Originalarbeit

Anforderungen an die Präzision und Reproduzierbarkeit von Testverfahren bei der Wirksamkeitsbeurteilung chemischer Desinfektionsverfahren

Marvin Rausch*[1], Felix Droop[1], Martin Exner[2], Jürgen Gebel[1], Carola Ilschner[1], Kira Roesch[1], Thomas Selhorst[3], Nico T. Mutters[1]

[1]Institut für Hygiene und Öffentliche Gesundheit des Universitätsklinikums Bonn; [2]Universität Bonn; [3]Bundesinstitut für Risikobewertung, Berlin

Korrespondierender Autor:
Dr. Marvin Rausch
Universitätsklinikum Bonn
Institut für Hygiene und Öffentliche Gesundheit
Desinfektionsmitteltestung
(Gebäude 63)
Venusberg-Campus 1
53127 Bonn

E-Mail: marvin.rausch@ukbonn.de

Interessenkonflikt:
Die Autoren erklären, dass kein Interessenkonflikt im Sinne der Richtlinien des International Committee of Medical Journal Editors (ICMJE) besteht.

Zitierweise:
Rausch M, Droop F, Exner M, Gebel J, Ilschner C, Roesch K, Selhorst T, Mutters NT. Anforderungen an die Präzision und Reproduzierbarkeit von Testverfahren bei der Wirksamkeitsbeurteilung chemischer Desinfektionsverfahren. HygMed 2024; 49(4): D10–D16.

Manuskriptdaten:
Eingereicht: 28.2.2024
Überarbeitete Version angenommen: 2.4.2024

Zusammenfassung

Einleitung: Für die Wirksamkeitsbeurteilung chemischer Desinfektionsverfahren sind präzise und reproduzierbare Testverfahren unerlässlich. Die Kommission für Krankenhaushygiene und Infektionsprävention beim Robert Koch-Institut (KRINKO) empfiehlt für die Beurteilung der sicheren Wirksamkeit eines Desinfektionsverfahrens für infektionssensible Bereiche zwei Prüfberichte aus voneinander unabhängigen akkreditierten Laboren sowie die zugehörigen Gutachten. In der vorliegenden Studie wird die statistische Relevanz dieser Anforderungen, die auch vom Verbund für Angewandte Hygiene (VAH) an die Zertifizierung von Desinfektionsmitteln gestellt wird, durch die umfassende Auswertung eines Ringversuches untersucht.

Material und Methoden: Die Untersuchung basiert auf Daten des VAH-Ringversuchs 2023-01 zur Bewertung der bakteriziden Wirksamkeit von Glutaraldehyd gegen *Staphylococcus aureus* mit dem quantitativen Suspensionsversuch nach DIN EN 13727:2015. Die umfassende statistische Analyse, einschließlich einer einfaktoriellen Varianzanalyse (ANOVA) und Simulationen zielte darauf ab, die Variabilität und Konsistenz der Testergebnisse über verschiedene Labore hinweg zu ermitteln.

Ergebnisse: Die Ergebnisse zeigen signifikante intralaboratorielle und auch interlaboratorielle Schwankungen. Die Simulationsstudie bestätigt, dass die Einbindung der Ergebnisse eines zweiten Labors die Genauigkeit der Ergebnisbewertung verbessert und das Risiko einer Fehleinschätzung erheblich reduziert. Dieser Effekt ist vor allem in der Simulation im sogenannten „intermediär wirksamen“ Bereich sichtbar: Zieht man nur die Entscheidung eines Labors heran, so wird das Produkt in 32,3% der Fälle als wirksam deklariert. Zieht man ein zweites Labor heran, dann beträgt diese Wahrscheinlichkeit nur noch 10,3%. Das daraus resultierende Odds-Ratio (OR) liegt bei 0,24. Somit wird die Wahrscheinlichkeit, das Produkt als wirksam zu deklarieren um 76% gesenkt.

Schlussfolgerungen: Die Ergebnisse stärken die Forderung nach zwei unabhängigen Prüfberichten für die Beurteilung der Wirksamkeit eines Desinfektionsverfahren und bestätigen die zentrale Rolle von Ringversuchen für die Qualitätssicherung im Bereich der Desinfektionsmittelprüfung. Eine hohe Varianz zwischen den Laboratorien zeigte sich insbesondere im intermediären Wirkbereich. Die statistische Analyse weiterer Ringversuche im Hinblick auf die Signifikanz der Einbeziehung von Ergebnissen aus einem zweiten Prüflabor sollte zur Bestätigung dieser Aussagen folgen.

Schlüsselwörter
- Desinfektionsmittelprüfung
- ANOVA
- Ringversuch
- Prüfbericht
- Wirksamkeit

Abstract

Requirements for the precision and reproducibility of test methods in the efficacy assessment of chemical disinfection processes

Introduction: Assessing the efficacy of chemical disinfectants requires precise and reproducible test procedures. The

© mhp Verlag GmbH 2024

German Commission on Hospital Hygiene and Infection Prevention at the Robert Koch Institute recommends two test reports from independent accredited laboratories with the associated expert reports for assured effectiveness of disinfectants used in areas which harbor a special risk for infections ("infection-sensitive areas"). In this paper, the statistical relevance of these requirements which are also stipulated by the Association for Applied Hygiene (VAH) for certification of disinfectants, is assessed by a comprehensive analysis of a ring trial.

Material and methods: The study is based on data from the ring trial 2023-01 to evaluate the bactericidal efficacy of glutaraldehyde against *Staphylococcus aureus* using the quantitative suspension test according to DIN EN 13727:2015. The comprehensive statistical analysis, including a one-way analysis of variance (ANOVA) and simulations, aimed to determine the variability and consistency of the test results across different laboratories.

Results: The results show significant intra-laboratory and inter-laboratory variations. The simulation study confirms that including the results from a second laboratory improves the accuracy of the result assessment and significantly reduces the risk of misjudgement. This effect is particularly visible in the simulation of the so-called "intermediary effective" range: if the decision of only one laboratory is applied, the product is declared effective in 32.3% of cases. If a second laboratory is consulted, this probability is only 10.3%. The resulting odds ratio (OR) is 0.24. Thus, the probability of declaring the product as effective is reduced by 76%.

Conclusions: The results strengthen the requirement for two independent test reports to assess the efficacy of a disinfection procedure and confirm the critical role of ring trials for quality assurance in disinfectant testing. A high variance between the laboratories was particularly evident in the intermediate efficacy range. The statistical analysis of further ring trials with regard to the significance of including results from a second test laboratory should follow to confirm these findings.

Keywords: *Disinfectant testing, ANOVA, ring trial, test report, efficacy*

Einleitung

Der Verbund für Angewandte Hygiene (VAH) fordert bei der Wirksamkeitsbeurteilung eines Desinfektionsverfahrens die Vorlage von zwei Prüfberichten aus zwei unabhängigen akkreditierten Prüflaboratorien. Desinfektionsmitteltests sind mikrobiologische Tests, die grundsätzlich einer gewissen biologischen statistischen Streuung unterliegen. Um das Maß solcher Abweichungen präzise bestimmen zu können, werden seit mehr als 15 Jahren VAH-Ringversuche mit einer großen Zahl erfahrener Laboratorien unter gleichen Prüfvorgaben durchgeführt und die Ergebnisberichte veröffentlicht. Auf europäischer Ebene wurde dies z.B. für die mandatierten Normen EN 13727 [1], EN 13624 [2] und die EN 14348 [3] unter dem Dach des CEN TC 216 – dem Technischen Komitee für chemische Desinfektionsmittel und Antiseptika des Europäischen Komitees für Normung – durchgeführt.

In der EN 14885 [4], der Norm, die die Anwendung und Interpretation der Normen im CEN TC 216 abdeckt, heißt es unter 5 a) und b), Präzision der Prüfverfahren (Wiederholungen), dass die Prüfverfahren in Ringversuchen überprüft wurden und die Präzision bzw. die jeweils festgelegten Bestehensgrenzen eine Bewertung des Produktes ermöglichen, ohne eine Bewertung der Ungenauigkeit der jeweiligen Testung durchführen zu müssen.

Dennoch werden zur Erfüllung der „bestmöglichen Praxis" folgende Maßnahmen zur Verbesserung der Präzision empfohlen:

- Vorhaltung eines Qualitätssicherungssystems (z.B. EN ISO/IEC 17025) und regelmäßige Teilnahme an Ringversuchen durch die Prüflaboratorien
- Wiederholung der Tests im jeweiligen Prüflabor
- Einbeziehung von internen Standards – also laborinternen Standards mit Referenzsubstanzen
- Durchführung der Prüfung in einem zweiten und/oder dritten Labor

In der EN 13727 [1] wird unter Anhang E (informativ) zur Präzision der Prüfergebnisse nach Auswertung eines Ringversuches im Zeitraum von 2000 bis 2002 mit zehn Laboratorien festgestellt, dass im günstigsten Fall vier Wiederholungen und im ungünstigsten Fall sechs oder mehr Wiederholungen in einem Labor erforderlich sind, um eine Präzision der Reduktion von ± 1 lg zu erreichen.

In der EN 13624 [2] heißt es im Anhang E (informativ) zur Präzision der Prüfergebnisse, dass nach Auswertung des Ringversuches im Zeitraum von 2001 bis 2003 mit 8 Laboratorien mit dem Prüforganismus *Aspergillus brasiliensis*, dem Wirkstoff Glutaraldehyd bei einer Einwirkzeit von 15 min, im günstigsten Fall vier Wiederholungen und im ungünstigsten Fall fünf oder mehr Wiederholungen in einem Labor erforderlich sind, um eine Präzision der Reduktion von ± 1 lg zu erreichen.

In den Normen EN 14348, EN 14561, EN 14562 und EN 14563 [3, 5–7], die auch im Rahmen dieser Ringversuche für harmonisierte Normen untersucht wurden, finden sich keine Angaben zur Präzision in den jeweiligen Anhängen.

Es bleibt festzuhalten, dass in den einzigen beiden statistischen Auswertungen, die sich in den EN-Normen finden, die Berechnungen belegen, dass im besten Fall eine 4-malige Wiederholung der Prüfung in einem Prüflabor erforderlich ist, um eine Präzision des Ergebnisses mit einer Standardabweichung von ± 1 lg zu generieren.

Der VAH hat sich zur Absicherung der Prüfergebnisse für die Reproduktion der Versuche im Labor selbst auf die praxisnahen Testverfahren – im europäischen Kontext als Phase-2-Stufe-2-Tests beschrieben – beschränkt, da diese Versuche praxisnähere Bedingungen als die Phase-2-Stufe-1-Tests abbilden.

Der VAH hat sich für die in der EN 14885 unter 5 a) und 5 b) vorgeschlagenen Maßnahmen (siehe Auflistung oben) zur Absicherung des Ergebnisses entschieden. Der VAH fordert folglich Prüfergebnisse der Phase 2 Stufe-1 und Phase 2 Stufe-2 (mit Reproduktionen der Phase 2-Stufe-2-Tests) von akkreditierten Laboratorien (z.B. gemäß DAkkS), die sich regelmäßig an Ringversuchen beteiligen und die Reproduktion der Prüfergebnisse in einem weiteren unabhängigen und akkreditierten Prüflabor. Die Akkreditierung, z.B. nach z.B. EN ISO/IEC 17025, garantiert eine vom Auftraggeber unabhängige Prüfung.

© mhp Verlag GmbH 2024

Diese Forderung nach zwei unabhängigen Prüfberichten von akkreditierten Laboratorien wird im Rahmen der Bewertung von Bioziden von der ECHA nicht umgesetzt (Guidance on the BPR: Volume II Parts B+C – Version 6,0 August 2023) [8]. Im Rahmen des Bewertungsverfahrens durch die ECHA wird die Wirksamkeit in der Regel auf Grundlage eingereichter Prüfberichte aus einem Labor vorgenommen. Eine Akkreditierung des jeweiligen Prüflabors wird unter Punkt 3.1.1. nicht gefordert, es reicht eine Bescheinigung, dass ein Qualitätsmanagementsystem installiert ist. Auch wird eine Unabhängigkeit des Prüflabors vom Hersteller des Desinfektionsmittels nicht vorgeschrieben. Theoretisch kann also der Antragsteller sein eigenes Produkt prüfen und diese Resultate für das Bewertungsverfahren einreichen. Auf eine Reproduktion der Testergebnisse im jeweiligen Prüflabor wird zwar unter 5.4.0.4.3 Data requirements – Repetitions hingewiesen, aber nicht explizit gefordert.

In dieser Arbeit soll anhand eines aktuellen Ringversuchs exemplarisch dargelegt werden, dass die Forderung nach zwei Prüfberichten zur Absicherung der Wirksamkeit eines chemischen Desinfektionsverfahrens statistisch zu begründen ist.

Material und Methoden

Datensatz:

Im VAH-Ringversuch 2023-01 wurde die bakterizide Wirksamkeit des Testprodukts A mit dem quantitativen Suspensionsversuch gegen *Staphylococcus aureus* nach DIN EN 13727:2015 zur Beurteilung der Laborleistung geprüft [9]. Das Prüfprodukt wurde aus einer einzelnen Charge für alle Teilnehmer im Mai 2023 vom VAH zur Verfügung gestellt. Ziel des Versuchs war es, die Reduktion für das Testprodukt A (Sigma-Aldrich Glutaraldehyd (GA) 50 Gew.-% wässrige Lösung mit Ref. 340855, CAS-Nummer 111-30-8, DMT 2023-016) bei drei verschiedenen Glutaraldehydkonzentrationen – 0,10% / 0,05% / 0,01% – 5 min – unter den vorgegebenen Testbedingungen zu ermitteln. Die laborübergreifende Reproduzierbarkeit des EN 13727-Testprotokolls und die laborübergreifende Reproduzierbarkeit der ermittelten bakteriziden Aktivität wurden überprüft. Basierend auf Vorversuchen des VAH-Referenzlabors sollte eine „unwirksame“ Konzentration (0,01% – 5 min), eine Konzentration im sog. „intermediären“ Bereich (0,05% – 5 min), der an der Wirksamkeitsgrenze liegt, und eine Konzentration in einem „wirksamen“ Bereich (0,1% – 5 min) gefunden werden. Insgesamt haben sich 31 Laboratorien, von denen drei nicht akkreditiert waren, beteiligt.

Methoden:

Um zu prüfen, ob das Labor einen signifikanten Einfluss auf die Reduktionen hat, wurde eine einfaktorielle Varianzanalyse (ANOVA) durchgeführt. Hier werden die Labormittelwerte untereinander verglichen und mit dem F-Test (α=5%) auf signifikante Unterschiede geprüft. Der Test wird als signifikant bewertet, wenn mindestens zwei Labore einen signifikanten Unterschied aufweisen.

Danach wurde der Anteil der Labore mit einer Aussage **wirksam** (lg-R > 5) und **nicht wirksam** (lg-R < 5) ermittelt. Es wurden anschließend alle Aussagen der Labore paarweise miteinander verglichen und notiert, ob zwei Labore die gleiche Aussage treffen oder nicht.

Mit Hilfe der paarweisen Vergleiche erhält man einen binomialverteilten Datensatz, der eine „1“ enthält, wenn die Labore zu einem ungleichen Ergebnis kommen und eine „0“ bei einem gleichen Ergebnis.

Um die Veränderung der Aussage bei Hinzunahme eines zweiten Labors zu analysieren, wurde die folgende Simulation 10.000-mal durchgeführt:

1. Wähle zufällig ein Labor und notiere die Aussage
2. Wähle zufällig ein zweites Labor und vergleiche die Aussage mit 1.
3. Unter 1. wird die Entscheidung des Labors (**wirksam** oder **unwirksam**) notiert und unter 2. wird nur **wirk-**

Tabelle 1: Statistische Parameter für die Reduktion von *Staphylococcus aureus* mit Glutaraldehyd nach EN 13727:2015.

Quantitativer Suspensionsversuch (EN 13727:2015)
Reduktion von *Staphylococcus aureus* - geringe Belastung

Produkt	Glutaraldehyd		
Konzentration / Zeitrelation	0,01% - 5 min	0,05% - 5 min	0,1% - 5 min
Anzahl Teilnehmer	28	31	28
Anzahl für Teilnehmer mit quantitativen Werten	28	31	28
Mean lg Reduktion ± 95% CI*	0,42 ± 0,15	4,27 ± 0,32	5,31 ± 0,09
Reproducibility SD	0,42	0,93	0,26
Repeatability SD	0,13	0,34	0,11

*CI: Konfidenzintervall; Repeatability SD: Standardabweichung der Wiederholbarkeit; Reproducibility SD: Standardabweichung der Reproduzierbarkeit

© mhp Verlag GmbH 2024

sam notiert, wenn beide Labore die Entscheidung haben, dass das Produkt **wirksam** ist. Sonst wird es als **unwirksam** deklariert.

4. Füge die Entscheidung nach folgendem Muster in die Vierfeldertafel ein:

Entscheidung	beide Labore	ein Labor
wirksam	A	b
unwirksam	C	d

Nachdem die Simulation durchgeführt wurde, erhält man eine Vierfeldertafel mit N=10.000 Einträgen. Für diese Vierfeldertafel wird dann ein Odds Ratio (OR) berechnet. Dieses liefert die Aussage, wie sich das Chancenverhältnis, ein wirksames Ergebnis zu erhalten bei Hinzunahme eines weiteren Labors verändert. Anschließend wird die Signifikanz des OR mit einem statistischen Test (α=5%), der aufgrund der hohen Simulationsdurchläufe nahezu immer zu einem signifikanten Ergebnis führt, geprüft.

Ergebnisse

ANOVA

Die der ANOVA zugrundeliegenden Daten basieren auf dem VAH-Ringversuch 2023-01 [9]. In Tabelle 1 werden die Ergebnisse zusammengefasst. Im VAH-Ringversuch 2023-01 wurde die bakterizide Wirksamkeit des Testproduktes Glutaraldehyd (GA) 50 Gew.-% wässrige Lösung mit dem quantitativen Suspensionsversuch gegen *Staphylococcus aureus* nach DIN EN 13727:2015 zur Bewertung der Laborleistung geprüft. Ziel des Versuchs war es, die Reduktion für das Testprodukt bei drei verschiedenen Testproduktkonzentrationen – 0,10% / 0,05% / 0,01% – 5 min – unter den vorgegebenen Testbedingungen zu ermitteln und unterschiedliche oder falsche Berechnungen zu identifizieren. Für drei der 31 Laboratorien ist aufgrund falscher Vorverdünnungen nur der 0,05%-Wert in die Auswertung einbezogen worden.

Aus der Analyse der Reduktionen für die verschiedenen Konzentrationen aus den nachfolgenden Abbildungen 1, 2 und 3 kann der Einfluss des Labors auf die Reduktionen in allen drei Wirksamkeitsbereichen (unwirksam, intermediär wirksam und wirksam) als signifikant bezeichnet werden (p-Werte alle < 0,001) [9].

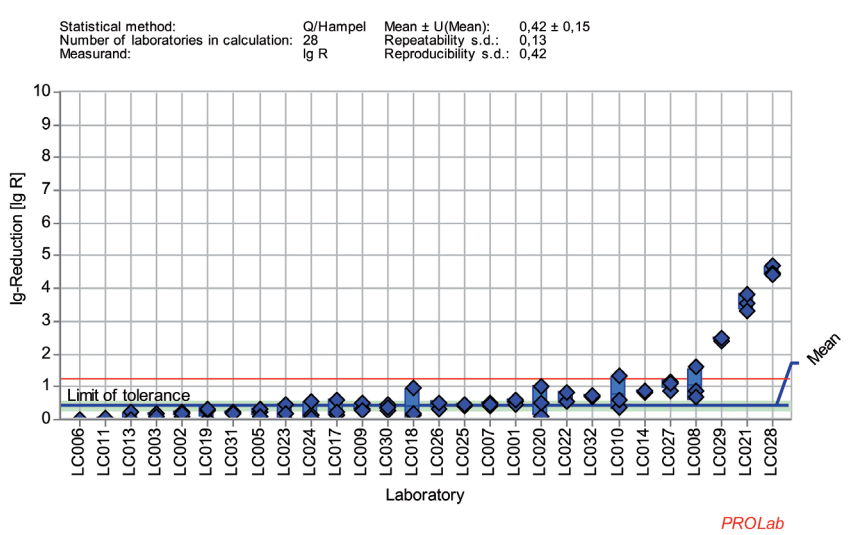

Abb. 1: Reduktion von *Staphylococcus aureus* nach EN 13727:2015 [Glutaraldehyd; 0,01% - 5 min] sortiert nach Labormittelwerten [9].

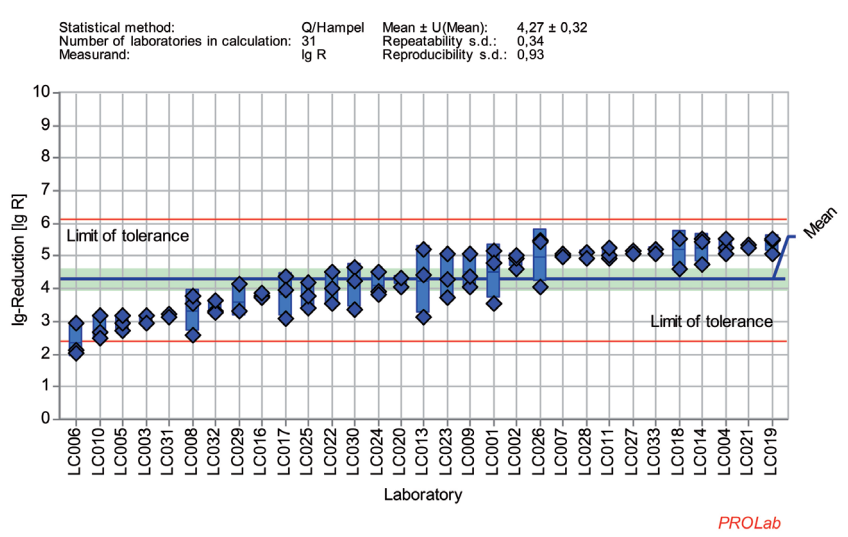

Abb. 2: Reduktion von *Staphylococcus aureus* nach EN 13727:2015 [Glutaraldehyd; 0,05% - 5 min] sortiert nach Labormittelwerten [9].

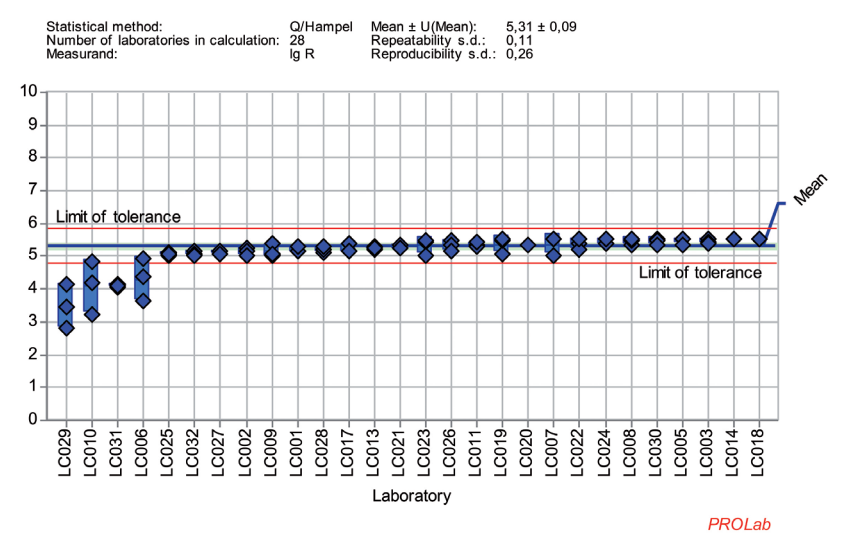

Abb. 3: Reduktion von *Staphylococcus aureus* nach EN 13727:2015 [Glutaraldehyd; 0,1% - 5 min] sortiert nach Labormittelwerten [9]

© mhp Verlag GmbH 2024

Simulation

Simulation „unwirksamer" Bereich – Glutaraldehyd 0,01% – 5 min

Der Anteil der Labore des VAH-Ringversuch 2023-01 mit einer Entscheidung „wirksam" (lg-R ≥ 5) liegt bei 0% und der Anteil der Labore mit einer Entscheidung „nicht wirksam" (lg-R < 5) liegt bei 100%. Auf eine Simulation wurde hier verzichtet, da zu 100% die Entscheidung „unwirksam" getroffen wird.

Simulation „intermediär wirksamer" Bereich – Glutaraldehyd 0,05% – 5 min

Es gab 10/31 (32,3%) Labore im VAH-Ringversuch 2023-01, die das Produkt als wirksam eingestuft haben und 21/31 (67,7%) Labore, die das Produkt als nicht wirksam deklarierten. Durch die Anzahl der Labore ergibt sich eine Anzahl an paarweisen Vergleichen von 465. Davon kommen 45,2% (210) zu einem ungleichen und 54,8% (255) zu einem gleichen Ergebnis.

Tabelle 2 zeigt die Vierfeldtertafel, die sich aus der Simulation für den „intermediär wirksamen" Bereich ergibt.

Zieht man nur die Entscheidung eines Labors heran, so wird das Produkt in 32,3% der Fälle als wirksam deklariert. Zieht man ein zweites Labor heran, dann beträgt diese Wahrscheinlichkeit nur noch 10,3%. Das daraus resultierende Odds-Ratio (OR) liegt bei 0,24. Somit wird die Chance, das Produkt als wirksam zu deklarieren, um 76% gesenkt, wenn man die Entscheidung eines zweiten Labors als Bewertung heranzieht.

Simulation „wirksamer" Bereich – Glutaraldehyd 0,1% – 5 min

Es gab 24/28 (85,7%) Labore im VAH-Ringversuch 2023-01, die das Produkt als wirksam eingestuft haben und 4/28 (14,3%) Labore, die das Produkt als nicht wirksam deklarierten. Durch die Anzahl der Labore ergibt sich eine Anzahl an paarweisen Vergleichen von 378. Davon kommen 96 (26,4%) zu einem ungleichen und 282 (74,6%) zu einem gleichen Ergebnis.

Aus der Simulation für den „wirksamen" Bereich ergibt sich die Vierfeldertafel in Tabelle 3.

Im wirksamen Bereich wird die Entscheidung „wirksam" bei einem Labor in 85,8% der Fälle getroffen. Durch die Hinzunahme eines zweiten Labors wird dieser Anteil auf 72,7% reduziert. Dies entspricht einem OR von 0,44, was einer Reduktion der Chance, ein Produkt als wirksam einzustufen, um 56% entspricht.

Diskussion

Die Präzision und Reproduzierbarkeit von Testverfahren bei der Beurteilung der Wirksamkeit chemischer Desinfektionsmittel sind von entscheidender Bedeutung, insbesondere vor dem Hintergrund der Biozidgesetzgebung, die niedrige Wirkstoffkonzentrationen zur Minimierung von Risiken für Anwender und Umwelt fordert. Die Biozidgesetzgebung der Europäischen Union (EU) stellt spezifische Anforderungen an die Sicherheit und Wirksamkeit von Desinfektionsmitteln, um sowohl den Anwenderschutz als auch den Umweltschutz zu gewährleisten (Europäische Kommission, Biozidverordnung (EU) Nr. 528/2012) [10]. Die Forderung nach minimalen Wirkstoffkonzentrationen resultiert aus dem Bestreben, die Exposition gegenüber potenziell schädlichen Chemikalien für den Anwender zu reduzieren. In diesem Kontext entwickelte Produkte befinden sich oft in einem Wirksamkeitsbereich, der nahe der kritischen Grenze zwischen eindeutiger Wirksamkeit und Unwirksamkeit liegt – in dieser Studie auch als intermediärer Bereich bezeichnet –, was eine präzise und zuverlässige Bewertung erschwert. Die Anforderungen des VAH stellen mit den geforderten zwei Prüfberichten aus unabhängigen, akkreditierten Prüflaboratorien und der Forderung nach Reproduktion der Ergebnisse der praxisnahen Prüfungen eine Möglichkeit dar, den Einfluss der inhärenten statistischen Variabilität mikrobiologischer Tests zwischen verschiedenen Laboratorien bei der Beurteilung der Wirksamkeit eines Desinfektionsmittels zu reduzieren.

Die Anforderungen der Deutschen Veterinärmedizinischen Gesellschaft (DVG) ähneln denen des VAH, indem sie eine Bestätigung der listungsrelevanten Ergebnisse eines vorgelegten Gutachtens durch eine begrenzte, repräsentative zweite Untersuchung (Eckwertprüfung) fordern. Der Umfang dieser Prüfung wird vom Ausschuss Desinfektion entsprechend den Ergebnissen des vorgelegten Gutachtens festgelegt [11]. Auch im Rahmen der OECD (Organisation für wirtschaftliche Zusammenarbeit und Entwicklung)-Methode zur Flächendesinfektion werden drei unabhängige Prüfdurchgänge in ein bis drei verschiedenen Prüflaboratorien gefordert [12]. Weitere Behörden, wie die United States Environmental Protection Agency (EPA) und die US Food and Drug Administration (FDA), fordern ebenfalls die Prüfung von drei Chargen der Prüfprodukte in drei unabhängigen Versuchen [13]. Diese Anforderungen spiegeln die allgemeine Tendenz wider, die Notwendigkeit einer soliden und reproduzierbaren Datenbasis für die Beurteilung der Wirksamkeit von Desinfektionsmitteln zu betonen, um die öffentliche Gesundheit und Sicherheit zu gewährleisten.

Die Europäische Chemikalienagentur (ECHA) scheint in dieser Hinsicht eine Ausnahme darzustellen, da sie für die Wirksamkeitsbeurteilung von Desinfektionsmitteln einen einzigen Prüf-

Tabelle 2: Vierfeldertafel der Simulation für den „intermediär wirksamen" Bereich

Entscheidung	beide Labore	ein Labor
wirksam	1026	3232
unwirksam	8974	6768

Tabelle 3: Vierfeldertafel der Simulation für den „wirksamen" Bereich

Entscheidung	beide Labore	ein Labor
wirksam	7277	8583
unwirksam	2723	1417

© mhp Verlag GmbH 2024

durchgang als ausreichend ansieht [8]. Die Komplexität der Testverfahren und die Notwendigkeit der Einhaltung strikter Standards, wie sie z.B. in den Normen vorgegeben sind, werden durch die Ergebnisse dieser Arbeit hervorgehoben. Aber selbst unter Einhaltung europäischer Normen und Richtlinien zeigen unsere Analysen, dass signifikante Unterschiede in den Ergebnissen verschiedener Labore auftreten können. Die Analyse der Ergebnisse aus dem VAH-Ringversuch 2023-01, speziell die ANOVA und Simulationen, liefern fundierte Erkenntnisse.

Im Rahmen der durchgeführten Varianzanalyse (ANOVA) wurde eine signifikante Variabilität/Heterogenität in den Reduktionsraten von *Staphylococcus aureus* mit Glutaraldehyd über die verschiedenen Labore hinweg festgestellt, was die Komplexität und die Herausforderung einer konsistenten Wirksamkeitsbeurteilung chemischer Desinfektionsmittel unterstreicht. Zur Quantifizierung der statistischen Signifikanz und der Zuverlässigkeit der ermittelten Odds Ratios wurden Konfidenzintervalle berechnet. Insbesondere im Grenzbereich zwischen Wirksamkeit und Unwirksamkeit bzw. im sog. intermediären Bereich hat die festgestellte Variabilität direkte und folgenreiche Auswirkungen auf die Zuverlässigkeit der Produktbewertungen. Die Simulationsergebnisse zeigen, dass im intermediären Wirkbereich die Einbindung eines zweiten unabhängigen Labors nicht nur eine breitere Datenbasis für die Entscheidungsfindung bietet, sondern auch die Wahrscheinlichkeit einer Fehlklassifikation eines unwirksamen Produktes als wirksam deutlich reduziert. In der durchgeführten Analyse betrug diese Reduktion des Fehlklassifikationsrisikos bis zu 76%. Dies zeigt, dass die Zweitmeinungsstrategie ein möglicher Weg ist, die Genauigkeit und Zuverlässigkeit der Wirksamkeitsbewertung auch bei eindeutig positiven Ergebnissen zu verbessern. Der positive Effekt der Zweitmeinung beschränkt sich somit nicht ausschließlich auf den intermediären Wirksamkeitsbereich.

In einer rechtlichen Bewertung der „Stellungnahme der Kommission für Krankenhaushygiene und Infektionsprävention (KRINKO) zu Anforderungen an Desinfektionsmittel für den Einsatz in infektionssensiblen Bereichen“ [14] von Krüger und Heidorn werden Fragen bezüglich des aktuellen Stands der Wissenschaft aufgeworfen [15]. Die Autoren argumentieren, dass der wissenschaftliche Beleg für die Forderung nach zwei unabhängigen Prüfberichten zur Bestätigung der Wirksamkeit von Oberflächendesinfektionsmitteln im Gesundheitswesen fehle. Die Untersuchungen dieser Arbeit liefern jedoch substantielle Belege, die diese Forderung der KRINKO stärken, indem sie die signifikante Variabilität der Testergebnisse zwischen verschiedenen Laboren aufzeigen und somit die Notwendigkeit für zwei unabhängige Bewertungen unterstreichen.

Diese Forderung wird ebenfalls durch die Untersuchung von Hilgren *et al.* (2022) [16] weiter gestärkt, deren Fokus auf der Reproduzierbarkeit und Präzision von mikrobiologischen Tests zur Bewertung der Wirksamkeit von Desinfektionsmitteln liegt. Sie beleuchtet spezifisch die Problematik der Reproduzierbarkeit und Variabilität der Testergebnisse bei der Anwendung eines ASTM-Standardverfahrens, das zur Übernahme als OECD-Verfahren vorgeschlagen wurde. Trotz der Anwendung standardisierter Testbedingungen wurde eine inakzeptable Reproduzierbarkeit der Ergebnisse zwischen den beteiligten Laboren festgestellt.

Die interlaboratorischen Studien von Parker et al. (2014) sowie von Allkja et al. (2021), an denen acht bzw. fünf Labore teilgenommen haben, dienten dazu, die Reproduzierbarkeit der MBEC (Minimum Biofilm Eradication Concentration)-Methode zu überprüfen und mit weiteren Methoden der Biofilm-Quantifizierung zu vergleichen. Diese Arbeiten zeigen nicht eindeutig, dass für eine robuste Bewertung immer zwei oder mehr Labore erforderlich sind [17, 18]. Im Gegensatz dazu zeigt die Studie von Goeres et al. (2019) zur Entwicklung und Validierung des Einzelröhrenverfahrens (Single Tube Method, STM) für die Wirksamkeitstests von antimikrobiellen Produkten gegen Biofilme zwar insgesamt eine gute Reproduzierbarkeit, dennoch wurden breite Schwankungen in den Ergebnissen zwischen verschiedenen Laboren beobachtet, insbesondere bei der Bewertung der logarithmischen Reduktionen der lebensfähigen Bakterien durch verschiedene antimikrobielle Mittel. Die mittlere Log-Reduktion der lebensfähigen Bakterien in der Studie, die Ergebnisse aus sechs Laboratorien umfasst, reichte von 2,32 bis 4,58, mit Reproduzierbarkeitsstandardabweichungen zwischen 0,89 und 1,67 [19].

Für die Standardisierung und Validierung eines neuen Tests oder einer neuen Methode für die Wirksamkeitsprüfung von Desinfektionsverfahren ist die Einbeziehung mehrerer Labore wichtig, um die Robustheit und allgemeine Anwendbarkeit der Methode zu demonstrieren. Die Einbindung mehrerer Labore ist besonders relevant, wenn die Ergebnisse für regulatorische Zwecke verwendet werden sollen oder wenn es um die Bewertung der Wirksamkeit von Produkten geht, die auf den Markt gebracht werden sollen.

Die Analyse weiterer Ringversuche untermauern diese Erkenntnisse. Es werden derzeit Daten aus Ringversuchen mit praxisnahen Tests (Phase 2/ Stufe 2) für eine weitere Publikation statistisch aufbereitet. Der Einfluss des Labors auf die Ergebnisse der Wirksamkeitstestung in der jetzt vorgelegten Auswertung des quantitativen Suspensionsversuchs lenkt den Blick insbesondere auf Produkte im intermediären Wirksamkeitsbereich und die Auswirkungen auf die Genauigkeit und Verlässlichkeit der Wirksamkeitsbeurteilung, selbst wenn europäische Normen und Richtlinien für die Testung eingehalten werden.

Die Balance zwischen Minimierungsgebot und zuverlässiger Wirksamkeit zur Sicherstellung der Infektionsprävention wird im Kontext der Biozidgesetzgebung zu einer zentralen Fragestellung, die eine sorgfältige Abwägung und Bewertung erfordert. Die fortlaufende wissenschaftliche und regulatorische Auseinandersetzung mit diesen Themen ist notwendig, um den Schutz der Anwender und der öffentlichen Gesundheit sicherzustellen. Ein weiterer kritischer Aspekt ist die mögliche Entwicklung von Toleranzen und ggfs. Resistenzen durch den Einsatz von Desinfektionsmitteln im subletalen Bereich. Die fortgesetzte Exposition von Mikroorganismen gegenüber unzureichenden Desinfektionsmittelkonzentrationen kann zur Selektion angepasster Stämme führen [20]. Die Entwicklung von Desinfektionsmitteln, die nicht nur aktuellen, sondern auch zukünftigen Herausforderungen in der Infektionsprävention gerecht werden, erfordert eine vorausschauende Forschung und Entwicklung. Dies un-

© mhp Verlag GmbH 2024

terstreicht die Bedeutung der Zusammenarbeit zwischen der Wissenschaft, Industrie und Regulierungsbehörden, um sicherzustellen, dass die Teststandards und Richtlinien mit den neuesten Erkenntnissen und technologischen Entwicklungen Schritt halten.

Schlussfolgerungen

Die statistische Analyse in der vorliegenden Arbeit zeigt, dass insbesondere für den intermediären Bereich – also der Bereich, der sich unterhalb des wirksamen Bereiches befindet – davon auszugehen ist, dass auch bei erfahrenen und akkreditierten Laboratorien und unter verhältnismäßig gut zu kontrollierenden Bedingungen eine hohe Varianz der Ergebnisse zwischen den Laboratorien gegeben ist. Da für die Zulassung von Bioziden grundsätzlich aus öko- und humantoxikologischer Sicht ein niedriger Wirkstoffgehalt günstiger beurteilt wird und zu bevorzugen ist, besteht die Gefahr, dass bei Berücksichtigung nur eines Prüfberichts die interlaboratorielle Schwankungsbreite der Ergebnisse zwischen Laboratorien nicht erfasst wird.

Aus Sicht der Infektionsprävention sowie unter öko- und humantoxikologischen Gesichtspunkten sollte es stets das Ziel sein, ein Desinfektionsverfahren nur dann anzuwenden, wenn seine Wirksamkeit sicher gewährleistet ist.

Die Einbeziehung der Prüfberichte aus einem zweiten Labor erhöht signifikant die Aussagekraft der Bewertung über die sichere Wirksamkeit eines Produkts. Der Einsatz eines Produkts oder Verfahrens mit unzureichender bzw. unzureichend abgesicherter Wirksamkeit ist auch im Hinblick auf die möglichen Nebenwirkungen nicht vertretbar.

In der Konsequenz erfordert die Zulassung und fortlaufende Überprüfung von Desinfektionsmitteln eine multidisziplinäre Herangehensweise, die sowohl mikrobiologische Wirksamkeitstests als auch ökotoxikologische und resistenzepidemiologische Bewertungen umfasst. Die Implementierung von Standards, die die Einbeziehung von Daten aus mehreren unabhängigen Quellen vorschreiben, trägt wesentlich zur Sicherstellung der Zuverlässigkeit und Effektivität von Desinfektionsverfahren bei und unterstützt die öffentliche Gesundheit sowie den Umweltschutz.

Literatur

1. DIN, DIN EN 13727 – Quantitativer Suspensionsversuch zur Bestimmung der bakteriziden Wirkung im humanmedizinischen Bereich – Prüfverfahren und Anforderungen (Phase 2, Stufe 1); Deutsche Fassung EN 13727:2012+A2:2015, in EN 13727:2012+A2:2015, TC216, Editor. 2015, European Committee for Standardization (CEN). p. 1–56.
2. DIN, DIN EN 13624 – Quantitativer Suspensionsversuch zur Bestimmung der fungiziden oder levuroziden Wirkung im humanmedizinischen Bereich – Prüfverfahren und Anforderungen (Phase 2, Stufe 1); Deutsche Fassung EN 13624:2021, in EN 13624, TC216, Editor. 2022, European Committee for Standardization (CEN). p. 1–63.
3. DIN, DIN EN 14348 – Quantitativer Suspensionsversuch zur Bestimmung der mykobakteriziden Wirkung chemischer Desinfektionsmittel im humanmedizinischen Bereich einschließlich der Instrumentendesinfektion. Prüfverfahren und Anforderungen (Phase 2/Stufe 1), in EN 14348, TC216, Editor. 2005, European Committee for Standardization (CEN). p. 1–38.
4. DIN, DIN EN 14885 – Anwendung Europäischer Normen für chemische Desinfektionsmittel und Antiseptika; Deutsche EN 14885:2022 + AC:2023, in EN 14885, TC216, Editor. 2023, European Committee for Standardization (CEN).
5. DIN, DIN EN 14561 – Quantitativer Keimträgerversuch zur Prüfung der bakteriziden Wirkung für Instrumente im humanmedizinischen Bereich – Prüfverfahren und Anforderungen (Phase 2, Stufe 2), in EN 14561, TC216, Editor. 2006, European Committee for Standardization (CEN). p. 1–39.
6. DIN, DIN EN 14562 – Quantitativer Keimträgerversuch zur Prüfung der fungiziden oder levuroziden Wirkung für Instrumente im humanmedizinischen Bereich – Prüfverfahren und Anforderungen (Phase 2, Stufe 2), in EN 14562, TC216, Editor. 2006, European Committee for Standardization (CEN). p. 1–41.
7. DIN, DIN EN 14563 – Quantitativer Keimträgerversuch zur Prüfung der mykobakteriziden oder tuberkuloziden Wirkung chemischer Desinfektionsmittel für Instrumente im humanmedizinischen Bereich – Prüfverfahren und Anforderungen (Phase 2, Stufe 2), in EN 14563, TC216, Editor. 2009, European Committee for Standardization (CEN). p. 1–36.
8. ECHA. Guidance on the Biocidal Products Regulation. Volume II: Efficacy. Parts B+C: Assessment and Evaluation. Version 6.0, August 2023. 2023, Europaen Chemicals Agency. p. 526.
9. VAH. Final Report VAH Ring Trial 2023-1. Chemical disinfectants and antiseptics. Quantitative suspension test for the evaluation of bactericidal activity (Phase 2, Step 1) – DIN EN 13727:2015, *S. aureus*. 2023, VAH: Bonn. p. 29. https://vah-online.de/files/download/Ringversuche/Ring%20trial_VAH%202023-1.pdf.
10. Europäisches Parlament und Rat der Europischen Union. Biozidverordnung (EU) Nr. 528/2012 des europäischen Parlaments und des Rates vom 22. Mai 2012 über die Bereitstellung auf dem Markt und die Verwendung von Biozidprodukten, in Amtsblatt der Europäischen Union. 2012.
11. Deutsche Veterinärmedizinische Gesellschaft (DVG). Voraussetzungen für die Desinfektionsmittelprüfung und Aufnahme in die Desinfektionsmittellisten der DVG – Stand: 12.02.2023. 2023. p. 12.
12. OECD. Guidance document on quantitative methods for evaluating the activity of microbicides used on hard non-porous surfaces, in Series on Biocides No. 6 – Series on testing and assessment No. 187. 2013.
13. U.S. Environmental Protection Agency, EPA. Antimicrobial Testing Methods & Procedures Developed by EPA's Microbiology Laboratory. 2024.
14. KRINKO. Stellungnahme der Kommission für Krankenhaushygiene und Infektionsprävention (KRINKO) zu Anforderungen an Desinfektionsmittel für den Einsatz in infektionssensiblen Bereichen. Epid Bull, 2023. 23: 22–26.
15. Krüger H, Heidorn H. Die KRINKO-Empfehlung zur Flächendesinfektion - Rechtliche Bewertung aus der Perspektive des Biozidrechts. StoffR, 2023. 3: p. 186–201.
16. Hilgren J et al. Reproducibilty problem with a proposed standard method to measure disinfection efficacy. bioRxiv, 2022. preprint.
17. Parker AE et al. Ruggedness and reproducibility of the MBEC biofilm disinfectant efficacy test. J Microbiol Methods, 2014. 102: p. 55–64.
18. Allkja J et al. Interlaboratory study for the evaluation of three microtiter plate-based biofilm quantification methods. Sci Rep, 2021. 11(1): p. 13779.
19. Goeres DM et al. Development, standardization, and validation of a biofilm efficacy test: The single tube method. J Microbiol Methods, 2019. 165: p. 105694.
20. Kampf G. Adaptive bacterial response to low level chlorhexidine exposure and its implications for hand hygiene. Microb Cell, 2019. 6(7): p. 307–320.

© mhp Verlag GmbH 2024