Journal of
Bat Research & Conservation

OPEN ACCESS
www.secemu.org
DOI: 10.14709/BarbJ.15.1.2022.09

ORIGINAL ARTICLE

Bat Noise Scrubber Bias Adjustment Using the Rogan-Gladen Estimator and Bayesian Inference

Peter Ommundsen[1]

[1] Salt Spring Island Community Bat Program, 456 Scott Point Drive, Salt Spring Island, BC, V8K 2R2 Canada

*Corresponding author:
bats@capewest.ca

DOI: https://doi.org/10.14709/BarbJ.15.1.2022.09

Keywords: Bat acoustic, echolocation, noise filter, noise scrubber, Rogan-Gladen estimator

received: November, 21st 2022
accepted: December, 15th 2022

ABSTRACT

The number of bat call files recorded in acoustic surveys may be used to assess comparative bat activity levels over time and among habitats. Acoustic signal processing can segregate bat files from noise files and thus quickly provide an estimate of the number of bat files in a large sample of recordings. However, false positive and false negative classifications may result in a biased estimate requiring adjustment, as inaccurate bat numbers may impact bat conservation decisions. Previous research has ranked software classification accuracy in comparison to the visual classification of spectrograms. Small classification errors can result in considerable bias in software-derived estimates of the number of bat call files in a sample. Estimation bias may not have a linear relationship to the percentage of files containing bats, requiring unique correction coefficients. The focus of this note is to 1) illustrate patterns of bias that may result from noise scrubbing and 2) to illustrate the application of two methods of bias adjustment, the Rogan-Gladen estimator and Bayesian inference. The expected bias of four noise scrubbing tools from the literature, each of different measured accuracy, was plotted over a simulated range of true bat file prevalence while holding constant the accuracy of each scrubber. Rogan-Gladen bias adjustment was accurate for all four noise scrubbers. Bayesian bias adjustment showed low overall error, with some inflation at very low bat file prevalence. Caveats in the use of both bias adjustment methods are discussed.

INTRODUCTION

Acoustic monitoring of bats is currently a conservation priority (Loeb et al. 2015), as bats face a host of risk factors, including climate change, disease, habitat loss, pesticide poisoning, and wind farming (Frick et al. 2019). Bat detector recordings may inform inferences regarding relative bat activity levels and relative population density (Dzal et al. 2010, Brooks 2011, Roche et al. 2011, Nocera et al. 2019, Pinzari et al. 2019, Phinney 2020, de Jong et al. 2021, Harrison 2021, Whiting et al. 2022).

A bat detector may record tens of thousands of recordings within a short period of time, and researchers may employ software to isolate digital files containing bats from those containing only noise. Such algorithms have been termed classifiers, filters, recognizers, categorizers, and scrubbers. The software may output separate subfolders containing the bat files and noise files, or it may output a list of bat files. In any event, the data permit a scrubber-derived estimate of the proportion of bat files within the sample of recordings. Such automated classification enables fast processing of large datasets, but misclassification can impact bat conservation decisions.

Past research has reported on the software's accuracy in classifying files as bats or noise (Clement et al. 2014, Perea & Tena 2020), where accuracy refers to the agreement with visual classification. However, high classification accuracy (low numbers of false positives and false negatives) does not necessarily translate to high accuracy in the estimation of the proportion of bat files in a sample of recordings. Modest classification errors can cause large errors in estimates derived from scrubber output. For example, in a batch of 10000 recordings containing 500 bat files and 9500 noise files, software that classified the bat files and noise files with 95% accuracy would yield an estimate of 950 bat files (475 true positives + 475 false positives), almost twice the true number. The misclassification error is 5%, but the estimation error is 90%. The apparent number of bat files sorted by noise scrubbing can be a biased estimate of true bat file prevalence.

If the same software, with an accuracy of 95%, were used to scan a sample of 10000 files that contained 1000 bat files, it would yield an estimate of 1400 bat files (950 true positives + 450 false positives), an estimation error of 40%. The underlying accuracy of the noise scrubber is the same as in the previous example (95%), but the estimation error differs. In this case, the pattern of bias is non-linear.

The aim of this note is to 1) illustrate patterns of bias in the number of files estimated to have bats present using a linear estimator, and to 2) illustrate two approaches to bias adjustment, the Rogan-Gladen estimator (Rogan & Gladen 1978) and the Bayesian estimator (Joseph et al. 1995). The goal is to show the results of applying these methods to the output of acoustic filtering rather than to review or recommend specific noise-scrubbing software.

MATERIAL AND METHODS

I calculated the performance statistics of four noise-filtering applications described in the literature. These were the Bat Call Identification 10.0 filter and Britzke-Murray filter, both used in the Analook W (Clement et al. 2014), the Sonobat Batch Scrubber 5.1 and the Kaleidoscope 4.5.4 (Perea & Tena 2020). Those authors tested the software on bat detector recordings that had been visually vetted. A different recording library was used to test each pair of scrubbers; therefore, comparisons between scrubbers are valid only within pairs. The statistics shown here are examples for illustration purposes only, as accuracy statistics are not constant but may vary with sample size, sampling error, soundscape quality, software settings, software versions, and recording conditions (Clement et al. 2014, Whiting et al. 2022). I used published data to calculate the classification accuracy probabilities of each scrubber for bat files (sensitivity) and noise files (specificity) (Bramer 2020).

Sensitivity is the proportion of a sample of visually confirmed bat recordings classified by the software as bat-positive. Specificity is the proportion of a sample of visually confirmed noise recordings classified as bat-negative by the software. Overall accuracy was calculated as (true positives + true negatives) / total recordings.

Noise scrubber estimator

I then calculated the expected pattern of bias in the output of each noise scrubber resulting from misclassification. These results were plotted over a simulated range of true bat file prevalence (x-axis) from zero to 100%. I used the linear estimator $y = ax + b$, where y is the expected proportion of scrubber-labelled bat files in a sample of recordings, x is the true proportion of bat files, b is the false positive rate (1 – specificity), and a is sensitivity – b. The use of the linear estimator allows comparison of the expected output of a scrubber over a wide range of bat prevalence while holding constant the level of accuracy for that scrubber. Sensitivity and specificity were thus treated as fixed values for purposes of an idealized graph, following Marchevsky (1979). In reality, the sensitivity and specificity of a noise scrubber would each vary around a mean value.

Bias adjustment to the scrubber output

In the noise scrubber output, the apparent proportion of bat files may be significantly biased by misclassification, so applying a bias adjustment to the scrubber output is highly appropriate. One adjustment method is to calculate x, the true proportion of bat files in the sample, which is estimated as $x = (y - b)/a$. This solution to x is the Rogan-Gladen estimator (Rogan & Gladen 1978), a method that is appropriate so long as the sum of the sensitivity and specificity is not 1, and so long as it does not generate negative values or values exceeding 1. To illustrate bias adjustment, I plotted, in Fig. 2, the Rogan-Gladen adjustment of the outputs shown in Fig. 1. These bias-adjusted results are plotted over a simulated range of true bat file prevalence from zero to 100%.

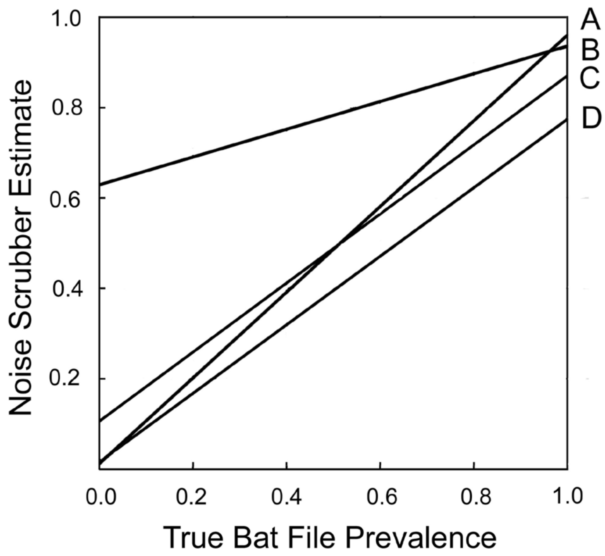

Fig. 1 – Bias of expected noise scrubbing estimates of the proportion of bat files in samples of recordings over a simulated true bat file prevalence range when the accuracy of each noise scrubber is held constant. A = BCID 10.0, B = Britzke & Murray, C = Sonobat Batch Scrubber 5.1, and D = Kaleidoscope 4.5.4.

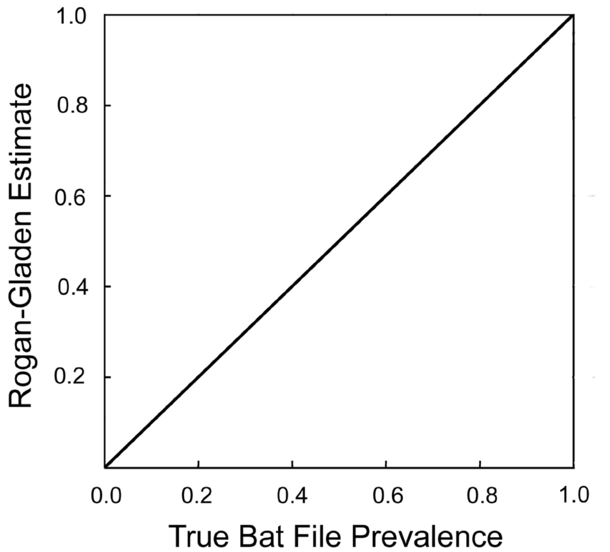

Fig. 2 - Rogan-Gladen bias adjustment, with overlapping estimates of the proportion of bat files in samples of recordings from the four noise scrubbers shown in Fig. 1. The graph is idealized as the accuracy of each scrubber is held constant.

If a Kaleidoscope noise scrubber sorts a sample of 10000 recordings into two groups, 2000 bat files and 8000 noise files, this scrubber-derived estimate of bat file prevalence (20%) can be adjusted as follows. Scrubber accuracy levels from Table 1 show a sensitivity of 0.774 and a specificity of 0.984. Thus, the point estimate would be 0.243 (24.3% were bat files).

$$Bat\ file\ prevalence = \frac{scrubber\ estimate - false\ positive\ rate}{true\ positive\ rate - false\ positive\ rate}$$

$$\frac{0.200 - 0.016}{0.774 - 0.016} = 0.243$$

Other examples of Rogan-Gladen point estimates of bat file prevalence are shown along with confidence intervals, calculated as $x \pm 1.96\sqrt{Var[x]}$, employing a modified Wald variance calculation as given by Rogan & Gladen (1978):

$$Var[x] = \frac{1}{(Se + Sp - 1)^2}\left[\frac{y(1-y)}{n} + \frac{Se(1-Se)x^2}{n_{Se}} + \frac{Sp(1-Sp)(1-x)^2}{n_{Sp}}\right]$$

Se = sensitivity, *Sp* = specificity, *n* = the number of acoustic files, and n_{Se} and n_{Sp} are the sample sizes that were used to validate sensitivity and specificity, respectively, and are shown in Table 1.

Table 1 also shows the mean absolute error (MAE) of each scrubber for simulated true bat file prevalence ranging from 0 to 100%. MAE is the mean of the absolute differences between true bat file prevalence (x) and the expected scrubber estimate (*y*) of bat file prevalence for each increment of bat file prevalence used in the simulation.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |\ y_i - x_i\ |$$

I also generated 188 Bayesian estimates of the proportion of bat files with 95% credible intervals using an Epitools Gibbs sampler (Sergeant 2018). The software employed in this analysis is hosted at https://epitools.ausvet.com.au/onetest. This tool is intended for estimating disease prevalence from the results of a diagnostic test when there are prior estimates of the sensitivity and specificity of that test. The tool can equally be applied to estimating bat file prevalence from scrubber results when prior estimates of scrubber sensitivity and specificity are available. The Gibbs sampler uses a Markov-chain Monte Carlo algorithm to generate posterior probability distributions for prevalence, sensitivity, and specificity. The default number of iterations is 25000, of which 5000 from the burn-in period are discarded to allow the model to converge on the true values. The output provides histograms of the posterior distributions for visual assessment of convergence. Epitools cites Joseph et al. (1995), who determined that as few as 100 – 200 cycles were sufficient for convergence. Initial alpha and beta parameters for sensitivity and specificity were derived from Table 1, and in the case of true prevalence, noninformative beta priors (1, 1) were used. I plotted the Bayesian adjustment of the four noise-scrubbing outputs shown in Fig. 1. These results are plotted over a simulated range of true bat file prevalence from zero to 100%.

RESULTS AND DISCUSSION

Bat acoustic surveys can generate numerous acoustic files within a short time span, and automated noise scrubbing can segregate acoustic files containing bat calls from those containing only noise. The bat file data may be used to infer trends in bat activity. The “scrubber estimate” of bat file prevalence may include false positive and false negative classification, and it is necessary to test software classification accuracy intermittently for a given field site with large samples of recordings of bats and noise. This is done by comparing software output to the visual classification of acoustic spectrograms, which is regarded as true bat and noise file numbers. Table 1 shows data from the literature for the classification performance of four noise scrubbers (Clement et al. 2014, Perea & Tena 2020). Also shown in Table 1, values are calculated for true positive (sensitivity), true negative (specificity), and false positive classification probabilities. Misclassification creates a bias in the scrubber estimate of bat file prevalence within a field sample of recordings, and this bias may be non-linear among estimates from the same noise scrubber, requiring unique correction coefficients. The scrubber classification accuracy values used here are only for illustration purposes and may differ for later versions of the same software applications when tested on the same recordings.

The expected pattern of bias among the four noise scrubbers is shown in Fig. 1 over a simulated range of bat prevalence from zero to 100%. The *y*-axis shows expected estimates of bat file prevalence within field samples of recordings while holding constant the measured accuracy of each scrubber (Table 1). The two pairs of scrubbers were accuracy-tested using different datasets. Therefore, comparisons are valid only within pairs. The BCID and Britzke-Murray filters show low estimation bias at high bat file prevalence, resulting from their high sensitivity (high true positive detection rate). The BCID also has low estimation bias throughout the range of bat file prevalence, as it has a high classification accuracy of noise files (high specificity), resulting in few false positives. In comparison, the Britzke-Murray has a high bias at low bat prevalence, where noise files are abundant, because it has a high false positive rate. The Britzke-Murray false positive rate is 63%, meaning that if there were zero bat files in a sample of recordings, the software estimate would imply that approximately 63% of the files were bat files. The accuracy of the Sonobat and Kaleidoscope scrubbers differs by only 2%, but the Sonobat has a significantly lower mean absolute error (Table 1). Their prevalence estimation plots are parallel, but they intersect true prevalence at different points. The Kaleidoscope has a high specificity and low sensitivity, so the percentage bias across the range of bat file prevalence is close to linear. The Sonobat’s lower specificity magnifies the bias at low bat file prevalence.

Examples of bias adjustment of scrubber point estimates are shown in Table 1. For example, if a Sonobat scrubber sorted the acoustic files in a field sample such that 20% were labelled as bat files, this output can be bias-adjusted, giving a Rogan-Gladen estimate of the true proportion of bat files of 12.3%. This adjustment of the scrubber bat file prevalence

Table 1 – Classification accuracy, estimation error, and examples of bias adjustment of four noise scrubbers. Accuracy data are from Clement et al. (2014) and Perea & Tena (2020).

	Scrubber name			
	BCID[a]	BM[b]	Sonobat[c]	Kaleidoscope[d]
Classification accuracy:				
Bat files tested	1556	1556	2828	2828
True positives	1493	1456	2459	2190
Sensitivity	0.960	0.936	0.870	0.774
Noise files tested	13801	13801	4685	4685
True negatives	13629	5124	4187	4611
Specificity	0.988	0.371	0.894	0.984
False positive rate	0.012	0.629	0.106	0.016
Overall accuracy	0.985	0.428	0.885	0.905
Mean absolute estimation error:				
Scrubber	0.017	0.289	0.060	0.106
Bayesian	<0.001	0.002	0.001	<0.001
Rogan-Gladen	0.000	0.000	0.000	0.000
Bias adjustment examples where n = 10000:				
Scrubber estimate	0.700	0.700	0.200	0.200
95 % CI	0.691 – 0.709	0.691 – 0.709	0.192 – 0.208	0.192 – 0.208
Rogan-Gladen estimate	0.726	0.231	0.123	0.243
95% CI	0.714 – 0.738	0.195 – 0.268	0.111 – 0.135	0.230 – 0.255
Bayesian estimate	0.726	0.231	0.123	0.243
95% CrI	0.714 – 0.738	0.191 – 0.269	0.108 – 0.138	0.231 – 0.255

a = Bat Call Identification 10.0, b = Britzke-Murray, c = Sonobat Batch Scrubber 5.1, d = Kaleidoscope 4.5.4

estimate of 0.200 was calculated from values in Table 1 as

$$\frac{0.200 - 0.106}{0.870 - 0.106} = 0.123$$

Rogan-Gladen adjusted scrubber results agreed with simulated true bat file prevalence values over a range of zero to 100% for all four scrubbers (Fig. 2). However, this estimator does have drawbacks. It cannot be used if the sensitivity and specificity total 100%. Also, it can result in a value that is negative or that exceeds unity. Flor et al. (2020) found this in 4% of 999979 simulated data sets. They suggest that if such cases are encountered in the field, bias adjustment (via truncation) should not be attempted, and that classification accuracy (sensitivity and specificity) should be re-evaluated.

Bayesian adjustment of expected scrubber estimates of bat file prevalence, when compared over a simulated true bat prevalence range of zero to 100%, showed very low mean absolute errors (Table 1 and Fig. 3), largely replicating the Rogan-Gladen results. However, at very low bat prevalence, there is divergence and inflation among Bayesian estimates (Fig. 3). It may be prudent to avoid using this tool when very low bat prevalence is expected in bat surveys. Flor et al. (2020) tested both the Rogan-Gladen estimator and Bayesian inference using 999979 simulated data sets. They

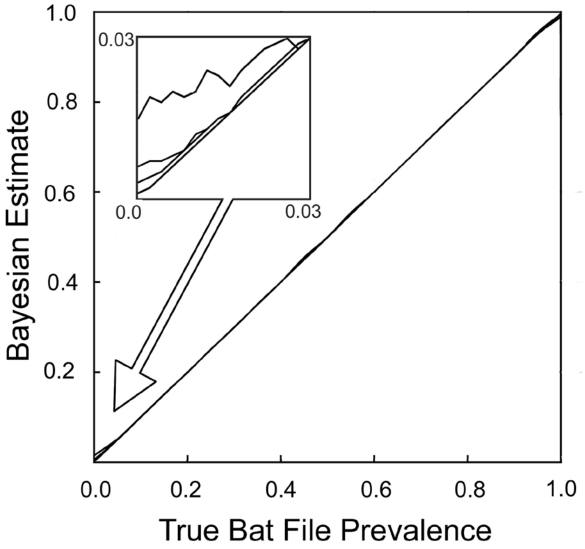

Fig. 3 – Bayesian bias adjustment of the proportion of bat files in samples of recordings from the four noise scrubbers shown in Fig. 1. The inset shows the variation among Bayesian estimates at very low bat file prevalence. The scrubbers depicted in the inset, from top to bottom, are Britzke & Murray, Sonobat Batch Scrubber 5.1, Kaleidoscope 4.5.4, and BCID 10.0.

found that Bayesian credible intervals overall gave better coverage and greater length than Rogan-Gladen confidence intervals.

The use of these bias-adjustment tools requires consideration of various caveats and assumptions. Prior to routine bias adjustment of noise scrubber output from field sampling, it is necessary to calibrate the classification accuracy of the software under prevailing field conditions with adequate sample sizes. Clement et al. (2014) emphasize that the sample of acoustic files used to assess the classification accuracy of any software visually must be representative of the target population. Accuracy data are situational, and classification accuracy may differ among acoustic surveys recorded under dissimilar conditions. Field sites may vary in noise sources and risk of generating false positive classifications. Some sites may include electronic noise, for example. If the bat call recognition threshold of a given software application is adjusted or changes with software versions (Goodwin & Gillam 2021), classification accuracy and bias adjustment may change and should be recalibrated. Raising the bat call recognition threshold may exclude weak and fragmented calls from distant animals (Knight & Bayne 2019), but if those are abundant, false negative classification probability may be high. If sampling bats in an open meadow, a noise scrubber with a fixed recognition threshold may show lower sensitivity than when sampling at a small pond, where bats are concentrated closer to the microphone and call quality is high.

If noise scrubbers are used in analyzing bat acoustic survey data, a conservative approach would visually validate accuracy of the software at each sampling station intermittently, consistently using the same recording equipment, same bat detector settings, same spectrogram viewer and settings, and the same visual classification criteria. It would not be appropriate to adopt values such as those in Table 1 as universally applicable.

CONCLUSIONS

If bat activity levels and population trends are to be inferred from comparative numbers of bat acoustic recordings, noise-scrubbing software applications may enable automated estimation of the proportion of bat files in field samples of recordings. However, misclassification by the software will result in biased estimates, and the bias may be non-linear. This may impact bat conservation decisions. If the software is tested for classification accuracy, those results can be incorporated into bias-adjustment tools. This report illustrates bias adjustment using the Rogan-Gladen estimator and Bayesian inference when applied to the simulated output of four bat noise scrubbers described in the literature, although both tools have some limitations.

ACKNOWLEDGEMENTS

I wish to thank Dr. Elena Tena López for sharing of data, and I thank two anonymous reviewers for help in greatly improving the manuscript. The Salt Spring Island Community Bat Program is funded by the Habitat Stewardship Program of the Government of Canada, the Habitat Conservation Trust Foundation, the Forest Enhancement Society of British Columbia, and public donations to the British Columbia Conservation Foundation.

REFERENCES

BRAMER, M. (2020). Principles of data mining. ed.: Springer-Verlag. London, UK, 571 pp. https://doi.org/10.1007/978-1-4471-4884-5

BROOKS, R. T. (2011). Declines in summer bat activity in central New England 4 years following the initial detection of white-nose syndrome. *Biodivers Conserv*, 20: 2537-2541. https://doi.org/10.1007/s10531-011-9996-0

CLEMENT, M. J., MURRAY, K. L., SOLICK, D. I. & GRUVER, J. C. (2014). The effect of call libraries and acoustic filters on the identification of bat echolocation. *Ecol Evol*, 4(17): 3482-3493. https://doi.org/10.1002/ece3.1201

DE JONG, J., LARA, M., HÅSTAD, O. & VICTORSSON, J. (2021). Activity pattern and correlation between bat and insect abundance at wind turbines in south Sweden. *Animals*, 11(11): 3269. https://doi.org/10.3390/ani11113269

DZAL, Y., MCGUIRE, L. P., VESELKA, N. & FENTON, M. B. (2010). Going, going, gone: the impact of white-nose syndrome on the summer activity of the little brown bat (*Myotis lucifugus*). *Biol Letters*, 7(3): 392-394. http://doi.org/10.1098/rsbl.2010.0859

FLOR, M., WEIß, M., SELHORST, T., MÜLLER-GRAF, C. & GREINER, M. (2020). Comparison of Bayesian and frequentist methods for prevalence estimation under misclassification. *BMC Public Health*, 20: 1135. https://doi.org/10.1186/s12889-020-09177-4

FRICK, W. F., KINGSTON, T. & FLANDERS, J. (2019). A review of the major threats and challenges to global bat conservation. *Ann Ny Acad Sci*, 1469(1): 5-25. https://doi.org/10.1111/nyas.14045

GOODWIN, K. R. & GILLAM, E. H. (2021). Testing accuracy and agreement among multiple versions of automated bat call classification software. *Wildlife Soc B*, 45(4): 690-705. https://doi.org/10.1002/wsb.1235

HARRISON, K. P. (2021). Bat use of upland ponds within a hardwood forest ecosystem, Southern Indiana. Ball State University, Munice, Indiana. Ball State University.

JOSEPH, L., GYORKOS, T. W. & COUPAL, L. (1995). Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *Am J Epidemiol*, 141(3): 263-272. https://doi.org/10.1093/oxfordjournals.aje.a117428

KNIGHT, E. C. & BAYNE, E. M. (2019). Classification threshold and training data affect the quality and utility of focal species data processed with automated audio-recognition software. *Bioacoustics*, 28(6): 539-554. https://doi.org/10.1080/09524622.2018.1503971

LOEB, S. C., RODHOUSE, T. J., ELLISON, L. E., LAUSEN, C. L., REICHARD, J. D., IRVINE, K. M., INGERSOLL, T. E., COLEMAN, J. T. H., THOGMARTIN, W. E., SAUER, J. R., et al. (2015). A plan for the North American Bat monitoring program (NABat). ed.: United States Department of Agriculture. Asheville, North Carolina, USA, 100 pp.

MARCHEVSKY, N. (1979). Re:"estimating prevalence from the results of a screening test". *Am J Epidemiol*, 109(6): 720-721. https://doi.org/10.1093/oxfordjournals.aje.a112736

NOCERA, T., FORD, W. M., SILVIS, A. & DOBONY, C. A. (2019). Patterns of acoustical activity of bats prior to and 10 years after WNS on Fort Drum Army Installation, New York. *GECCO*, 18: e00633. https://doi.org/10.1016/j.gecco.2019.e00633

PEREA, S. & TENA, E. (2020). Different bat detectors and processing software... Same results? *JBRC*, 13(1): 4-8. https://doi.org/10.14709/BarbJ.13.1.2020.01

PHINNEY, L. (2020). Long-term decline in bat activity using passive acoustic monitoring and an equipment correction factor in Nova Scotia, Canada. University of Waterloo, Waterloo, Ontario, Canada. University of Waterloo.

PINZARI, C., PECK, R., ZINN, T., GROSS, D., MONTOYA-AIONA, K., BRINCK, K., GORRESEN, M. & BONACCORSO, F. (2019). Hawaiian hoary bat (*Lasiurus cinereus semotus*) activity, diet and prey availability at the Waihou Mitigation Area, Maui. ed.: Hawai`i Cooperative Studies Unit. Hilo, Hawai`I, 60 pp.

ROCHE, N., LANGTON, S., AUGHNEY, T., RUSS, J. M., MARNELL, F., LYNN, D. & CATTO, C. (2011). A car-based monitoring method reveals new information on bat populations and distributions in Ireland. *Anim Conserv*, 14(6): 642-651. https://doi.org/10.1111/j.1469-1795.2011.00470.x

ROGAN, W. J. & GLADEN, B. (1978). Estimating prevalence from the results of a screening test. *Am J Epidemiol*, 107(1): 71-76. https://doi.org/10.1093/oxfordjournals.aje.a112510

SERGEANT, E. S. G. (2018) Epitools Epidemiological Calculators. Ausvet.

WHITING, J. C., DOERING, B. & AHO, K. (2022). Can acoustic recordings of cave-exiting bats in winter estimate bat abundance in hibernacula?. *Ecol Indic*, 137: 108755. https://doi.org/10.1016/j.ecolind.2022.108755