabhängig und in den Referenzlabors für jeden Filter und jede Faserart identisch verteilt sind. Die Teststatistik wird sowohl zur Überprüfung der Zählergebnisse für die langen Fasern als auch für die kurzen Fasern herangezogen.

Im folgenden wird zunächst die benötigte Notation eingeführt. Sei dazu für eine Faserlängenklasse

X_{ij} - die Zufallsvariable für das Zählergebnis des Prüflabors für die j–te Faserart im i–ten Filter, $i = 1, 2, 3$, $j = 1, 2, 3$,

und

Y_{lij} - die Zufallsvariable für das Zählergebnis des l–ten Referenzlabors für die j–te Faserart im i–ten Filter, $l = 1, 2$, $i = 1, 2, 3$, $j = 1, 2, 3$.

Für diese Zufallsvariablen werden nun die folgenden Annahmen getroffen:

(A1) Für $i = 1, 2, 3$ und $j = 1, 2, 3$ gilt: $X_{ij} \sim Po(\lambda_{ij})$, und die Zufallsvariablen sind stochastisch unabhängig.

(A2) Für $l = 1, 2$, $i = 1, 2, 3$ und $j = 1, 2, 3$ gilt: $Y_{lij} \sim Po(\mu_{lij})$, und die Zufallsvariablen sind stochastisch unabhängig.

(A3) Die Zufallsvariablen X_{ij} sind stochastisch unabhängig von den Zufallsvariablen Y_{1ij} und Y_{2ij} für $i = 1, 2, 3$ und $j = 1, 2, 3$.

Bezeichne im folgenden P_{ij} die Zufallsvariable für das transformierte Zählergebnis aus dem Prüflabor und R_{lij} die Zufallsvariable für das transformierte Zählergebnis aus dem l–ten Referenzlabor für $l = 1, 2$, $i = 1, 2, 3$ und $j = 1, 2, 3$.

Zur Herleitung der Teststatistiken zum Vergleich der Zählergebnisse aus dem Prüflabor und der Referenzlabors wird zusätzlich zu (A1) – (A3) in diesem Abschnitt angenommen, daß die Zählergebnisse der beiden Referenzlabors bezogen auf Filter und Faserart identisch verteilt sind. D. h. in diesem Abschnitt gilt zusätzlich die folgende Annahme:

(A4) Für $i = 1, 2, 3$ und $j = 1, 2, 3$ gilt: $E(Y_{1ij}) = E(Y_{2ij}) = \mu_{ij}$.

Zunächst wird für jedes $i = 1, 2, 3$ und $j = 1, 2, 3$ die Zufallsvariable

$$T_{ij} := P_{ij} - \frac{1}{2}\ (R_{1ij} + R_{2ij}) \tag{1}$$

betrachtet, d. h. der Abstand des transformierten Zählergebnisses im Prüflabor vom Mittelwert der transformierten Zählergebnisse der beiden Referenzlabors. Unter der Elementarhypothese, daß der Erwartungswert des Zählergebnisses im Prüflabor mit dem Erwartungswert der Zählergebnisse in den Referenzlabors übereinstimmt, d. h.

$$H_0^{ij} \ :\ \ \lambda_{ij} = \mu_{ij} \quad , \tag{2}$$

gilt dann unter den Annahmen (A1) – (A4) für die Verteilung von T_{ij} approximativ

$$\sqrt{\frac{8}{3}} \cdot T_{ij} \overset{appr.}{\sim} N(0, 1) \tag{3}$$

bzw.

$$\frac{8}{3} \cdot T_{ij}^2 \overset{appr.}{\sim} \chi_1^2. \tag{4}$$

Mit Hilfe von (3) bzw. (4) läßt sich dann ein Test zur Überprüfung von H_0^{ij} bestimmen.

Aufgrund der Annahmen (A1) – (A3) sind die Zufallsvariablen T_{ij}, $i = 1, 2, 3$, $j = 1, 2, 3$, stochastisch unabhängig. Damit läßt sich für den Vergleich der neun Einzelzählergebnisse des Prüflabors mit den entsprechenden Zählergebnissen der beiden Referenzlabors bezüglich des Testproblems

$$H_0 :\ \ \lambda_{ij} = \mu_{ij}\ \forall i, j \qquad \text{gegen} \qquad H_1 :\ \ \exists i', j'\ \lambda_{i'j'} \neq \mu_{i'j'} \tag{5}$$

die folgende Teststatistik heranziehen:

$$T_1 := \frac{8}{3} \cdot \sum_{i=1}^{3} \sum_{j=1}^{3} T_{ij}^2, \tag{6}$$

für die unter der Globalhypothese H_0 aus (5) gilt:

$$T_1 \overset{appr.}{\sim} \chi_9^2 \, . \tag{7}$$

Basierend auf der Teststatistik T_1 läßt sich somit die folgende Entscheidungsregel für das Testproblem aus (5) formulieren:

$$\text{Verwerfe } H_0 \text{ zum Niveau } \alpha, \text{falls gilt:} \qquad T_1 > \chi^2_{9,1-\alpha} \, . \tag{8}$$

Da aufgrund der Wurzeltransformation die Varianz der Zufallsvariablen mit $\frac{1}{4}$ bekannt ist, ist der Test basierend auf T_1 der gleichmäßig beste invariante Test für H_0 (vgl. LEHMANN (1986), Seite 377), falls die beteiligten Zufallsvariablen normalverteilt sind. Da es sich im vorliegenden Fall um eine approximative Normalverteilung handelt, ist mit Hilfe einer Simulationsstudie das tatsächliche Niveau des Tests (8) mit dem nominellen Niveau für verschiedene Parameterkonstellationen verglichen worden. Dabei wurden für jeden Meßfilter und für jede Faserart die Realisationen mittels der Funktion RANPOI in SAS 6.10 unter Windows 3.11 erzeugt. Die Simulation des tatsächlichen Niveaus beruht für jede Parameterkonstellation auf 10000 Wiederholungen. Tabelle 1 zeigt für zehn verschiedene Parameterkonstellationen die simulierten tatsächlichen Niveaus für die nominellen Niveaus von $\alpha = 0.01, 0.05$ und 0.1.

Aus der Tabelle 1 wird ersichtlich, daß das Testverfahren das vorgegebene Testniveau recht gut einhält, wobei die simulierten Anteile für das fälschliche Verwerfen von H_0 in der Regel etwas größer als die vorgegebenen Anteile sind.

Neben den Einzelzählergebnissen wurden auch die Summen der Zählergebnisse für die drei Faserarten jeweils über die drei Filter betrachtet, um dem unterschiedlichen Schweregrad der Filter pro Labor Rechnung zu tragen. Es wurden also zusätzlich die Zufallsvariablen

$$X_{.j} := \sum_{i=1}^{3} X_{ij} \sim Po(\lambda_{.j}), \; j = 1, 2, 3 \; ,$$

und

$$Y_{l.j} := \sum_{i=1}^{3} Y_{2ij} \sim Po(\mu_{.j}), \; l = 1, 2, \; j = 1, 2, 3 \; ,$$

betrachtet. Mit Hilfe der oben beschriebenen Wurzeltransformation erhält man die transformierten Zählergebnisse $P_{.j}$ für das Prüflabor und $R_{l.j}$ für das l–te Referenzlabor, $l = 1, 2$, $j = 1, 2, 3$.

Tabelle 1:
Simulation des tatsächlichen Niveaus (in %) für die Testentscheidung (8) bei verschiedenen Parameterkonstellationen der Poisson–Verteilungen

Parameterkonstellation									Nominelles Niveau $\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
λ_{11}	λ_{21}	λ_{31}	λ_{12}	λ_{22}	λ_{32}	λ_{13}	λ_{23}	λ_{33}	χ^2–Test	χ^2–Test	χ^2–Test
3	3	3	3	3	3	3	3	3	0.79	4.45	9.43
4	4	4	4	4	4	4	4	4	1.06	5.33	10.34
5	5	5	5	5	5	5	5	5	1.25	5.57	10.56
10	10	10	10	10	10	10	10	10	1.15	5.24	10.25
3	3	3	4	4	4	5	5	5	1.01	5.17	10.13
3	3	3	5	5	5	10	10	10	1.06	5.12	10.05
3	4	5	6	7	8	9	10	11	1.10	5.26	10.26
5	5	5	6	6	6	7	7	7	1.23	5.51	10.46
5	5	5	10	10	10	15	15	15	1.13	5.36	10.41
10	10	10	15	15	15	20	20	20	1.11	5.19	10.17

Für das Testproblem

$$H_{0,\Sigma}: \quad \lambda_{.j} = \mu_{.j},\ j = 1,2,3 \qquad \text{gegen} \qquad H_{1,\Sigma}: \quad \exists j' \in \{1,2,3\}\ \lambda_{.j'} \neq \mu_{.j'} \tag{9}$$

erhält man mit

$$T_{.j} := P_{.j} - \frac{1}{2}\left(R_{1.j} + R_{2.j}\right) \tag{10}$$

in analoger Herleitung wie oben die Teststatistik

$$T_1^{\Sigma} := \frac{8}{3} \cdot \sum_{j=1}^{3} T_{.j}^2 \,, \tag{11}$$

für die unter der Hypothese $H_{0,\Sigma}$ aus (9) gilt:

$$T_1^{\Sigma} \overset{appr.}{\sim} \chi_3^2 \,. \tag{12}$$

3. Systematische Abweichungen zwischen den Referenzlabors

Da die Referenzlabors nicht gewährleisten können, daß sie 'richtig' bzw. 'identisch' messen, besteht die Möglichkeit, daß die Annahme (A4) nicht erfüllt ist und somit eine systematische Verzerrung zwischen den Zählergebnissen der Referenzlabors pro Filter und Faserart vorliegt, d. h. es gilt für $i = 1, 2, 3$, $j = 1, 2, 3$:

$$Y_{1ij} \sim Po\,(\mu_{1ij}) \qquad \text{und} \qquad Y_{2ij} \sim Po\,(\mu_{2ij}) \qquad \text{mit} \qquad \mu_{1ij} \neq \mu_{2ij}, \tag{13}$$

und daraus folgt dann

$$R_{1ij} \overset{appr.}{\sim} N\left(\sqrt{\mu_{1ij} + \frac{1}{8}}, \frac{1}{4}\right) \qquad \text{und} \qquad R_{2ij} \overset{appr.}{\sim} N\left(\sqrt{\mu_{2ij} + \frac{1}{8}}, \frac{1}{4}\right), \tag{14}$$

wobei mit den Erwartungswerten der Zufallsvariablen R_{lij}, $l = 1, 2$ auch der Steinersche Verschiebungssatz gilt (vgl. auch Anscombe (1948), Hartung et al. (1995), S. 350).

Angenommen, daß die systematischen Verzerrungen für alle Zählergebnisse bezüglich Filter und Faserart vorhanden sind, so muß die Globalhypothese H_0 aus (5) dahingehend geändert werden, daß die Erwartungswerte der Zählergebnisse des Prüflabors jeweils in einem Intervall liegen, dessen Grenzen jeweils durch die Erwartungswerte der Zählergebnisse der beiden Referenzlabors bestimmt werden. Da jedoch nicht bekannt ist, welches Referenzlabor bezüglich eines Filters und einer Faserart den größeren bzw. kleineren Erwartungswert besitzt, geht das Testproblem (5) über in das Testproblem

$$H_0^* : \lambda_{ij} \in [\xi_{1ij}, \xi_{2ij}] \;\forall i, j \qquad \text{gegen} \qquad H_1^* : \exists i', j' \; \lambda_{i'j'} \notin [\xi_{1i'j'}, \xi_{2i'j'}], \tag{15}$$

wobei $\xi_{1ij} = \min\{\mu_{1ij}, \mu_{2ij}\}$ und $\xi_{2ij} = \max\{\mu_{1ij}, \mu_{2ij}\}$.

Für die Zufallsvariable T_{ij}, $i = 1, 2, 3$, $j = 1, 2, 3$, aus (1) ergibt sich unter diesen Voraussetzungen die folgende Verteilungsaussage:

$$\sqrt{\frac{8}{3}} \cdot T_{ij} \overset{appr.}{\sim} N\left(\sqrt{\lambda_{ij} + \frac{1}{8}} - \frac{1}{2}\left(\sqrt{\mu_{1ij} + \frac{1}{8}} + \sqrt{\mu_{2ij} + \frac{1}{8}}\right), 1\right), \tag{16}$$

so daß für die Summe der quadrierten Zufallsvariablen gilt

$$\frac{8}{3} \cdot \sum_{i=1}^{3} \sum_{j=1}^{3} T_{ij}^2 \overset{appr.}{\sim} \chi^2_{9,\delta} \tag{17}$$

mit dem Nichtzentralitätsparameter

$$\delta = \sum_{i=1}^{3} \sum_{j=1}^{3} \left(\sqrt{\lambda_{ij} + \frac{1}{8}} - \frac{1}{2}\left(\sqrt{\mu_{1ij} + \frac{1}{8}} + \sqrt{\mu_{2ij} + \frac{1}{8}}\right)\right)^2. \tag{18}$$

Unter der Nullhypothese H_0^* aus (15) erreicht der Nichtzentralitätsparameter δ seinen maximalen Wert, falls der Erwartungswert des Prüflabors für alle Filter und alle Faserarten jeweils mit einem Erwartungswert der Referenzlabors übereinstimmt, d. h. es gilt $\lambda_{ij} = \mu_{1ij}$ oder $\lambda_{ij} = \mu_{2ij}$ für alle $i = 1, 2, 3$ und $j = 1, 2, 3$. Der maximale Wert des Nichtzentralitätsparameter δ unter H_0^* beträgt dann

$$\delta_{max} = \frac{1}{4} \cdot \sum_{i=1}^{3} \sum_{j=1}^{3} \left(\sqrt{\mu_{1ij} + \frac{1}{8}} - \sqrt{\mu_{2ij} + \frac{1}{8}} \right)^2 . \tag{19}$$

Eine Entscheidungsregel zur Überprüfung der Nullhypothese H_0^* lautet somit:

$$\text{Verwerfe } H_0^* \text{ zum Niveau } \alpha, \text{ falls gilt} \qquad \frac{8}{3} \cdot \sum_{i=1}^{3} \sum_{j=1}^{3} T_{ij}^2 > \chi^2_{9,\delta_{max},1-\alpha} . \tag{20}$$

Da der Nichtzentralitätsparameter δ_{max} unbekannt ist, muß dieser geeignet geschätzt werden. Dazu wird zunächst pro Filter und Faserart die Differenz der transformierten Zählergebnisse der beiden Referenzlabors betrachtet, d. h.

$$D_{ij} = R_{1ij} - R_{2ij} , \qquad , i = 1, 2, 3, \; j = 1, 2, 3 . \tag{21}$$

In der vorliegenden Situation gilt somit

$$\mathrm{E}(D_{ij}) = \left(\sqrt{\mu_{1ij} + \frac{1}{8}} - \sqrt{\mu_{2ij} + \frac{1}{8}} \right) \qquad \text{und} \qquad \mathrm{Var}(D_{ij}) = \frac{1}{2}, \tag{22}$$

so daß für die Statistik $\frac{1}{4} \cdot \sum_{i=1}^{3} \sum_{j=1}^{3} D_{ij}^2$ unter den Annahmen (A1)–(A3) gilt

$$\mathrm{E} \left(\frac{1}{4} \cdot \sum_{i=1}^{3} \sum_{j=1}^{3} D_{ij}^2 \right) = \frac{9}{8} + \delta_{max} . \tag{23}$$

Ein naheliegender Schätzer für den Nichtzentralitätsparameter δ_{max} ist somit

$$\hat{\delta}_{max} = \max \left\{ 0, \frac{1}{4} \sum_{i=1}^{3} \sum_{j=1}^{3} D_{ij}^2 - \frac{9}{8} \right\} . \tag{24}$$

Damit ergibt sich die folgende Entscheidungsregel:

$$\text{Verwerfe } H_0^* \text{ zum Niveau } \alpha, \text{ falls gilt:} \qquad \frac{8}{3} \cdot \sum_{i=1}^{3} \sum_{j=1}^{3} T_{ij}^2 > \chi^2_{9,\hat{\delta}_{max},1-\alpha} . \tag{25}$$

In der Tabelle 2 sind nun die simulierten tatsächlichen Niveaus für die beiden Tests mit den Entscheidungsregeln (8) und (25) bei Vorliegen von systematischer Abweichung zwischen den Zählergebnissen der Referenzlabors dargestellt. Dabei wurde die Simulation so

Tabelle 2:
Simulation des tatsächlichen Niveaus (in %) für die Testentscheidungen (8) und (25) bei Vorliegen von systematischen Abweichungen zwischen den Zählergebnissen in den Referenzlabors

Verzerrung Δ	Nominelles Niveau					
	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
	Test (8)	Test (25)	Test (8)	Test (25)	Test (8)	Test (25)
0.1	1.08	0.95	5.28	4.75	10.43	9.53
0.2	1.24	1.06	5.87	5.17	11.39	10.22
0.3	1.55	1.27	6.87	5.88	12.91	11.27
0.4	2.08	1.60	8.37	6.85	15.23	12.82
0.5	2.88	2.05	10.43	8.07	18.40	14.81
0.6	4.03	2.60	13.46	9.62	22.27	17.11
0.7	5.63	3.29	17.24	11.45	27.11	19.58
0.8	7.92	4.11	21.91	13.65	32.92	22.26
0.9	10.95	5.10	27.48	15.86	39.44	25.35
1.0	15.05	6.27	33.92	18.34	46.86	28.62

angelegt, daß die Verteilungen der Zählergebnisse für alle Faserarten und alle Filter aus dem Prüflabor mit denen aus einem Referenzlabor identisch sind und die Verteilungen der Zählergebnisse aus dem anderen Referenzlabor für alle Faserarten und alle Filter um den jeweils in der Tabelle 2 als Verzerrung Δ angegebenen Wert verschoben sind, d. h. der Wert Δ beinhaltet für alle Faserarten und Filter die 'ungünstigste' Konstellation zwischen den Referenzlabors. Dabei bezeichnet der Wert Δ den Abstand zwischen den Erwartungswerten der transformierten Zufallsvariablen R_{1ij} und R_{2ij}, $i = 1, 2, 3$, $j = 1, 2, 3$, d. h. $\Delta = \left|\sqrt{\mu_{1ij} + \frac{1}{8}} - \sqrt{\mu_{2ij} + \frac{1}{8}}\right|$.
Aufgrund der Konstruktion von (25) liegen die simulierten Niveaus von (25) stets unterhalb der simulierten Niveaus von (8) bezogen auf die Hypothese H_0^* aus (15). Jedoch überschreiten beide Tests erheblich das vorgegebene Niveau für die Überprüfung von H_0^*. Die Verbesserung durch den Test (25) ist nur bis zu einer Verzerrung von 0.3 zufriedenstellend. Beachtet man jedoch, daß eine Verzerrung von $\Delta = 0.3$ der transformierten Zählergebnisse bedeutet, daß die Erwartungswerte der nicht–tranformierten Zählergebnisse z. B. zwischen 2 und 3 bzw. zwischen 10 und 12 liegen, so scheint die Verbesserung durch die Entscheidungsregel (25) im praxisrelevanten Rahmen zu liegen.

Für die Überprüfung der Fasersummen bei Vorliegen von systematischen Abweichungen zwischen den Referenzlabors wird das Testproblem (9) wie folgt umgeschrieben:

$$H^*_{0,\Sigma} : \lambda_{.j} \in [\xi_{1.j}, \xi_{2.j}]\ \forall j \qquad \text{gegen} \qquad H^*_{1,\Sigma} : \exists j\ \lambda_{.j} \notin [\xi_{1.j}, \xi_{2.j}], \tag{26}$$

wobei $\xi_{1.j} = \min\{\mu_{1.j}, \mu_{2.j}\}$ und $\xi_{2.j} = \max\{\mu_{1.j}, \mu_{2.j}\}$.

Für die Teststatistik (11) gilt nun

$$\frac{8}{3} \cdot \sum_{j=1}^{3} T^2_{.j} \overset{appr.}{\sim} \chi^2_{3,\delta(\Sigma)} \tag{27}$$

mit dem Nichtzentralitätsparameter

$$\delta(\Sigma) = \sum_{j=1}^{3} \left(\sqrt{\lambda_{.j} + \frac{1}{8}} - \frac{1}{2} \left(\sqrt{\mu_{1.j} + \frac{1}{8}} + \sqrt{\mu_{2.j} + \frac{1}{8}} \right) \right)^2 . \tag{28}$$

Unter der Nullhypothese $H^*_{0,\Sigma}$ nimmt der Nichtzentralitätsparameter $\delta(\Sigma)$ seinen maximalen Wert an, falls für die Faserarten jeweils der Parameter aus dem Prüflabor mit einem der beiden Parameter aus den Referenzlabors übereinstimmt, d. h. es gilt unter $H^*_{0,\Sigma}$ aus (26)

$$\delta_{max}(\Sigma) = \frac{1}{4} \sum_{j=1}^{3} \left(\sqrt{\mu_{1.j} + \frac{1}{8}} - \sqrt{\mu_{2.j} + \frac{1}{8}} \right)^2 . \tag{29}$$

Analog zur Herleitung von (24) läßt sich in dieser Situation ein Schätzer für den Nichtzentralitätsparameter $\delta_{max}(\Sigma)$ bestimmen gemäß

$$\hat{\delta}_{max}(\Sigma) = \max \left\{ 0, \frac{1}{4} \cdot \sum_{j=1}^{3} D^2_{.j} - \frac{3}{8} \right\} \tag{30}$$

mit $D_{.j} := R_{1.j} - R_{2.j}$.

Im folgenden Teil dieses Abschnitts wird zusätzlich zur systematischen Verzerrung der Zählergebnisse der Referenzlabors pro Filter und Faserart eine mögliche positve Korrelation dieser Zählergebnisse im Vergleich zum Zählergebnis des Prüflabors in Betracht gezogen. Bezeichne dazu ρ diese Korrelation zwischen den transformierten Zählergebnissen R_{1ij} und R_{2ij}, $i = 1, 2, 3$, $j = 1, 2, 3$, dann gilt bei Benutzung der Zufallsvariablen T_{ij} aus (1) die folgende Verteilungsaussage:

$$\frac{8}{3 + \rho} \cdot \sum_{i=1}^{3} \sum_{j=1}^{3} T^2_{ij} \overset{appr.}{\sim} \chi^2_{9,\delta} \tag{31}$$

mit δ aus (18).

Zur Überprüfung von H_0^* aus (15) müssen somit adäquate Schätzungen für die Korrelation ρ und den Nichtzentralitätsparameter δ bestimmt werden. Für die Zufallsvariable D_{ij} aus (21), $i = 1,2,3,\ j = 1,2,3$, hängen bei Vorliegen von Korrelation und systematischer Abweichung jedoch Erwartungswert und Varianz von den beiden Parametern ab. Es gilt nämlich

$$D_{ij} \overset{appr.}{\sim} N\left(\sqrt{\mu_{1ij} + \frac{1}{8}} - \sqrt{\mu_{2ij} + \frac{1}{8}}, \frac{1}{2}\,(1-\rho)\right), \tag{32}$$

d. h.

$$\mathrm{E}\left(\sum_{i=1}^{3}\sum_{j=1}^{3} D_{ij}^2\right) = 9 \cdot \frac{1-\rho}{2} + 4 \cdot \delta_{max} \tag{33}$$

mit δ_{max} aus (19), so daß mit Hilfe dieser Zufallsvariablen keine geeignete Schätzung der beiden unbekannten Parameter durchgeführt werden kann.

Da die transformierten Zählergebnisse R_{1ij} und R_{2ij} gleiche Varianzen besitzen, gilt für die Kovarianz zwischen T_{ij} und D_{ij}, $i = 1,2,3,\ j = 1,2,3$,

$$\mathrm{Cov}(T_{ij}, D_{ij}) = -\frac{1}{2}\Big(\mathrm{Var}(R_{1ij}) - \mathrm{Var}(R_{2ij})\Big) = 0,$$

so daß unter Normalverteilungsapproximation davon ausgegangen werden kann, daß die Zufallsvariablen stochastisch unabhängig sind. Dies bedeutet, daß auch die Zufallsvariablen

$$T_1 = \frac{8}{3} \cdot \sum_{i=1}^{3}\sum_{j=1}^{3} T_{ij}^2 \qquad \text{und} \qquad D = \sum_{i=1}^{3}\sum_{j=1}^{3} D_{ij}^2 \tag{34}$$

als stochastisch unabhängig angesehen werden können. Somit gilt für den Quotienten dieser beiden Zufallsvariablen bei Vorliegen von Korrelation, aber *keinen* systematischen Abweichungen unter H_0 aus (5)

$$T_2 := \frac{\dfrac{4}{3+\rho} \cdot \sum_{i=1}^{3}\sum_{j=1}^{3} T_{ij}^2}{\dfrac{1}{1-\rho} \cdot \sum_{i=1}^{3}\sum_{j=1}^{3} D_{ij}^2} \overset{appr.}{\sim} F_{9,9}. \tag{35}$$

Der Nenner in (35) mißt in gewissem Maße die mögliche Verzerrung zwischen den Zählergebnissen der Referenzlabors, so daß im folgenden die F–Statistik (35) dahingehend modifiziert wird, daß Korrelation und systematische Abweichung hinreichend berücksichtigt werden. Die Doppelsumme im Nenner von (35) wird klein, falls kaum systematische Abweichungen, aber Korrelationen auftreten. Der Vorfaktor wirkt diesem Verhalten entgegen. Um dieses Phänomen geeignet abfangen zu können, soll der Nenner durch eine positive reelle Zahl nach unten beschränkt werden. Der Zähler der F–Teststatistik (35) soll dahingehend

Tabelle 3:
Simulation des tatsächlichen Niveaus (in %) für die Testentscheidung (38) bei verschiedenen Konstellationen der Verzerrung Δ und der Korrelation ρ zum nominellen Niveau von $\alpha = 0.05$

Verzerrung Δ	0	0.1	0.2	0.3	0.4	0.5
Korrelation ρ						
0	0.96	0.95	0.87	0.79	0.71	0.60
0.10	1.48	1.45	1.32	1.18	1.05	0.84
0.25	2.73	2.66	2.49	2.22	1.85	1.41
0.30	3.36	3.29	3.09	2.68	2.21	1.65
0.40	4.96	4.89	4.55	3.95	3.14	2.31
0.50	7.11	7.01	6.64	5.79	4.55	3.23
0.75	13.23	13.44	13.81	13.35	11.20	7.33
0.90	14.73	15.15	16.40	18.37	18.65	13.10

modifiziert werden, daß eine Schätzung der Korrelation einfließt. Aus (33) ergibt sich mit D aus (34) die Gleichung

$$\rho = 1 - \frac{2}{9} \cdot \Big(E(D) - 4 \cdot \delta_{max} \Big),$$

so daß ein möglicher Schätzer für ρ gegeben ist durch

$$\tilde{\rho} = \max\{0, 1 - \frac{1}{9} \cdot D\}. \tag{36}$$

Die Wahl einer unteren Schranke für den Nenner der F–Teststatistik wird motiviert über ein Quantil der Verteilung der Nennerstatistik (für $\rho = 0$ und $\delta = 0$), wobei die geeignete Wahl eines solchen Quantils gewisser Erfahrungswerte bedarf. Im folgenden wird das 0.1–Quantil der Nennerstatistik, welches der Hälfte des 0.1–Quantils der χ^2–Verteilung mit neun Freiheitsgraden entspricht, als untere Grenze gewählt und der Korrelationskoeffizient dort gleich Null gesetzt, so daß die modifizierte Teststatistik (35) das folgende Aussehen bekommt:

$$T_{2,\tilde{\rho}} := \frac{\frac{4}{3+\tilde{\rho}} \cdot \sum_{i=1}^{3} \sum_{j=1}^{3} T_{ij}^2}{\max\{2.084, D\}} \, . \tag{37}$$

Die Testentscheidung läßt sich somit wie folgt fomulieren:

$$\text{Verwerfe } H_0^* \text{ aus (15) zum Niveau } \alpha \text{, falls gilt} \qquad T_{2,\tilde{\rho}} > F_{9,9,1-\alpha} \, . \tag{38}$$

In der Tabelle 3 sind nun für verschiedene Konstellationen von systematischen Abweichungen und Korrelationen zwischen den transformierten Zählergebnissen der Referenzlabors die simulierten tatsächlichen Niveaus der Testentscheidung (38) für das nominelle Niveau von $\alpha = 0.05$ dargestellt. Man sieht, daß für Korrelationen bis zu $\rho = 0.3$ für alle betrachteten Verzerrungen der Test recht konservativ ist, während er für $\rho = 0.4$ bei geringer Verzerrung bis zu $\Delta = 0.2$ und für $\rho = 0.5$ bei Verzerrungen um $\Delta = 0.3$ bis $\Delta = 0.4$ das vorgegebene Niveau recht gut einhält. Bei den weiteren betrachteten Konstellationen ist der Test häufig sehr antikonservativ. Für die konkrete Anwendung hält der Test somit das nominelle Niveau zufriedenstellend ein, falls moderate Korrelationen und moderate systematische Abweichungen zwischen den transformierten Zählergebnissen der Referenzlabors vorliegen.

Für die Überprüfung der Fasersummen bei Vorliegen von Korrelation und systematischer Abweichung bezüglich der transformierten Zählergebnisse aus den Referenzlabors läßt sich in analoger Argumentation wie bei der Herleitung von (35) die Teststatistik T_2^{Σ} bestimmen mit

$$T_2^{\Sigma} = \frac{\dfrac{4}{3+\rho^*} \cdot \sum\limits_{j=1}^{3} T_{\cdot j}^2}{\dfrac{1}{1-\rho^*} \cdot \sum\limits_{j=1}^{3} D_{\cdot j}^2} \overset{appr.}{\sim} F_{3,3} \ , \tag{39}$$

wobei ρ^* die Korrelation zwischen den Summen der Zählergebnisse aus den beiden Referenzlabors bezeichne. Zur Überprüfung von $H_{0,\Sigma}^*$ muß die Teststatistik T_2^{Σ} in ähnlicher Weise wie die Teststatistik T_2 modifiziert werden.

Das Hauptaugenmerk bei der Konstruktion obiger Tests liegt natürlich in der Kontrolle des Fehlers 1. Art, damit das Prüflabor nicht ungerecht beurteilt wird. Der Test basierend auf $T_{2,\tilde{\rho}}$ ist jedoch für moderate Korrelationen und Verzerrungen recht konservativ, so daß sich berechtigterweise die Frage stellt, ob dieser Test überhaupt genügend Macht besitzt, um Unterschiede zwischen den Referenzlabors und dem Prüflabor entdecken zu können. Daher wird im folgenden für einige Parameterkonstellationen von ρ und Δ die Gütefunktion des Tests basierend auf $T_{2,\tilde{\rho}}$ aus (37) an einigen Stützstellen simuliert, deren Ergebnisse für $\rho = 0,\ 0.2,\ 0.5$ und $\Delta = 0,\ 0.2,\ 0.4$ der Tabelle 4 zu entnehmen sind. Die betrachteten Stützstellen der entsprechenden Gütefunktionen sind dabei so zu interpretieren, daß sie die Abweichung des Erwartungswertes des transformierten Zählergebnisses des Prüflabors vom Mittelwert der Erwartungswerte der transformierten Zählergebnisse der beiden Referenzlabors angeben, d. h. das simulierte tatsächliche Testniveau befindet sich demnach an der Stelle $\Delta/2$.

Der Tabelle 4 ist nun zu entnehmen, daß bei tatsächlichen Testniveaus bis zu 2 % bei nominellem Niveau 5 % die Gütefunktionen der entsprechenden Tests nicht besonders stark ansteigen. Sogar wenn alle transformierten Zählergebnisse des Prüflabors jeweils aus einer Verteilung stammen, deren Erwartungswert um den Wert 1 größer ist als der Mittelwert

Tabelle 4:
Simulierte Werte (in %) der Gütefunktion des Tests aus (38) mit nominellem Niveau $\alpha = 0.05$ bei verschiedenen Konstellationen der Korrelation ρ und Verzerrung Δ an ausgewählten Stützstellen
(fett gedruckte Werte geben jeweils das simulierte tatsächliche Testniveau an)

		Stützstellen der Gütefunktion							
ρ	Δ	0	0.1	0.2	0.3	0.4	0.5	0.75	1
0	0	**0.96**	1.10	1.59	2.69	4.70	8.03	24.88	51.86
0.2	0	**2.22**	2.52	3.48	5.47	8.92	14.40	37.67	66.74
0.5	0	**7.11**	7.89	10.27	14.64	21.54	31.29	62.88	88.00
0	0.2	0.77	**0.89**	1.30	2.20	3.93	6.74	21.63	49.96
0.2	0.2	1.79	**2.02**	2.84	4.52	7.41	11.99	32.77	61.06
0.5	0.2	5.98	**6.52**	8.47	12.31	18.39	27.03	56.67	83.32
0	0.4	0.41	0.47	**0.71**	1.25	2.28	4.00	13.80	34.31
0.2	0.4	0.93	1.05	**1.52**	2.47	4.16	6.87	20.96	45.50
0.5	0.4	3.01	3.86	**4.47**	6.64	10.29	15.68	38.30	66.61

der transformierten Zählergebnisse der Referenzlabors, entdeckt der Test in höchstens 2/3 der Fälle diesen Unterschied. Dies bedeutet, daß bei geringen Korrelationen und geringen Verzerrungen das Prüflabor recht wohlwollend beurteilt wird. Liegt das tatsächliche Testniveau im Bereich von 5 %, so entdeckt der Test Abweichungen des Prüflabors mit zufriedenstellender Wahrscheinlichkeit.

4. Entscheidungsfolge

Basierend auf den Resultaten der Abschnitte 2 bzw. 3 werden für den Vergleich der Zählergebnisse des Prüflabors mit den Zählergebnissen der beiden Referenzlabors insgesamt vier Tests durchgeführt, jeweils für die langen und kurzen Fasern ein Test mit den einzelnen Zählergebnissen als auch ein Test mit den Zählergebnissen der Fasersummen. Die Ergebnisse dieser vier Tests müssen nun zu einem sinnvollen Gesamtergebnis kombiniert werden. Bei allen vier Tests neigen die Teststatistiken unter der Alternative zu größeren Werten, so daß die P–Werte der einzelnen Tests auf einheitliche Art geeignet zusammengefaßt werden können. Dabei ist jedoch zu beachten, daß die beiden Teststatistiken, die jeweils die Zählergebnisse für die kurzen bzw. die langen Fasern bewerten, nicht unabhängig sind.

Bezeichnen im folgenden p_L und p_L^Σ die P–Werte aus den Tests für die langen Fasern sowie p_K und p_K^Σ die entsprechenden P–Werte aus den Tests für die kurzen Fasern, so werden aufgrund der Abhängigkeiten der Teststatistiken zunächst die P–Werte für die langen und kurzen Fasern getrennt zusammengefaßt. Dabei wird der P–Wert aus dem Test zur Überprüfung der Zählergebnisse der einzelnen Fasern doppelt so stark gewichtet wie der P–Wert aus dem Test zur Überprüfung der Fasersummen. Die zusammengefaßten P–Werte für die beiden Faserklassen ergeben sich in Anlehnung an die gewichtete 'Inverse Normal–Methode', vgl. HEDGES / OLKIN (1985), Seite 39, als gewichtete Mittelung der Probits der einzelnen P–Werte zu

$$1 - p_{L,ges} = \Phi\left\{\frac{2}{3}\cdot\Phi^{-1}(1-p_L) + \frac{1}{3}\cdot\Phi^{-1}(1-p_L^\Sigma)\right\} \tag{40}$$

bzw.

$$1 - p_{K,ges} = \Phi\left\{\frac{2}{3}\cdot\Phi^{-1}(1-p_K) + \frac{1}{3}\cdot\Phi^{-1}(1-p_K^\Sigma)\right\} \quad , \tag{41}$$

wobei Φ die Verteilungsfunktion der Standardnormalverteilung bezeichnet.

Die nach (40) und (41) bestimmten P–Werte sind unabhängig und können demgemäß meta–analytisch kombiniert werden. Da die kurzen Fasern in der Regel schwerer zu finden sind als die langen, wird bei der meta–analytischen Zusammenfassung der P–Wert $p_{K,ges}$ nur halb so stark gewichtet wie der P–Wert $p_{L,ges}$. Mit der originalen gewichteten 'Inversen Normal–Methode' ergibt sich somit ein Gesamt–P–Wert gemäß

$$1 - p_{ges} = \Phi\left\{\frac{3}{\sqrt{5}}\cdot\left(\frac{1}{3}\cdot\Phi^{-1}(1-p_{K,ges}) + \frac{2}{3}\cdot\Phi^{-1}(1-p_{L,ges})\right)\right\}. \tag{42}$$

Die Entscheidung, ob das Prüflabor die Richtlinie erfolgreich umgesetzt hat, wird durch den Vergleich des in (42) bestimmten Gesamt–P–Werts mit einem vorgegebenen Niveau α getroffen.
Ist $p_{ges} \geq \alpha$, so erfüllt das Prüflabor die Kriterien. Ist jedoch $p_{ges} < \alpha$, so mißt das Prüflabor signifikant anders als die beiden Referenzlabors und erfüllt somit nicht die Kriterien zur korrekten Umsetzung der Richtlinie VDI 3492 Blatt 1.

Anmerkung:

Die Autoren danken dem Gutachter der Zeitschrift ALLGEMEINES STATISTISCHES ARCHIV für seine Verbesserungsvorschläge zu einer früheren Fassung des Manuskripts.

Literatur:

Anscombe, F. J. (1948): The transformation of Poisson, Binomial and Negative–Binomial data; Biometrika, 35, 246–254.

Hartung, J. / Elpelt, B. / Klösener, K.–H. (1995): Statistik. Lehr- und Handbuch der angewandten Statistik; München: Oldenbourg Verlag, 10. Auflage.

Hedges, L. V. / Olkin, I. (1985): Statistical Methods for Meta–Analysis; Orlando: Academic Press.

Höfert, N. / König, R. / Grefen, K. / Rödelsperger, K. / Teichert, U. (1996a): Messen faserförmiger Partikel — Erster Ringversuch nach Richtlinie VDI 3492 Blatt 1, Teil 1; Gefahrstoffe – Reinhaltung der Luft, 56, 11–15.

Höfert, N. / König, R. / Grefen, K. / Rödelsperger, K. / Teichert, U. (1996b): Messen faserförmiger Partikel — Erster Ringversuch nach Richtlinie VDI 3492 Blatt 1, Teil 2; Gefahrstoffe – Reinhaltung der Luft, 56, 63–68.

Lehmann, E. L. (1986): Testing Statistical Hypotheses; New York: Wiley, Second Edition.

Autorenanschrift:

Prof. Dr. Joachim Hartung
Dr. Guido Knapp
Lehrstuhl Statistik mit Anwendungen
im Bereich der Ingenieurwissenschaften
Fachbereich Statistik der Universität Dortmund
Vogelpothsweg 87

D–44221 Dortmund

Email: hartung@omega.statistik.uni-dortmund.de
knapp@omega.statistik.uni-dortmund.de