

How many laboratories are necessary? Analytical assessment of precision and reproducibility in disinfectant efficacy testing for quantifying the impact of laboratory number on result reliability

Wie viele Labore sind notwendig? Analytische Bewertung der Präzision und Reproduzierbarkeit bei der Prüfung der Wirksamkeit von Desinfektionsmitteln zur Quantifizierung des Einflusses der Laboranzahl auf die Zuverlässigkeit der Ergebnisse

Abstract

Background: Reliable assessment of the efficacy of chemical disinfection procedures requires robust experimental precision and reproducibility. In order to obtain a certificate from the Association for Applied Hygiene (VAH) proving the efficacy of a disinfectant procedure, test results from two independent accredited laboratories must be submitted for evaluation. In a previous study, Monte-Carlo simulations based on VAH ring trial data demonstrated that this two-laboratory stipulation significantly reduces the probability of false-positive efficacy statements. The present study extends that work by developing an analytical solution to quantify the influence of intra- and interlaboratory variability on efficacy assessment outcomes.

Methods: The analysis is based on data from four VAH ring trials covering bactericidal, yeasticidal, and mycobactericidal activities as tested according to EN 13727, EN 13624, EN 17387 and EN 14563. Analytical probability models were derived to calculate the likelihood that one, two, three, or four laboratories independently classify a disinfectant as effective under defined test conditions. The results were compared across different efficacy ranges and interpreted in relation to the existing two-laboratory validation framework.

Results: The analytical solution confirms that including a second laboratory significantly enhances the robustness of efficacy classification by reducing the probability of false-positive outcomes in the intermediate efficacy range by up to 90%. Beyond two laboratories, the benefit is observed to drop sharply: Inclusion of a third or fourth laboratory further lowers the probability of consistent positive classifications, but the marginal gain in reliability is small relative to the associated increase in cost, effort, and organizational complexity. The model thus reveals an asymptotic relationship between the number of laboratories and classification certainty.

Conclusion: The analytical findings corroborate the empirical findings from VAH ring trials, demonstrating that the two-laboratory stipulation represents an optimal balance between statistical reliability and practical feasibility. Expanding beyond two laboratories provides minimal additional certainty and may increase disagreement rates due to compounding intralaboratory variation. Future standardization efforts should therefore prioritize the reduction of intralaboratory variability rather than expanding interlaboratory replication.

Kira-Marie Roesch^{1,2}

Marvin Rausch^{1,2}

Felix Droop¹

Martin Exner^{2,3}

Stefanie Gemein⁴

Carola Ilschner^{1,2}

Axel Kramer^{2,5}

Thomas Selhorst⁶

Miranda Suchomel^{2,7}

Nico T. Mutters^{1,2}

Jürgen Gebel^{1,2}

1 University Bonn, University hospital Bonn, Institute for Hygiene and Public Health, Bonn, Germany

2 VAH – Association for Applied Hygiene, Germany

3 University of Bonn, Germany

4 Referenzinstitut für Bioanalytik, Bonn, Germany

5 Institute of Hygiene and Environmental Medicine, University Medicine Greifswald, Germany

6 Federal Institute for Risk Assessment, Berlin, Germany

7 Institute for Hygiene and Applied Immunology, Medical University of Vienna, Vienna, Austria

Keywords: disinfectant efficacy testing, precision and reproducibility, analytical modelling, ring trials, interlaboratory variability, intralaboratory variation, confirmation probability, quality assurance

Zusammenfassung

Hintergrund: Eine zuverlässige Bewertung der Wirksamkeit chemischer Desinfektionsverfahren erfordert eine hohe experimentelle Präzision und Reproduzierbarkeit. Der Verbund für Angewandte Hygiene (VAH) fordert als Voraussetzung für eine Produktzertifizierung Testergebnisse von zwei unabhängigen akkreditierten Laboren. In einer früheren Studie haben Monte-Carlo Simulationen auf der Grundlage von VAH-Ringversuchsdaten gezeigt, dass diese Forderung zur Prüfung durch zwei Labore die Wahrscheinlichkeit falsch-positiver Wirksamkeitsaussagen erheblich verringert. Die vorliegende Studie erweitert diese Arbeit durch die Entwicklung einer analytischen Lösung zur Quantifizierung des Einflusses der Variabilität innerhalb und zwischen Laboren auf die Ergebnisse der Wirksamkeitsbewertung.

Methoden: Die Analyse basiert auf Daten aus vier VAH-Ringversuchen zur Bestimmung der bakteriziden, levuroziden und mykobakteriziden Wirkung gemäß den europäischen Testmethoden EN 13727, EN 13624, EN 17387 und EN 14563. Es wurden analytische Wahrscheinlichkeitsmodelle abgeleitet, um die Wahrscheinlichkeit zu berechnen, dass ein, zwei, drei oder vier Laboratorien ein Desinfektionsmittel unter definierten Testbedingungen unabhängig voneinander als wirksam einstufen. Die Ergebnisse wurden über verschiedene Wirksamkeitsbereiche hinweg verglichen und im Zusammenhang mit dem bestehenden Validierungsrahmen für zwei Laboratorien interpretiert.

Ergebnisse: Die analytische Lösung bestätigt, dass die Einbeziehung eines zweiten Labors die Robustheit der Wirksamkeitsklassifizierung deutlich verbessert, indem sie die Wahrscheinlichkeit falsch-positiver Ergebnisse im intermediären Wirksamkeitsbereich um bis zu 90% reduziert. Bei mehr als zwei Laboren nimmt der Nutzen jedoch stark ab: Die Einbeziehung eines dritten oder vierten Labors senkt die Wahrscheinlichkeit konsistenter positiver Klassifizierungen weiter, aber der marginale Gewinn an Zuverlässigkeit ist im Verhältnis zu den damit verbundenen höheren Kosten, höherem Aufwand und größerer organisatorischer Komplexität gering. Das Modell zeigt somit eine asymptotische Beziehung zwischen der Anzahl der Labore und der Klassifizierungssicherheit.

Schlussfolgerung: Die analytischen Ergebnisse untermauern die Erkenntnisse aus VAH-Ringversuchen und zeigen, dass die Zwei-Labor Entscheidung ein optimal ausgewogenes Verhältnis von statistischer Zuverlässigkeit und notwendigem Aufwand darstellt. Die Erweiterung über zwei Labore hinaus bietet nur minimale zusätzliche Sicherheit und kann aufgrund der sich verstärkenden intralaboratorischen Variation zu einer Erhöhung der Diskrepanzen führen. Zukünftige Standardisierungsbemühungen sollten daher der Verringerung der Variabilität innerhalb der Labore Vorrang vor der Ausweitung der Replikation zwischen den Labors einräumen.

Schlüsselwörter: Wirksamkeitsprüfung von Desinfektionsmitteln, Präzision, Reproduzierbarkeit, analytische Modellierung, Ringversuch, interlaboratorielle Variabilität, Qualitätssicherung

Introduction

Reliable assessment of the efficacy of chemical disinfection procedures requires a high level of precision and reproducibility, as even minor methodological or biological variations can substantially influence test outcomes and thus the efficacy of the disinfection measure. The current approach adopted by the Association for Applied Hygiene (VAH), which mandates test results from two independent accredited laboratories [1], has proven to be an important quality assurance measure. The previous study, “Requirements for the precision and reproducibility in the efficacy testing of chemical disinfection procedures” [2], demonstrated through Monte-Carlo simulations based on VAH ring trial data that the inclusion of a second laboratory substantially reduces the probability of misclassification and results in more conservative and reproducible efficacy evaluations.

While the prior publication focused on simulation-based estimations of uncertainty, the present study emphasizes an analytical statistical approach to further elucidate the mechanisms underlying inter- and intralaboratory variability. By deriving closed-form probability expressions, the analysis describes how variance structures within and between laboratories determines the likelihood that one, two, three, or more laboratories independently classify a procedure as effective. This approach allows the confirmation probability of efficacy assessments to be expressed explicitly as a function of laboratory number and underlying variability.

A central focus of the present study is the behaviour of efficacy classifications within decision-critical result ranges, that is, concentration–time conditions close to the regulatory efficacy threshold. In these ranges, laboratory-related variability exerts a disproportionate influence on classification outcomes, such that additional laboratories may potentially have a significant impact on the probability of achieving concordant efficacy decisions. The analytical model therefore permits a systematic evaluation of how expanding the test design from one to multiple laboratories affects classification certainty and agreement.

This issue is not only of theoretical but also of practical as well as of regulatory importance. On the one hand, including more laboratories may enhance the validity of efficacy assessments within decision-critical result ranges. On the other hand, should an increasing number of laboratories be consecutively involved in the decision-making process, it can be assumed that the disinfection procedure in question will ultimately be rejected, on the condition that a minimum of one of the total numbers of laboratories involved classifies the disinfectant procedure as ineffective. An analytical examination of these relationships allows for a deeper understanding of the balance between methodological robustness, statistical significance, and economic feasibility.

The present work therefore builds directly upon the VAH dataset and prior findings from ring trials to provide an analytically derived model for precision and reproducibility in disinfectant efficacy testing. The overall goal is to determine whether increasing the number of test laboratories yields a measurable improvement in the reliability of efficacy classification or whether, beyond a certain threshold, the incremental benefit no longer justifies the additional complexity, cost, and resource requirements. These results are expected to inform future standardization efforts and contribute to the evidence-based refinement of regulatory frameworks within the European and international context.

Method

Data basis and study context

The present analysis builds upon the same dataset used in our previous study [2], which evaluated interlaboratory variability in disinfectant efficacy testing using data from four VAH ring trials [3]. These ring trials, organized by the VAH, were conducted over several years to assess the precision and reproducibility of standardized quantitative test methods for chemical disinfectants and antiseptics in accordance with European standards. The goal of the VAH ring trial program is to ensure method reliability, evaluate laboratory performance under harmonized test conditions, and provide empirical data to support quality assurance and accreditation within microbiological efficacy testing.

VAH ring trials and test design

The analytical evaluation presented here uses data from four VAH ring trials, each representing distinct microorganism categories and methodological conditions (Table 1) [3], [4], [5], [6], [7].

All participating laboratories received identical test products from a single production batch, provided and distributed by the VAH. Each laboratory performed quantitative efficacy tests according to the selected European standard under controlled and standardized conditions. Reduction factors (R; Ig-reductions) were determined for each test product and condition. The datasets thus capture both intra- and interlaboratory sources of variation under simulated-use testing conditions representative of the VAH certification framework.

Analytical solution

Variability arises both intralaboratory and interlaboratory, raising the question of whether efficacy assessments based on a single laboratory are sufficiently reliable. To increase the robustness of the evaluation process, it is advisable to involve multiple independent laboratories.

Table 1: Test parameter for each VAH ring trial with number of participating laboratories.

VAH ring trial	VAH R2023-01*	VAH R2020-01*	VAH R2022-01*	VAH R2019-02*
Method	EN 13723	EN 13624	EN 17387	EN 14563
Test organism	<i>Staphylococcus aureus</i> (ATCC 6538)	<i>Candida albicans</i> (ATCC 10231)	<i>Enterococcus hirae</i> (ATCC 10541)	<i>Mycobacterium terrae</i> (ATCC 15755)
Phase	2/1	2/1	2/2	2/2
Test type	Suspension test	Suspension test	Surface test without mechanical action	Carrier test for instruments
Number of participating laboratories	31	30	27	16

*Code of ring trial

Concordant results across laboratories reduce the risk of misclassification.

In this context, the study compares two decision rules:

- one-laboratory rule: This regulatory approach designates a single laboratory—selected at random—as the sole authority for the efficacy assessment
- two-laboratory rule: Under this rule, a substance is classified as effective only if both independent laboratories unanimously report it as effective
- X-laboratory rule: Under this rule, a substance is classified as effective only if X independent laboratories unanimously report it as effective

The proportion of laboratories of a ring trial is N . The laboratories with stated the product effective ($R \geq 5$ lg for bactericidal efficacy and $R \geq 4$ lg for yeasticidal efficacy) and not effective ($R < 5$ lg accordingly $R < 4$ lg) was then determined.

Under the one-laboratory rule, an effective assessment corresponds to the event that the first laboratory drawn belongs to the laboratories which classified the disinfection procedure as effective (N_{eff}). The corresponding confirmation probability p_1 is

$$p_1 = \frac{N_{eff}}{N}$$

Under the two-laboratory rule, an effective assessment corresponds to the event that, after one laboratory has been drawn and classified the procedure as effective, a second laboratory drawn without replacement from the remaining pool also classifies the procedure as effective. The corresponding confirmation probability therefore reflects this sequential sampling process. Owing this procedure, the corresponding confirmation probability is

$$p_2 = \frac{N_{eff}}{N} \cdot \frac{N_{eff} - 1}{N - 1}$$

Where

$$CP = \frac{(N_{eff} - 1)}{(N - 1)}$$

defines the confirmation probability (CP) of the second laboratory.

Accordingly, $1 - CP$ quantifies the probability that an effective classification under the one-laboratory rule is not confirmed when a second laboratory is required.

For a specific number of laboratories $k > 2$ the confirmation probability of an effective final classification p_k is:

$$p_k = \frac{N_{eff}}{N} \prod_{j=1}^{k-1} \frac{(N_{eff} - j)}{(N - j)}$$

Results

Intralaboratory instead of interlaboratory variability

In a previous paper, we showed that statistical analyses using one-way analysis of variance (ANOVA) demonstrated that the laboratory itself exerts a statistically significant influence on the measured reduction values across all four ring trials ($p < 0.001$ for all test types and efficacy ranges) [2]. This finding confirmed substantial interlaboratory variation, reflecting differences in execution, equipment, environmental conditions, and operator-specific factors despite adherence to standardized protocols. On this basis, the present study no longer focuses on variation between laboratories (interlaboratory) but rather on variability within a laboratory (intralaboratory), the variation that occurs in repeated measurements within a single laboratory. Using an analytical statistical framework, this work aims to derive a closed-form solution that describes how such intralaboratory variability affects the precision and reliability of efficacy classification when expanding from one to multiple laboratory datasets.

The analytical model used in this study therefore provides a mathematical complement to the simulation-based findings of the prior publication, quantifying the trade-off between increasing the number of laboratories involved and the resulting statistical certainty of the efficacy decision.

Analytical solution of the ineffective range

For the ineffective range of each ring trial, the proportion of laboratories in VAH ring trials with a decision of ineffective is 100%. No simulation was carried out here, as the decision ineffective was unanimous (Table 2).

Table 2: Number of laboratories in the different VAH ring trials and ranges which classified the tested product as effective.

Range	VAH R2023-01*	VAH R2020-01*	VAH R2022-01*	VAH R2019-02*
Ineffective	0.01%–5 min	0.05%–15 min	30.0%–5 min	1.0%–15 min
	0/28 (0%)	0/30 (0%)	0/27 (0%)	0/16 (0%)
Intermediate effective	0.05%–5 min	0.2%–15 min	40.0%–5 min	1.0%–30 min
	10/31 (32.3%)	8/30 (26.7%)	3/27 (11.1%)	3/16 (18.8%)
Effective	0.1%–5 min	0.3%–15 min	60.0%–5 min	1.0%–60 min
	24/28 (85.7%)	20/30 (66.7%)	24/27 (88.9%)	12/16 (75.0%)

*Code of ring trial

Analytical solution of the intermediate effective range

For the intermediate effective range of each ring trial, the proportion of laboratories in VAH ring trials with a decision of effective is shown in Table 2. The results of the analytical solution for one, two, three and four laboratories can be found in Table 3 and Figure 1 calculated according to Equation 1–4.

Probability for classification as effective in the different ring trials

- R2023-01: The analytical solution shows that the probability that a single laboratory classifies a procedure as effective is 32.3%. The probability that two independently selected laboratories both classify the same procedure as effective is 9.7%, and the probability that three laboratories do so is 2.7% and with four laboratories 0.7% (see Table 3 and Figure 1). The probability that a second, independently selected laboratory also classifies a procedure as effective, given that the first laboratory has already classified it as effective is 30.0%.
- R2020-01: The analytical solution in comparison describes the probability of effective results for one laboratory of 26.7%, reducing to 6.4% with a second laboratory and further reducing to 1.3% and 0.2% with inclusion of a third or fourth laboratory (see Table 3 and Figure 1). The probability that a second, independently selected laboratory also classifies a procedure as effective, given that the first laboratory has already classified it as effective is 24.1%.
- R2022-01: The analytical solution in comparison describes the probability of effective results for one laboratory of 11.1%, reducing to 0.9% with a second laboratory and further reducing to 0.0% with inclusion of a third or fourth laboratory (see Table 3 and Figure 1). The probability that a second, independently selected laboratory also classifies a procedure as effective, given that the first laboratory has already classified it as effective is 7.8%.
- R2019-02: The analytical solution in comparison describes the probability of effective results for one laboratory of 18.8%, reducing to 2.5% with a second laboratory and further reducing to 0.2% or 0.0% with

inclusion of a third or fourth laboratory (see Table 3 and Figure 1). The probability that a second, independently selected laboratory also classifies a procedure as effective, given that the first laboratory has already classified it as effective is 13.3%.

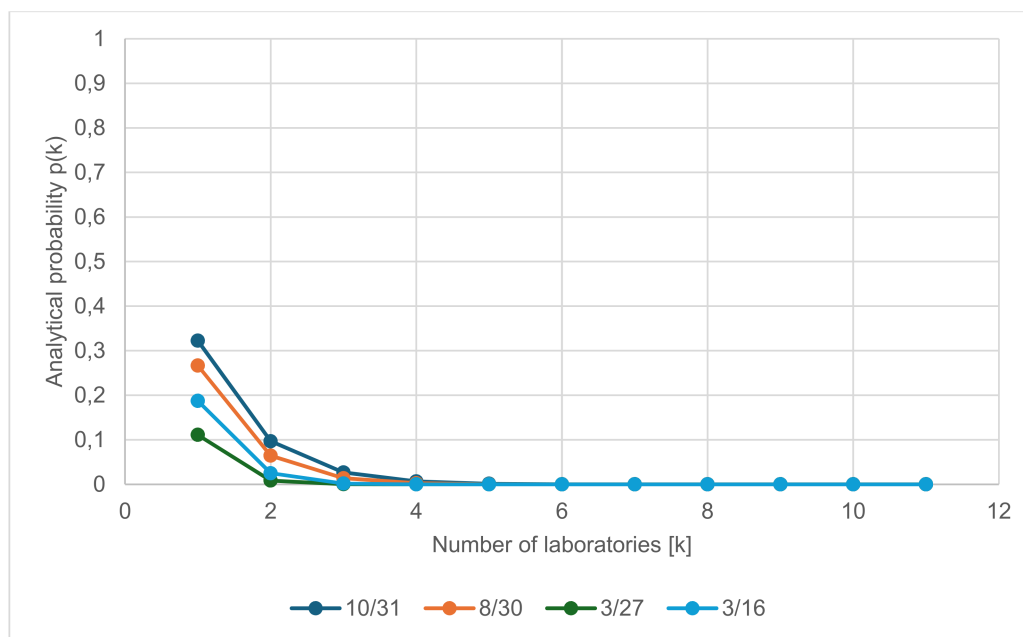
Analytical solution of the effective range

For the effective range of each ring trial, the proportion of laboratories in VAH ring trials with a decision of ineffective is shown in Table 2. The results of the analytical solution for one, two, three or four laboratories can be found in Table 4 and Figure 2 calculated according to Formula 1–4.

- R2023-01: The analytical solution in comparison describes the probability of effective results for one laboratory of 85.7%, reducing to 73.0% with a second laboratory and further reducing to 61.9% or 51.9% with inclusion of a third or fourth laboratory (see Table 4 and Figure 2). The probability that a second, independently selected laboratory also classifies a procedure as effective, given that the first laboratory has already classified it as effective is 85.2%.
- R2020-01: The analytical solution in comparison describes the probability of effective results for one laboratory of 66.7%, reducing to 43.7% with a second laboratory and further reducing to 28.1% or 17.7% with inclusion of a third or fourth laboratory (see Table 4 and Figure 2). The probability that a second, independently selected laboratory also classifies a procedure as effective, given that the first laboratory has already classified it as effective is 65.5%.
- R2022-01: The analytical solution in comparison describes the probability of effective results for one laboratory of 88.9%, reducing to 78.6% with a second laboratory and further reducing to 69.2% or 60.5% with inclusion of a third or fourth laboratory (see Table 4 and Figure 2). The probability that a second, independently selected laboratory also classifies a procedure as effective, given that the first laboratory has already classified it as effective is 88.5%.
- R2019-02: The analytical solution in comparison describes the probability of effective results for one laboratory of 75.0%, reducing to 55.0% with a second laboratory and further reducing to 39.3% or 27.2% with inclusion of a third or fourth laboratory (see Table 4 and Figure 2). The probability that a second, inde-

Table 3: Results of the analytical solution for the intermediate effective range of each VAH ring trial [p=probability, CP=confirmation probability of 1 to 2].

Code of ring trial	p1	p2	p3	p4	CP
R2023-01	0.323	0.097	0.027	0.007	0.300
R2020-01	0.267	0.064	0.013	0.002	0.241
R2022-01	0.111	0.009	0.000	0.000	0.078
R2019-02	0.188	0.025	0.002	0.000	0.133

**Figure 1: Analytical probability of laboratories that show effective results within the intermediate effective range of each VAH ring trial in the scheme Neff/N [dark blue=R2023-01, orange=R2020-01, green=R2022-01, light blue=R2019-02].****Table 4: Results of the analytical solution for the effective range of each VAH ring trial. [p=probability, CP=confirmation probability from 1 to 2].**

Code of ring trial	p1	p2	p3	p4	CP
R2023-01	0.857	0.730	0.619	0.519	0.852
R2020-01	0.667	0.437	0.281	0.177	0.655
R2022-01	0.889	0.786	0.692	0.605	0.885
R2019-02	0.750	0.550	0.393	0.272	0.733

pendently selected laboratory also classifies a procedure as effective, given that the first laboratory has already classified it as effective is 73.3%.

Analytical solution of a theoretical data set

To see how the probability of equal outcomes changes depending on what the outcome of a dataset looks like, Figure 3 shows this graphically for illustration purposes.

Discussion

The analytical evaluation presented here builds upon the findings of the VAH ring trials and the simulation-based results of the previous publication [2]. While the prior

analysis demonstrated by means of Monte-Carlo simulation that the inclusion of a second independent laboratory significantly reduces the probability of overly optimistic efficacy classifications, the present study provides a closed-form analytical solution that allows the same relationships to be expressed and interpreted mathematically. The VAH ring trial program, which systematically investigates precision and reproducibility across multiple laboratories, has shown that variability in disinfection efficacy testing is not solely random but influenced by identifiable factors intrinsic to each laboratory [8], [9]. In the earlier analysis, one-way ANOVA revealed that the laboratory itself significantly affects the outcome ($p < 0.001$ in all ring trials), confirming the presence of substantial interlaboratory variation even under standardized testing conditions. Building on this foundation, the current analytical approach identifies intralaboratory

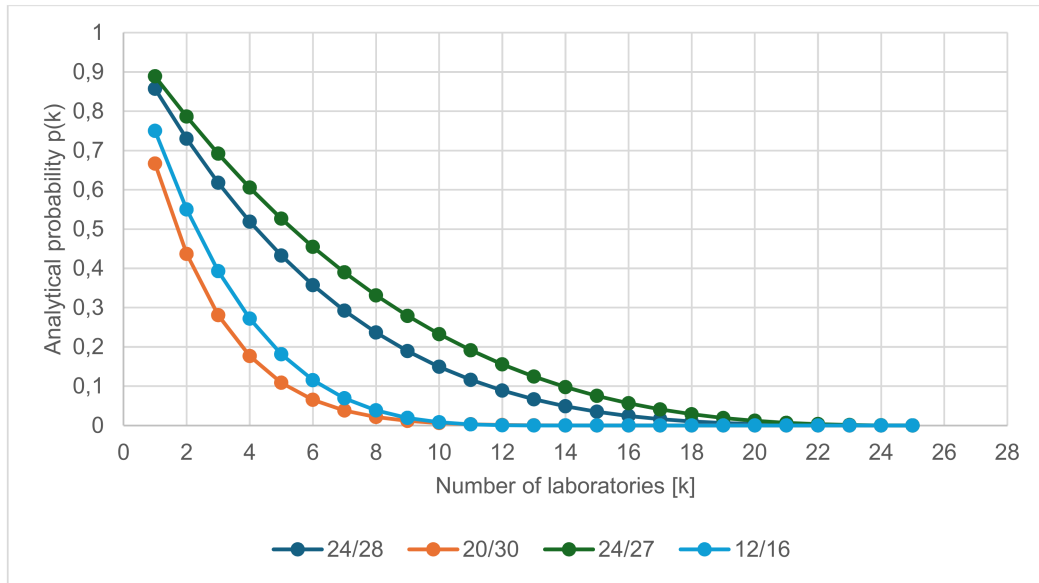


Figure 2: Analytical probability of laboratories that show effective results within the effective range of each VAH ring trial in the scheme N_{eff}/N [dark blue=R2023-01, orange=R2020-01, green=R2022-01, light blue=R2019-02]

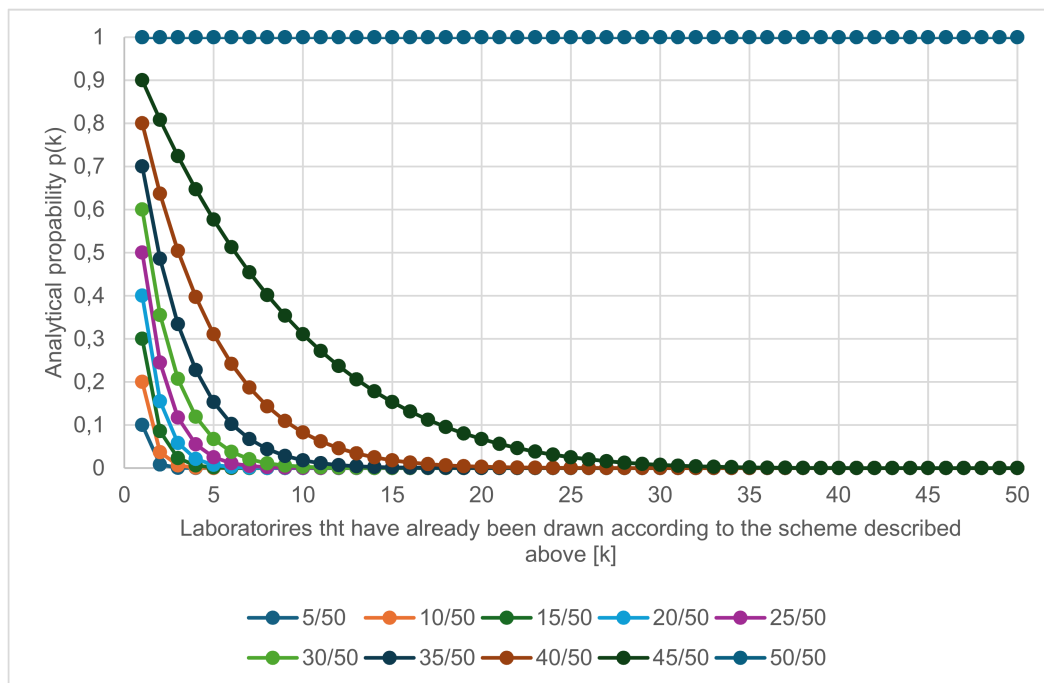


Figure 3: Analytical probability of laboratories that show effective results within a theoretical data set in steps of 5 effective laboratories till 50 laboratories in the scheme N_{eff}/N .

variability as a key determinant of the overall uncertainty in efficacy classification and quantifies how this uncertainty propagates when the results from additional laboratories are considered in the decision process regarding efficacy.

Across all four ring trials, the analytical results confirm that including a second independent laboratory significantly increases the robustness of efficacy classification by reducing the probability of a product being incorrectly assessed as effective in the intermediate range (Table 3). For instance, in trial R2023-01, the probability that a single laboratory classified a product as effective was

32.3%, whereas the probability that two laboratories independently agreed on such a classification dropped to 9.7%. These results corroborates the interpretation that multi-laboratory evaluation serves as an effective corrective mechanism for random or systematic laboratory-level bias.

The analytical results for the effective range reveal a similar but less pronounced trend. The likelihood that two laboratories independently confirm efficacy decreases moderately compared with single-laboratory assessment, reflecting the higher inherent stability of test results in clearly effective conditions (Table 4). For example, in

R2023-01 the probability of an effective classification decreased from 85.7% for one laboratory to 73.0% for two laboratories. The CP that a second laboratory confirms the first laboratory's positive result was 85.2%, indicating that even under optimal conditions, full agreement is not guaranteed. Similarly, CP values of 88.5% (R2022-01) and 73.3% (R2019-02) demonstrate that residual experimental uncertainty persists across all efficacy levels.

When extending the analytical model to include three and four laboratories, the diminishing marginal effect of additional testing becomes evident. The probability that three laboratories independently agree on an efficacy classification decreases sharply across all trials. While the CP correspondingly increases, suggesting that the result becomes more conservative from a statistical point of view, the rate of matching classifications decreases. In the intermediate effective range, the inclusion of a third laboratory reduced the probability of an efficacy classification to between 0.0% (R2022-01) and 2.7% (R2023-01). The probability that three laboratories independently confirm a positive result is therefore extremely low, underscoring how additional replication exponentially decreases the likelihood of unanimous "effective" assessments in borderline efficacy conditions. The probability if four laboratories need to be effective is less than 1 of each of the ring trials in the intermediate range (see Table 3).

For the effective range, the same trend persists but to a lesser extent. The probability of efficacy classification for three laboratories ranged from 28.1% (R2020-01) to 69.2% (R2022-01), while projections for four laboratories indicate further moderate reductions to 17.7%–60.5%, depending on the microbial spectrum and test type. The analytical solution thus highlights a law of diminishing returns: beyond two laboratories, the additional gain in reliability is progressively outweighed by an increased probability of non-concordant outcomes caused by random intralaboratory fluctuations. This relationship reflects the mathematical property of multiplicative probability reduction, where independent errors or variations increase with the number of laboratories.

From a methodological and regulatory point of view, these results suggest that the requirement of two independent accredited laboratories, as specified by VAH and the Commission for Infection Prevention and Hygiene in Healthcare and Nursing (KRINKO) [10] and similar approaches adopted by the German Veterinary Medical Society (DVG) [11], the Organization for Economic Cooperation and Development (OECD) [12], the Austrian Society for Hygiene, Microbiology and Preventive Medicine (ÖGHMP) [13], the U.S. Environmental Protection Agency (EPA) and the U.S. Food and Drug Administration (FDA) [14], represents an optimal balance between analytical robustness and practical feasibility. This aspect is particularly relevant in the context of the European Biocidal Products Regulation (BPR), which promotes active substance minimization and thus leads to products being positioned close to the efficacy threshold. The present data demonstrate that precisely in these decision-critical

ranges, variability between laboratories is highest, increasing the risk of inconsistent classifications when relying on a single laboratory. The requirement for two independent laboratories therefore represents a scientifically justified and practically appropriate standard [15]. While expanding to three or four laboratories theoretically yields a more conservative decision threshold, the corresponding increase in cost, time, and organizational complexity is unlikely to be justified by the marginal improvement in statistical confidence. Moreover, excessive replication may paradoxically amplify the perception of uncertainty by reducing the frequency of unanimous positive results, despite no change in the underlying efficacy of the product. These effects are confirmed by the theoretical data set, presented in Figure 3, where scenarios with different numbers of effective results are calculated from a total of 50 potential data sets or laboratories. The fewer effective data sets there are, the more difficult it becomes for two or more identical results to be generated independently of each other.

In a broader context, the analytical model underscores the necessity of distinguishing between variability that can be reduced through methodological standardization (e.g., consistent inoculum preparation, media quality, and procedural training) and variability that is inherent to biological test systems and thus cannot be fully eliminated. While ring trials remain an indispensable tool for quality assurance and standardization, the analytical framework introduced here offers a complementary instrument for predicting and managing the uncertainty inherent in multi-laboratory efficacy assessments.

Ultimately, this study demonstrates that the analytical solution confirms the two-laboratory rule as an empirically and theoretically justified standard for disinfectant efficacy evaluation. More than two laboratories, while potentially beneficial in special cases, appears to yield progressively smaller improvements in classification reliability. In this respect, the present study complements prior simulation-based evidence with an exact analytical foundation, offering a rigorous mathematical justification for the existing VAH approach and providing basis for refining future guidelines on disinfection efficacy testing.

Conclusions

The analytical results presented here provide a quantitative foundation for understanding the relationship between laboratory number and the reliability of disinfectant efficacy classification. Building on extensive VAH ring trial data, the analytical model demonstrates that while the inclusion of a second independent laboratory substantially improves classification robustness and minimizes the likelihood of false-positive efficacy assessments, further expansion to three or four laboratories yields only marginal additional benefit.

Beyond two laboratories, the gain in statistical certainty approaches an asymptotic limit, while the probability of discordant outcomes increases disproportionately due

to compounding intralaboratory variation. These findings confirm that the two-laboratory stipulation, as currently required by the VAH and recommended by KRINKO [1], [10] and other organisations, represents an empirically and analytically optimal standard for ensuring reproducibility without an unnecessary burden of testing costs and complexity.

From a regulatory perspective, this analytical framework underscores the importance of directing future quality assurance endeavours towards reducing intralaboratory variation. This can be achieved through standardization of test conditions, the provision of operator training, and the harmonization of test procedures, rather than expanding the number independent laboratories and expert opinions. Strengthening reproducibility within laboratories is deemed to be more efficacious with regard to enhancing the precision of efficacy testing than merely increasing interlaboratory replication beyond two laboratories.

In summary, the study provides further theoretical and empirical justification, in line with existing literature, for maintaining the two-laboratory rule as the cornerstone of reliable disinfectant efficacy assessment. The analytical solution offers a mathematically sound complement to simulation-based evidence and supports evidence-driven standardization in accordance with European and international disinfection testing frameworks.

Notes

Authors' ORCIDs:

- Roesch K: <https://orcid.org/0009-0000-3494-9191>
- Rausch M: <https://orcid.org/0000-0003-1562-4337>
- Exner M: <https://orcid.org/0000-0002-6383-7866>
- Gemein S: <https://orcid.org/0009-0006-7329-9404>
- Ilschner C: <https://orcid.org/0009-0006-4083-7405>
- Kramer A: <https://orcid.org/0000-0003-4193-2149>
- Selhorst T: <https://orcid.org/0000-0003-0969-1533>
- Suchomel M:
<https://orcid.org/0000-0001-8758-9652>
- Mutters NT:
<https://orcid.org/0000-0002-0156-9595>
- Gebel J: <https://orcid.org/0000-0001-9328-3174>

Funding

This work was funded by the Association for Applied Hygiene (VAH), Germany.

Acknowledgments

The authors are grateful for the input provided by all laboratories which participated in the VAH ring trials.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

The authors Roesch K and Rausch M contributed equally.

References

1. VAH Disinfectant Commission, VAH. Requirements and Methods for VAH Certification of Chemical Disinfection Procedures. Continuously updated comprehensive edition since April 2015. 2015 Apr [last accessed: 2026 Feb 17]. Available from: <https://vah-online.de/de/fuer-laboratorien>
2. Roesch KM, Rausch M, Droop F, Exner M, Ilschner C, Kramer A, Selhorst T, Suchomel M, Mutters NT, Gebel J. Requirements for the precision and reproducibility in the efficacy testing of chemical disinfection procedures. *GMS Hyg Infect Control*. 2026 Mar 20;21:Doc33. DOI: 10.3205/dgkh000642
3. VAH. List of Ring Trial Final Reports (as of 31-12-2025). Available from: https://vah-online.de/files/download/Publikationsverzeichnis/Publikationsverzeichnis_Ringversuche_VAH_31.Dezember2025.pdf
4. Gebel J, Roesch KM, Steering Committee of VAH Ring Trials, Disinfectant Commission. 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, with *Staphylococcus aureus*. Bonn: VAH am Institut für Hygiene und Öffentliche Gesundheit der Universität Bonn; 2023 Nov 10. Available from: https://vah-online.de/files/download/Ringversuche/Ring%20trial_VAH%202023-1.pdf
5. Gebel J, Roesch KM, Steering Committee of VAH Ring Trials, Disinfectant Commission. Final report VAH ring trial 2020-01. Chemical disinfectants and antiseptics - Quantitative suspension test for the evaluation of yeasticidal activity - (Phase 2, Step 1) *Candida albicans*. Bonn: VAH am Institut für Hygiene und Öffentliche Gesundheit der Universität Bonn; 2021 Apr 26. Available from: https://vah-online.de/files/download/Ringversuche/Ring%20trial_VAH_2020-1.pdf
6. Gemein S, Gebel J, Steering Committee of VAH Ring Trials, Disinfectant Commission. Final report VAH ring trial 2022-01. Chemical disinfectants and antiseptics - Quantitative test method on non-porous surfaces without mechanical action - DIN EN 17387:2021 respectively VAH method 14.1: 2015 (Phase 2, Step 2) *Enterococcus hirae*. Bonn: VAH am Institut für Hygiene und Öffentliche Gesundheit der Universität Bonn; 2022 Sep 13. Available from: https://vah-online.de/files/download/Ringversuche/Ring%20trial_VAH_2022_1.pdf
7. Gebel J, Steering Committee of VAH Ring Trials, Disinfectant Commission. Final report VAH ring trial 2019-02. Chemical disinfectants and antiseptics - Quantitative carrier test - (Phase 2, Step 2) *Mycobacterium terrae*. Bonn: VAH am Institut für Hygiene und Öffentliche Gesundheit der Universität Bonn; 2020 Apr 21. Available from: https://vah-online.de/files/download/Ringversuche/Ring_trial_VAH_2019_2.pdf
8. 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. *Hyg Med*. 2024;49(4):D10-D6.
9. Suchomel M, Kampf G, Gebel J, Droop F, Christiansen B, Roesch KM. How reliable are test results from 17 laboratories on the basis of EN 1500 for a hand rub based on 80% (w/w)? *J Hosp Infect*. 2024 May;147:188-96. DOI: 10.1016/j.jhin.2024.01.022

10. Kommission für Krankenhaushygiene und Infektionsprävention (KRINKO). Stellungnahme der Kommission für Krankenhaushygiene und Infektionsprävention (KRINKO) zu Anforderungen an Desinfektionsmittel für den Einsatz in infektionshygienisch sensiblen Bereichen]. *Epid Bull.* 2023 Jun 8;23:22-6. DOI: 10.25646/11517
11. DVG. Voraussetzungen für die Desinfektionsmittelprüfung und Aufnahme in die Desinfektionsmittellisten der DVG. Stand: 12.02.2023]. 2023 Feb 12. Available from: https://www.desinfektion-dvg.de/fileadmin/FG_Desinfektion/Dokumente/Fuer_Gutachter/Pruefrichtlinien/3-Voraussetzungen_12Feb2023.pdf
12. OECD. Guidance Document on Quantitative Methods for Evaluating the Activity of Microbicides Used on Hard Non-Porous Surfaces. Series on Testing and Assessment No. 187. 2013 Jun 21. (Series on Biocides No. 6). Available from: https://www.oecd.org/content/dam/oecd/en/publications/reports/2013/06/guidance-document-on-quantitative-methods-for-evaluating-the-activity-of-microbicides-used-on-hard-non-porous-surfaces_b2be38a2/cb91d642-en.pdf
13. ÖGHMP. Expertenverzeichnis der Österreichischen Gesellschaft für Hygiene, Mikrobiologie und Präventivmedizin (ÖGHMP). last updated: 2026. Available from: <https://experten.oeghmp.at/pages/vorwort>
14. EPA. Antimicrobial Testing Methods & Procedures Developed by EPA's Microbiology Laboratory - SOP?MB-15-05: Standard Operating Procedure for the AOAC Sporicidal Activity of Disinfectants Test (Bacillus subtilis x porcelain carrier). Washington, DC: EPA; 2021.
15. Regulation (EU) No 528/2012 of the European Parliament and of the Council of 22 May 2012 concerning the making available on the market and use of biocidal products. *Off J Eur Union.* 2012;L167:1-123.

Corresponding author:

Kira-Marie Roesch, M.Sc.
 Institute for Hygiene and Public Health, University Hospital
 Bonn, Venusberg-Campus 1, Building 64, Room 21, 2nd
 floor, 53127 Bonn, Germany; Phone: +49 228
 287-11982
kira-marie.roesch@ukbonn.de

Please cite as

Roesch KM, Rausch M, Droop F, Exner M, Gemein S, Ilchner C, Kramer A, Selhorst T, Suchomel M, Mutters NT, Gebel J. How many laboratories are necessary? Analytical assessment of precision and reproducibility in disinfectant efficacy testing for quantifying the impact of laboratory number on result reliability. *GMS Hyg Infect Control.* 2026;21:Doc46.
 DOI: 10.3205/dgkh000655, URN: urn:nbn:de:0183-dgkh0006558

This article is freely available from

<https://doi.org/10.3205/dgkh000655>

Published: 2026-06-05

Copyright

©2026 Roesch et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.