



Opinion: Methodological Shortcomings in the Study on a Prophage-based PCR Test for Lyme Borreliosis

Freek R. van de Schoor^{1†}, M. E. Baarsma^{2†}, Mariska M. G. Leeflang³, Volker Fingerle^{4,5}, Gabriele Margos^{4,5}, Joppe W. Hovius^{2,5} and Alje P. van Dam^{2,5*}

¹ Radboudumc, Department of Internal Medicine, Radboudumc Center for Infectious Diseases and Radboud Institute of Health Sciences, Nijmegen, Netherlands, ² Center for Experimental and Molecular Medicine, Amsterdam Institute for Infection and Immunology, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands, ³ Epidemiology and Data Science, Amsterdam Public Health, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands, ⁴ German National Reference Centre for Borrelia, Bavarian Health and Food Safety Authority, Oberschleibheim, Germany, ⁵ ESCMID Study Group for Lyme Borreliosis, Basel, Switzerland

OPEN ACCESS

Edited by:

Mattias Collin,
Lund University, Sweden

Reviewed by:

Catherine Ayn Brissette,
University of North Dakota,
United States

*Correspondence:

Alje P. van Dam
a.p.vandam@amsterdamumc.nl

†These authors share first authorship

Keywords: *Borrelia burgdorferi*, Lyme borreliosis, PCR, diagnosis, prophage

We read the article by Shan et al. (2021) with great interest, as new diagnostic tests for Lyme borreliosis (LB) are urgently needed (Cruickshank et al., 2018; Dessau et al., 2018). The article represents a proof of principle paper and an initial validation of an already commercially available test [Phelix Phage Borrelia—R.E.D. Laboratories (redlabs.be)]. We have various concerns regarding the study design, novelty of the approach, technical aspects of the assay, statistical analyses, and the conclusions, which must be addressed. Of note, several statements in the introduction are speculative and not supported by the references, but unfortunately, the word limit of our opinion does not allow us to elaborate on this.

GENE SEQUENCES AND PCR RESULTS

The concept of targeting genetic material from bacteriophages rather than from bacteria for clinical diagnosis is intriguing and—while not entirely new—it is still relevant today (Amouriaux et al., 1993). In a previous publication, Amouriaux et al. (1993) describe a similar approach targeting a plasmid region with sequence overlap to sequences used in the current publication (Figure 1A). However, the authors do not prove that bacteriophages are present and circulating in human blood. Therefore, the difference in sensitivity between the 16S PCR and *terL* PCR could actually be due to the difference in sensitivity between using a single-copy (16S) and a multi-copy (*terL*) target. This principle is well-known in bacteriology (Roosendaal et al., 1993). In addition, the genetic variation between and within the *Borrelia burgdorferi* sensu lato (s.l.) species of cp32 bacteriophage sequences is not discernible from the manuscript. In the alignment shown in Figure 2 in the article by Shan et al., the authors use cp32 genes of *B. burgdorferi* sensu stricto (s.s.), but not of other species. *B. burgdorferi* s.s. B31 has the highest number of cp32 ($n = 13$) in comparison to 16S rRNA, a single copy chromosomal locus. We would be very interested to see how *Borrelia afzelii* and *Borrelia garinii* would have performed in analyses using spiked blood samples, as these

Specialty section:

This article was submitted to
Infectious Agents and Disease,
a section of the journal
Frontiers in Microbiology

Received: 26 October 2021

Accepted: 08 November 2021

Published: 13 December 2021

Citation:

Van de Schoor FR, Baarsma ME, Leeflang MMG, Fingerle V, Margos G, Hovius JW and Van Dam AP (2021) Opinion: Methodological Shortcomings in the Study on a Prophage-based PCR Test for Lyme Borreliosis. *Front. Microbiol.* 12:802131. doi: 10.3389/fmicb.2021.802131

are the most common genospecies causing clinical symptoms in Europe, but have—according to the authors' Table 1—fewer cp32 plasmids ($n = 8$, $n = 4$, respectively). In their paper, the geographical origin of patients is not described, but the authors state that “patients were diagnosed by Dr. LT,” referring to Louis Teulières, who is based in Paris, France, where—like in the rest of the European continent—LB is caused mainly by *B. afzelii* and *B. garinii* (Stanek et al., 2012). However, the test method is based on the *terL* gene derived from the North American strain of *B. burgdorferi* s.s. B31. Furthermore, the authors include extremely low-positive signals in their results. Of the 23 healthy controls, 21 showed a positive signal in the *terL* PCR in at least one of the 12 samples. Whereas incidental carriage of *Borrelia*-DNA in blood of healthy persons, as suggested by the authors, might occur, it is highly unlikely that this would be found in over 90% of the population. This strongly suggests that at least some low-positive results represent unspecific signals or signals which are a result of DNA cross-contamination.

SELECTION CRITERIA LB PATIENTS

Another concern is the patient selection and interpretation of the clinical data. The manuscript lacks any description of patient characteristics, and does not report inclusion or exclusion criteria. Absence of clear eligibility criteria may indicate selection bias. Criteria for patient selection in an LB-related diagnostic test accuracy study should be clear and unambiguous, for example based on European guidelines (Mygland et al., 2010; Stanek et al., 2011; Hofmann et al., 2017). The authors refer to the ILADS guideline (Cameron et al., 2014), which in itself does not contain any diagnostic criteria. Without unambiguous criteria, one cannot ensure that these individuals were in fact patients with LB (Stanek et al., 2011; Lantos et al., 2021). It is also unclear what is meant by “early LD” and “late LD.” Would Lyme neuroborreliosis (LNB) be classified as early or late LD, for example (Koedel and Pfister, 2017)? Were there any LNB patients at all? If so, how were they diagnosed?

STATISTICAL ANALYSIS

We attempted to replicate the analyses presented by the authors in their Figure 7, using SPSS version 25. The authors describe having used Mann-Whitney U-tests to compare early LB patients, late LB patients, and healthy volunteers (HVs). However, they do not describe precisely how the results from the different groups have been compared. Their original dataset contains six test results on whole blood (WB) and six on serum for each participant, but it is unclear whether they analyzed all results, if they analyzed the mean per participant or used any transformation of the data. We replicated the Mann-Whitney U-tests to test for a difference between the different participant groups, using the mean values of the six iterations of each test per participant. While the authors' reported means and the means calculated by us were identical, our p -values were inconsistent with those reported by the authors. Subsequent analyses using other aggregate functions (such as medians) as

input for our statistical tests did not result in p -values consistent with those reported by the authors either (data not shown). In contrast, when we used the six iterations of the *terL* assay per participant separately, as if they were independent values, the levels of statistical significance match those reported by the authors. By doing so the authors seem to have artificially inflated their statistical power by increasing their sample size six-fold. This may have resulted in identical mean values, but incorrect and much lower p -values. More so, the results from serum and WB samples from one participant are not independent, as both measurements were done in the same person. Therefore, a Wilcoxon Signed-Rank test would have been more appropriate to compare serum and WB within one patient group. The authors do not describe what statistical test they used, but if this was a Mann-Whitney U-test as described in the Methods section, then this is inappropriate.

THE CONCLUSIONS ARE NOT SUPPORTED BY THE DATA

The aforementioned considerations cast substantial doubt on the reliability of the results, but—when interpreted with caution—do not undermine the value of the authors' hypothesis. Unfortunately, the conclusions drawn by the authors from the results are inappropriate. The authors state that their assay can distinguish early LB, late LB, and HVs. These conclusions are not supported by the data.

The mean/median copy numbers may be significantly different at a group level—even though we have shown in this manuscript's **Figures 1B,C** and **Supplementary Table 1** that they are not for most comparisons—but that does not imply diagnostic power. Only if there is little or no overlap between numeric values, will the assay be able to distinguish a patient from a non-patient. A simple scatterplot of the data shows there is a high degree of overlap between the groups. Subsequent ROC-analysis **Figures 1C,D** on the mean/median copy numbers shows that—when a minimally acceptable specificity of 90% is applied—the maximally attainable sensitivity is 62% (WB-MEAN: HV vs. early, cutoff at 1.275) or 57% (WB-MEAN: HV vs. late, cutoff at 1.283). This is worse than single-tier or modified two-tiered testing (MTTT) serology in EM and far worse than any type of serology in late LB (Leefflang et al., 2016; Waddell et al., 2016; Branda et al., 2017). Additional ROC-analyses show that the ability to discriminate between early vs. late LB is even (much) worse (data not shown). Please note that these analyses were performed with a small number of samples (early: $n = 13$; late: $n = 42$; HV: $n = 23$). It is much more likely that the assay lacks specificity and that many HVs had false-positive results, rather than suffer from asymptomatic *B. burgdorferi* infection, as the authors claim.

We must also point out that the manuscript suffers from flawed circular reasoning and over-interpretation. The fact that the groups differ with respect to the primary study parameter does not prove that they are LB patients or HVs. Participants' status as belonging to either group is the starting point for investigating potential differences in *terL* levels, not a conclusion

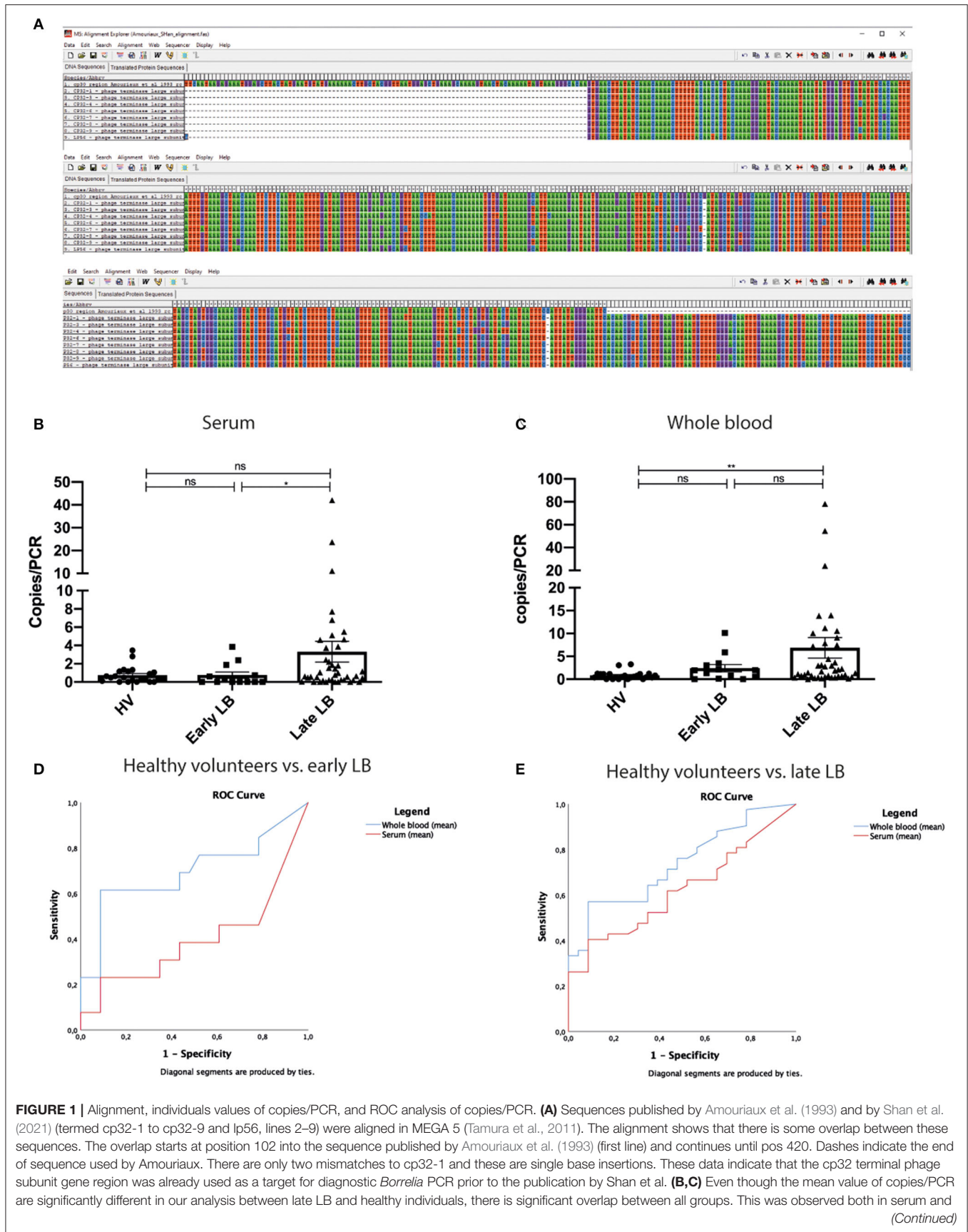


FIGURE 1 | Alignment, individuals values of copies/PCR, and ROC analysis of copies/PCR. **(A)** Sequences published by Amouriaux et al. (1993) and by Shan et al. (2021) (termed cp32-1 to cp32-9 and lp56, lines 2–9) were aligned in MEGA 5 (Tamura et al., 2011). The alignment shows that there is some overlap between these sequences. The overlap starts at position 102 into the sequence published by Amouriaux et al. (1993) (first line) and continues until pos 420. Dashes indicate the end of sequence used by Amouriaux. There are only two mismatches to cp32-1 and these are single base insertions. These data indicate that the cp32 terminal phage subunit gene region was already used as a target for diagnostic *Borrelia* PCR prior to the publication by Shan et al. **(B,C)** Even though the mean value of copies/PCR are significantly different in our analysis between late LB and healthy individuals, there is significant overlap between all groups. This was observed both in serum and (Continued)

FIGURE 1 | in WB. **(D,E)** ROC curves of mean values WB and serum. **(C)** Comparison of healthy volunteers to early LD. Whole blood AUC = 0.697(95%CI 0.496–0.899). Serum AUC = 0.400 (95%CI 0.187–0.612). **(D)** Comparison of healthy volunteers and late LD. Whole blood AUC = 0.738 (95%CI 0.618–0.858). Serum AUC = 0.622 (95%CI 0.487–0.757). LB, Lyme borreliosis; WB, whole blood; ROC, receiver operating characteristic; AUC, area under the curve; CI, confidence interval; ns, non-significant; * $p < 0.05$, ** $p < 0.01$, calculated using independent-samples Mann-Whitney U-test for comparing participant groups; dependent-samples Wilcoxon Signed Rank test for comparing whole-blood vs. serum within a participant group.

that can be drawn when these groups are indeed shown to be different on this outcome. The authors postulate that their test could be used to monitor LB treatment outcomes, yet, this study does not report on any follow-up samples or treatment outcome to support this claim. They further state that the Ter-qPCR could be used to indicate which treatment option may work best, however, the choice of treatment option is not supported by any of the data in this article.

CONCLUSIONS

We conclude that while this technique might be promising, the paper provides more questions than answers and contains a large number of inaccuracies. We would be interested to see the Ter-qPCR be validated on a cohort of clearly described LB patients and healthy controls from both North America and Europe before we could draw any conclusions on the diagnostic performance of the Ter-qPCR.

AUTHOR CONTRIBUTIONS

FS and MEB: conceptualization, methodology, software, formal analysis, resources, and writing—original draft. FS, MEB, and

GM: investigation. FS, MEB, ML, VF, GM, and JH: writing—review and editing. JH: supervision and funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

FS, MEB, and JH are funded by the Netherlands Organization for Health Research and Development (ZonMw, Project Number 522050001). JH's work was partially funded through the European Regional Development Fund and the Interreg North Sea Region Programme 2014–2020 as part of the NorthTick project (Reference Number J-No: 38-2-7-19).

ACKNOWLEDGMENTS

Robert-Koch Institute for funding the NRZ Borrelia.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2021.802131/full#supplementary-material>

REFERENCES

- Amouriaux, P., Assous, M., Margarita, D., Baranton, G., and Saint Girons, I. (1993). Polymerase chain reaction with the 30-kb circular plasmid of *Borrelia burgdorferi* B31 as a target for detection of the Lyme borreliosis agents in cerebrospinal fluid. *Res. Microbiol.* 144, 211–219. doi: 10.1016/0923-2508(93)90046-5
- Branda, J. A., Strle, K., Nigrovic, L. E., Lantos, P. M., Lepore, T. J., Damle, N. S., et al. (2017). Evaluation of modified 2-tiered serodiagnostic testing algorithms for early Lyme disease. *Clin. Infect. Dis.* 64, 1074–1080. doi: 10.1093/cid/cix043
- Cameron, D. J., Johnson, L. B., and Maloney, E. L. (2014). Evidence assessments and guideline recommendations in Lyme disease: the clinical management of known tick bites, erythema migrans rashes and persistent disease. *Expert Rev. Anti Infect. Ther.* 12, 1103–1135. doi: 10.1586/14787210.2014.940900
- Cruikshank, M., O'Flynn, N., and Faust, S. N. (2018). Lyme disease: summary of NICE guidance. *BMJ* 361, k1261. doi: 10.1136/bmj.k1261
- Dessau, R. B., van Dam, A. P., Fingerle, V., Gray, J., Hovius, J. W., Hunfeld, K. P., et al. (2018). To test or not to test? Laboratory support for the diagnosis of Lyme borreliosis: a position paper of ESGBOR, the ESCMID study group for Lyme borreliosis. *Clin. Microbiol. Infect.* 24, 118–124. doi: 10.1016/j.cmi.2017.08.025
- Hofmann, H., Fingerle, V., Hunfeld, K. P., Huppertz, H. I., Krause, A., Rauer, S., et al. (2017). Cutaneous Lyme borreliosis: guideline of the German Dermatology Society. *Ger. Med. Sci.* 5, Doc14. doi: 10.3205/000255
- Koedel, U., and Pfister, H. W. (2017). Lyme neuroborreliosis. *Curr. Opin. Infect. Dis.* 30, 101–107. doi: 10.1097/QCO.0000000000000332
- Lantos, P. M., Rumbaugh, J., Bockenstedt, L. K., Falck-Ytter, Y. T., Aguero-Rosenfeld, M. E., Auwaerter, P. G., et al. (2021). Clinical practice guidelines by the Infectious Diseases Society of America, American Academy of Neurology, and American College of Rheumatology: 2020 Guidelines for the prevention, diagnosis, and treatment of Lyme disease. *Neurology* 96, 262–273. doi: 10.1212/WNL.00000000000011151
- Leeflang, M. M., Ang, C. W., Berkhout, J., Bijlmer, H. A., Van Bortel, W., Brandenburg, A. H., et al. (2016). The diagnostic accuracy of serological tests for Lyme borreliosis in Europe: a systematic review and meta-analysis. *BMC Infect. Dis.* 16, 140. doi: 10.1186/s12879-016-1468-4
- Mygland, A., Ljostad, U., Fingerle, V., Rupprecht, T., Schmutzhard, E., and Steiner, I. (2010). EFNS guidelines on the diagnosis and management of European Lyme neuroborreliosis. *Eur. J. Neurol.* 17, 8–16, e1–e4. doi: 10.1111/j.1468-1331.2009.02862.x
- Roosendaal, R., Walboomers, J. M., Veltman, O. R., Melgers, I., Burger, C., Bleker, O. P., et al. (1993). Comparison of different primer sets for detection of *Chlamydia trachomatis* by the polymerase chain reaction. *J. Med. Microbiol.* 38, 426–433. doi: 10.1099/00222615-38-6-426
- Shan, J., Jia, Y., Teulière, L., Patel, F., and Clokie, M. R. J. (2021). Targeting multicopy prophage genes for the increased detection of *Borrelia burgdorferi* Sensu Lato (s.l.), the causative agents of Lyme disease, in blood. *Front. Microbiol.* 12, 651217. doi: 10.3389/fmicb.2021.651217
- Stanek, G., Fingerle, V., Hunfeld, K. P., Jaulhac, B., Kaiser, R., Krause, A., et al. (2011). Lyme borreliosis: clinical case definitions for diagnosis and management in Europe. *Clin. Microbiol. Infect.* 17, 69–79. doi: 10.1111/j.1469-0691.2010.03175.x
- Stanek, G., Wormser, G. P., Gray, J., and Strle, F. (2012). Lyme borreliosis. *Lancet* 379, 461–473. doi: 10.1016/S0140-6736(11)60103-7

- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121
- Waddell, L. A., Greig, J., Mascarenhas, M., Harding, S., Lindsay, R., and Ogden, N. (2016). The accuracy of diagnostic tests for Lyme disease in humans, a systematic review and meta-analysis of north american research. *PLoS ONE* 11, e0168613. doi: 10.1371/journal.pone.0168613

Conflict of Interest: MEB and JH: LB diagnostics in collaboration with various companies, although none of that work involved molecular detection of *B. burgdorferi* sensu lato. MEB and JH have not received any personal compensation from any of said companies, nor were any of said companies involved in any aspect of the current manuscript. VF: Research support: RKI/BMG, ESCMID, ECDC, StMG/STMUG, INSTAND. Lecturing activities (honoraria, travel expenses): DIAMEDIS, Diasorin, Mikrogen, Seramun, Siemens, HLR. Consulting activities EQA schemes (no honoraria): QCMD, INSTAND, ECDC.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Van de Schoor, Baarsma, Leeflang, Fingerle, Margos, Hovius and Van Dam. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.