3-23-2021

Discussion of "Is group testing ready for prime-time in disease Identification?" by Haber, Malinovsky, and Albert, Statistics in Medicine, 2021

Brad J. Biggerstaff

Received: 23 March 2021 | Accepted: 25 March 2021

DOI: 10.1002/sim.8989

COMMENTARY

Statistics in Medicine WILEY

Discussion of "Is group testing ready for prime-time in disease Identification?" by Haber, Malinovsky, and Albert, *Statistics in Medicine*, 2021

Brad J. Biggerstaff

Division of Vector-Borne Diseases, Centers for Disease Control and Prevention, Fort Collins, Colorado, USA

Correspondence
Brad J. Biggerstaff, Division of Vector-Borne Diseases, Centers for Disease Control and Prevention, Fort Collins, CO.
Email: bkb5@cdc.gov

I congratulate the authors on an interesting, instructive, and timely paper. Certainly, including the testing for SARS-CoV-2 example in a paper on a topic—group testing—of wide interest during the current COVID-19 pandemic emphasizes the relevance of the topic in public health and diagnostic medicine generally. But the thoughtful inclusion of the remaining examples—HIV, HPV, and cancer biomarker detection—illustrates that the issues considered are important even in less urgent times. In this note of discussion I will comment on two particular aspects in the paper that I found particularly informative, and then I will comment on two other, related areas that I believe the conclusions inform.

I found the direct use of explicit constraints on $Se(k)$ and $Sp(k)$ in design optimization natural and, as the authors say, "very easy to interpret and explain to non-statisticians." Incorporating these constraints at the outset of design optimization should have the further, philosophical advantage that consideration of and specification of these must (ideally) be made by subject matter experts *before* cost evaluations are undertaken, so that screening and diagnostic needs rather than cost constraints do not unduly drive (or result in inadvertant or active manipulation of) performance requirements.

The authors' investigation of the use of uncertain values or estimates for $Se(k)$ and $Sp(k)$ is instructive and important, both for use of the Dorfman procedure considered here and for when considering other testing strategies. The analysis presented demonstrates clearly some of the potential pitfalls of pretending to know too much. In this analysis, emphasis is on sampling directly from the population under study for the validation component in determining the use or not of group testing in a particular application. While this ideal situation should be employed whenever practical, as the authors note in the Discussion, more economical alternatives might be considered based on the technology underlying the assays or on population characteristics. To balance costs and evaluation, one might, for example, consider hybrids or combinations of both direct population sampling and laboratory contrived samples with well-controlled properties, such as the target-spiked samples the authors mention. Evaluation of such would almost certainly be needed on a case-by-case basis, but thorough consideration might be expected to lead to both improved testing efficiency and economic gains.

As presented, costs are characterized using the total number of tests performed per individual and total numbers of tests needed to warrant group testing. The authors do note that, "some authors include a cost for incorrect classification," but they then go to say that this "may be difficult to motivate from a medical or public health perspective." Detailed economic evaluation of interventions and screening tools are often encouraged (eg, References 1,2) to justify decisions and recommendations in public health, especially in times of decreasing budgets and increasing burdens. While such analyses can be difficult and intricate and would obviously require a great deal of detailed costing information, enlistment of health economists to collaborate in such an evaluation should provide more refined, complete, and defensible analyses. (Examples of readily available background information on economic analyses of this sort include the U.S. Department of

Health and Human Services [aspe.hhs.gov/system/files/pdf/74686/report.pdf] the U.S. Centers for Disease Control and Prevention [www.cdc.gov/policy/polaris/economics/cost-effectiveness.html] and the World Health Organization [www.who.int/choice/publications/p_2003_generalised_cea.pdf].) The good news here, of course, is that while such extensive analyses were not considered by the authors here, their demonstration of the need to correctly calculate group testing performance and numbers of tests required, especially in the presence of differential misclassification, moves the needle toward more rigorous evaluation. Implementation of these recommendations should, therefore, directly and more correctly inform more comprehensive economic analyses.

Finally, while the authors' focus was on group testing for identification, some of the lessons learned also apply to group testing for estimation. Correct specification of test performance parameters is essential in estimation applications that account for imperfect tests, especially if values are treated as known and so uncertainty in these is not incorporated into inferential statements. In applications involving estimation of virus prevalence in mosquito populations based on samples of trapped mosquitoes, for example, test performance characteristics may be included in estimation based on recent work[3] applying Firth's correction to effectively eliminate bias in estimation using imperfect tests and allowing for group size-dependent sensitivities and specificities (ie, for differential misclassification). Consider a simple, realistic, but contrived example with 20 groups (pools) each of sizes 10, 25, 50, 100, with 1, 0, 3, 5 of these groups positive, respectively. The percent difference in estimates of the population proportion positive based on this varied more than 8%, from 2.8 to 3.1 per 1000, when group-specific sensitivies were changed from (1.0, 0.975, 0.95, 0.9) to (1, 0.95, 0.9, 0.8). The difference using the latter sensitivities was over 15% when compared with perfect testing. Though they may appear small, such differences in estimates may impact governmental agency decisions regarding community interventions for mosquito abatement. Laboratory work (eg, Reference 4) characterizing sensitivies and specificies of assays for mosquito testing by group size can be used to inform such estimation. Another key recommendation of the authors that applies to estimation as well is to be sure to define carefully the population to which test characteristics apply. In the estimation of virus infection prevalence in mosquitoes, careful attribution of testing characteristics to both virus and mosquito species would be essential for proper implementation in estimation.

ORCID

Brad J. Biggerstaff https://orcid.org/0000-0002-3105-3530

REFERENCES

1. Iragorri N, Spackman E. Assessing the value of screening tools: reviewing the challenges and opportunities of cost-effectiveness analysis. *Public Health Rev*. 2018;39:17. https://doi.org/10.1186/s40985-108-0093-8.
2. Rabarison KM, Bish CL, Massoudi MS, et al. Economic evaluation enhances public health decision making. *Front Public Health*. 2015;3:164. https://doi.org/10.3389/fpubh.2015.00164.
3. Hepworth G, Biggerstaff BJ. Bias correction in estimating proportions by imperfect pooled testing. *J Agric Biol Environ Stat*. 2021;26(1):90-104. https://doi.org/10.1007/s13253-020-00411-5.
4. Burkhalter KL, Horiuchi K, Biggerstaff BJ, Savage HM, Nasci RS. Evaluation of a rapid analyte measurement platform and real-time reverse-transcriptase polymerase chain reaction assay West Nile virus detection system in mosquito pools. *J Am Mosq Control Assoc*. 2014;30(1):21-30.

How to cite this article: Biggerstaff BJ. Discussion of "Is group testing ready for prime-time in disease Identification?" by Haber, Malinovsky, and Albert, *Statistics in Medicine*, 2021. *Statistics in Medicine*. 2021;40:3887–3888. https://doi.org/10.1002/sim.8989