multiple cutoffs per study in a single analysis, modeling the empirical distribution function and thus avoiding paradox behavior. Like Levis et al. (1), we performed the analysis twice, first for the cutoffs that were publicly reported and then for all data. The results for selected cutoff values are Table 1. We found that sensitivities and specificities were similar for both analyses and well approximated the results of the separate IPD meta-analyses by Levis et al. (1). For the sensitivities, our confidence intervals tended to be narrower than those by Levis et al., reflecting our use of data for all available studies (and cutoffs) simultaneously.

We conclude that by using the multiple-cutoffs model we were able to anticipate the IPD analysis even when using only the part of the data that was selected for publication. Investigators facing primary studies presenting multiple cutoffs should consider analyzing all available data using a model developed particularly for this situation.

ACKNOWLEDGMENTS

This work was supported by the German Research Foundation (grant number DFG RU 1747/1-2) to G.R.

Conflict of interest: none declared.

REFERENCES

1. Levis B, Benedetti A, Levis AW, et al. Selective cutoff reporting in studies of diagnostic test accuracy: a comparison of conventional and individual-patient-data meta-analyses of the Patient Health Questionnaire-9 depression screening tool. *Am J Epidemiol*. 2017;185(10):954–964.
2. Macaskill P, Gatsonis C, Deeks J, et al. *Handbook for DTA Reviews* (Chapter 10). http://methods.cochrane.org/sites/methods.cochrane.org.sdt/files/public/uploads/Chapter%2010%20-%20Version%201.0.pdf. Published December 23, 2010. Accessed May 16, 2017.
3. Rücker G, Schumacher M. Summary ROC curve based on a weighted Youden index for selecting an optimal cutpoint in meta-analysis of diagnostic accuracy. *Stat Med*. 2010;29(30):3069–3078.
4. Rücker G, Schumacher M. Procalcitonin as a diagnostic marker for sepsis [letter]. *Lancet Infect Dis*. 2013;13(12):1012–1013.
5. Hamza TH, Arends LR, van Houwelingen HC, et al. Multivariate random effects meta-analysis of diagnostic tests with multiple thresholds. *BMC Med Res Methodol*. 2009;9:73.
6. Putter H, Fiocco M, Stijnen T. Meta-analysis of diagnostic test accuracy studies with multiple thresholds using survival methods. *Biom J*. 2010;52(1):95–110.
7. Riley RD, Takwoingi Y, Trikalinos T, et al. Meta-analysis of test accuracy studies with multiple and missing thresholds: a multivariate-normal model. *J Biomet Biostat*. 2014;4:12.
8. Steinhauser S, Schumacher M, Rücker G. Modelling multiple thresholds in meta-analysis of diagnostic test accuracy studies. *BMC Med Res Methodol*. 2016;16(1):97.

Gerta Rücker[1], Susanne Steinhauser[2], and Martin Schumacher[1]
(e-mail: ruecker@imbi.uni-freiburg.de)
[1] *Institute for Medical Biometry and Statistics, Medical Faculty and Medical Center – University of Freiburg, Freiburg, Germany*
[2] *Institute of Medical Statistics, Informatics and Epidemiology, University of Cologne, Cologne, Germany*

DOI: 10.1093/aje/kwx275; Advance Access publication: August 23, 2017

THREE AUTHORS REPLY

Rücker et al. (1) responded to our study of selective cutoff reporting in studies of diagnostic test accuracy (2), agreeing that cutoff selection is indeed a problem in primary studies and meta-analyses of diagnostic test accuracy. They described a model that they have developed—the multiple-cutoffs model—that allows for the inclusion of multiple cutoffs per study in a single analysis of aggregate data, modeling the distribution function rather than each point of the receiver operating characteristic curve separately (3). Using their multiple-cutoffs model, they reanalyzed our data in 2 ways: first, they included only cutoffs that were published in the original primary studies; second, they included all data from our individual participant data (IPD) meta-analysis, using all cutoffs for all studies (1). Rücker et al. found that, based on their model, sensitivities and specificities were similar for both sets of analyses and well approximated the results of our bivariate random-effects IPD meta-analyses, which included all cutoffs for all studies (1). Rücker et al. also found that, using their model, confidence intervals for sensitivity estimates tended to be narrower than our confidence intervals, and they attributed this finding to fact that that their model uses data for all available studies and cutoffs simultaneously (1). They concluded that, by using their model, they were able to approximate our IPD results, even when using diagnostic accuracy data only from published cutoffs (1).

It would be highly advantageous to be able to use a modelling approach to approximate the performance of diagnostic tests across thresholds when some primary studies do not report all relevant cutoff data. The degree to which this can be done accurately, however, depends on the validity of the assumptions of the model. In the studies included in our IPD meta-analysis, most studies (11 of 13) published accuracy results for the standard Patient Health Questionnaire–9 cutoff score of 10, which was also the strongest-performing cutoff for maximizing combined sensitivity and specificity (2). Accuracy data from primary studies were missing symmetrically on either side of this cutoff. Thus, although we identified what appeared to be biased reporting of results from some cutoffs and not others, the pattern of missing accuracy data may not have been typical because of its symmetry and because the cutoff threshold that is recognized as standard in the field also seems to be the best-performing cutoff.

There are other examples from study-level meta-analyses of depression screening tools where many included studies do not report data from the cutoff threshold that is considered standard, presumably because that cutoff performed poorly. For instance, in the largest existing meta-analysis of the diagnostic accuracy of the Hospital Anxiety and Depression Scale for detecting major depressive disorder (4), the authors attempted to assess accuracy for the standard cutoff score of 8, but results from this cutoff were published for just over half of otherwise eligible studies.

The model developed by Rücker et al. is promising. However, it involves several unknowns, and it will be important to test how well it replicates the results of full IPD data when accuracy data are published only for a limited set of cutoffs in the original primary studies. In particular, it should perform well in meta-analyses that may not be anchored with robust data at the best-performing cutoff, and where included datasets have more skewed patterns of missing accuracy data.

ACKNOWLEDGMENTS

B.L., A.B., and B.D.T. contributed to drafting the response to the letter to the editor.

B.L. was supported by a Canadian Institutes of Health Research (CIHR) Frederick Banting and Charles Best Canada Graduate Scholarship doctoral award. A.B. and B.D.T. were supported by Fonds de recherche du Québec – Santé (FRQS) researcher salary awards.

Conflict of interest: none declared.

REFERENCES

1. Rücker G, Steinhauser S, Schumacher M. Re: "Selective cutoff reporting in studies of diagnostic test accuracy: a comparison of traditional and individual patient data meta-analyses of the Patient Health Questionnaire-9 depression screening tool." *Am J Epidemiol*. 2017;186(7):894–895.
2. Levis B, Benedetti A, Levis AW, et al. Selective cutoff reporting in studies of diagnostic test accuracy: a comparison of conventional and individual-patient-data meta-analyses of the Patient Health Questionnaire-9 depression screening tool. *Am J Epidemiol*. 2017;185(10):954–964.
3. Steinhauser S, Schumacher M, Rücker G. Modelling multiple thresholds in meta-analysis of diagnostic test accuracy studies. *BMC Med Res Methodol*. 2016;16(1):97.
4. Brennan C, Worrall-Davies A, McMillan D, et al. The Hospital Anxiety and Depression Scale: a diagnostic meta-analysis of case-finding ability. *J Psychosom Res*. 2010;69(4):371–378.

Brooke Levis[1,2], Andrea Benedetti[2,3,4], and Brett D. Thombs[1,2,3,5,6,7]
(e-mail: brett.thombs@mcgill.ca)
[1] *Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, QC, Canada*
[2] *Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, QC, Canada*
[3] *Department of Medicine, McGill University, Montréal, QC, Canada*
[4] *Respiratory Epidemiology and Clinical Research Unit, McGill University Health Centre, Montréal, QC, Canada*
[5] *Department of Psychology, McGill University, Montréal, QC, Canada*
[6] *Department of Psychiatry, McGill University, Montréal, QC, Canada*
[7] *Department of Educational and Counselling Psychology, McGill University, Montréal, QC, Canada*

DOI: 10.1093/aje/kwx276; Advance Access publication: August 23, 2017

RE: "THE HIDDEN EPIDEMIC OF FIREARM INJURY: INCREASING FIREARM INJURY RATES DURING 2001–2013"

In a recent issue of the *Journal*, Kalesan et al. (1) made the case for a hidden epidemic of firearm injury in the United States during the period from 2001 to 2013. They concluded that "[t]he epidemic of firearm violence, driven largely by nonfatal injuries, is an important public health problem" and referred to the increase in nonfatal injuries as a "public health emergency" (1, p. 552). Over the course of their 12-year study period, there was a 2.5% increase in the crude rate of deaths from firearms (the net result of a large reduction in the firearm homicide rate coupled with an increase in the firearm suicide rate) and a supposed 20.4% increase in the nonfatal injury rate (due entirely to the trend in assault-related injury). It is this unexpected increase in the nonfatal injury rate that undergirds their principal conclusions. As it turns out, however, the surveillance data from which they computed the trends in nonfatal firearm injuries are flawed, and the apparent upward trend is an artifact of these flaws. Well-supported adjustments to the apparent trend in nonfatal injuries resulting from firearm assaults eliminate the upward trend, as we demonstrated in a recent article (2).

Kalesan et al. estimated trends in nonfatal injuries that were primarily based on a nationally representative survey of hospital emergency departments. The National Electronic Injury Surveillance System–All Injury Program is managed by the Consumer Product Safety Commission (3). Annual estimates from 2001 onward are publically available on a website maintained by the Centers for Disease Control and Prevention (the Web-Based Injury Statistics Query and Reporting System, or WISQARS) (4). A closely related source of data on nonfatal firearm injuries is the Firearms Injury Surveillance System (NEISS-FISS), which is based on a somewhat expanded sample and includes more detail; in particular, it distinguishes between unintentional injuries and injuries of undetermined intent (5). We used the NEISS-FISS sample; because of data availability and other considerations, our analysis was focused on the period of 2003–2012. When examining the data as reported by Centers for Disease Control and Prevention, we found that the estimated trends in gunshot injuries from firearm assaults were very similar to those reported by Kalesan et al.; there was a 49% increase in the