18 September 2009

Mike Mitchell
Montana Cooperative Wildlife Research Unit
205 Natural Sciences Building
University of Montana
Missoula, MT 59812
Ph: (406) 243-4390
Email: mike.mitchell@umontana.edu

Temporal validation of an estimator for successful breeding pairs of wolves in the U.S. northern Rocky Mountains

RH: Test of Breeding Pair Estimator

MICHAEL S. MITCHELL, U. S. Geological Survey, Montana Cooperative Wildlife Research Unit, 205 Natural Sciences Building, University of Montana, Missoula, MT 59812, USA

JUSTIN A. GUDE, Montana Fish, Wildlife & Parks, 1420 E. 6th Ave., Helena, MT 59620, USA

DAVID E. AUSBAND, Montana Cooperative Wildlife Research Unit, 205 Natural Sciences Building, University of Montana, Missoula, MT 59812, USA

CAROLYN A. SIME, Montana Fish, Wildlife & Parks, 1420 E. 6th Ave., Helena, MT 59620, USA

EDWARD E. BANGS, U.S. Fish and Wildlife Service, 585 Shepard Way, Helena, MT 59601, USA

MICHAEL D. JIMENEZ, U.S. Fish and Wildlife Service, P.O. Box 2645, Jackson, Wyoming, 83001, USA

CURT M. MACK, Gray Wolf Recovery Project, Nez Perce Tribe, P.O. Box 1922 McCall, ID 83638, USA

THOMAS J. MEIER, National Park Service, Denali National Park and Preserve, Denali Park, AK 99755

M. STEVEN NADEAU, Idaho Department of Fish and Game, 600 S. Walnut, Boise, ID 83707, USA

DOUGLAS W. SMITH, Yellowstone Center for Resources, P.O. Box 168, Yellowstone National Park, WY 82190, USA

ABSTRACT Model-based predictors derived from historical data are rarely evaluated before they are used to draw inferences. We performed a temporal validation, (i.e., assessed the performance of a predictive model using data collected from the same population after the model was developed) of a statistical predictor for the number of successful breeding pairs of wolves (*Canis lupus*) in the northern Rocky Mountains (NRM). We predicted the number of successful breeding pairs, β, in Idaho, Montana, and Wyoming based on the distribution of pack sizes observed through monitoring in 2006 and 2007 ($\hat{\beta}$), and compared these estimates to the minimum number of successful breeding pairs, β_{MIN}, observed through intensive monitoring. β_{MIN} was consistently included within the 95% confidence intervals of $\hat{\beta}$ for all states in both years (except Idaho in 2007), generally following the pattern $\hat{\beta}_L$ (lower 95% prediction interval for $\hat{\beta}$) $< \beta_{\mathrm{MIN}} < \hat{\beta}$. This evaluation of $\hat{\beta}$ estimates for 2006 and 2007 suggest it will be a robust model-based method for predicting successful breeding pairs of NRM wolves in the future, provided influences other than those modeled in $\hat{\beta}$ (e.g., disease outbreak, severe winter) do not have a strong effect on wolf populations. Managers can use $\hat{\beta}$ models with added confidence as part of their post-delisting monitoring of wolves in NRM.

KEY WORDS estimator, temporal validation, gray wolf, Idaho, northern Rocky Mountains, Montana, successful breeding pair, Wyoming

INTRODUCTION

Gray wolves in the NRM were classified under the Endangered Species Act through 2008 as either endangered in the Northwest Montana Recovery Area (NWMT) where wolves recolonized naturally beginning in 1979, or experimental, non-essential in the Central Idaho and Greater Yellowstone Experimental Population Areas (CIEPA and GYEPA respectively) where wolves were reintroduced in 1995 and 1996. Throughout recovery of the NRM wolf population, monitoring has been conducted to evaluate progress toward 2 recovery goals: 1) 300 wolves, and 2) 30 successful breeding pairs, defined by the U.S. Fish and Wildlife Service as packs containing at least 1 adult male and 1 adult female with ≥ 2 pups on 31 December of a given year (USFWS 1994). Monitoring to document progress toward recovery has been intensive, resulting in near-census quality data on wolf abundance and number of breeding pairs. Because the wolf population has exceeded recovery goals since 2002, the USFWS has delisted wolves in the NRM (USFWS 2009). Following delisting, federal funds for intensive monitoring will no longer be available, but Idaho, Montana, and Wyoming will be required to ensure numbers of wolves and successful breeding pairs remain above recovery criteria. Cost-effective and accurate alternatives to intensive monitoring are thus needed to ensure the wolf population in the NRM remains recovered after delisting.

To assist with post-delisting monitoring, Mitchell et al. (2008) used monitoring data through 2005 to develop a statistical model for predicting the number of successful breeding pairs, β, based on the distribution of pack sizes within a wolf population. They showed how demographic trends and human-caused mortality affected these predictions differently for 6 analysis areas within the NRM (Idaho, ID; NWMT, Southwest Montana adjacent to CIEPA, SWMT-CIEPA; Southwest Montana adjacent to GYEPA, SWMT-GYEPA; Wyoming, WY; Yellowstone National Park, YNP; Mitchell et al. 2008); they concluded models appropriate to the demography and human-caused mortality experienced by a wolf population must be used to derive predictions of the total number of successful breeding pairs in each population.

Our objective in this study was to temporally validate how well $\hat{\beta}$ could predict the number of successful breeding pairs in the NRM for years subsequent to those used to construct the predictive model. Temporal validation evaluates the ability of a statistical model to predict future conditions for the population from which the model was derived, subsequent to the observations used to generate the model (Altman and Royston 2000; also referred to as "historical transportability" in Justice et al. 1999). If a model performs well in a temporal validation, this lends support to the robustness of the model. We thus used $\hat{\beta}$ to predict the number of breeding pairs for Idaho, Montana, and Wyoming in 2006 and 2007 and compared these estimates to the minimum number of successful breeding pairs, β_{MIN}, documented for each year through intensive monitoring (data unavailable at the time analyses contained in Mitchell et al. 2008 were conducted; UFWS et al. 2007, 2008). The purpose of our effort was to provide another evaluation of whether $\hat{\beta}$ would be a reliable alternative to intensive monitoring for documenting successful breeding pairs following delisting.

STUDY AREA

The NWMT, CIEPA, and GYEPA federal recovery area boundaries overlap the states of Montana, Idaho, and Wyoming (Mitchell et al. 2008). Wolf populations within each recovery area experienced different levels of isolation, protection, management, and exposure to humans, based largely on geography and administrative boundaries (Mitchell et al. 2008, USFWS 2009). Much of ID was federally-designated wilderness; surrounding forested lands were a mix of public and private timber lands. Wolves in Idaho were managed as a non-essential, experimental population (i.e., receiving a lower level of protection under the Endangered Species Act, thus increasing management flexibility; USFWS 1994); the majority of mortality was due to removal in response to wolf-livestock conflicts and to poaching. Lands in NWMT were primarily public or corporate-owned and managed for timber production. Wolves in NWMT were managed as an endangered population; poaching and vehicle collisions exceeded legal removals. Land ownership in SWMT-CIEPA, SWMT-GYEPA, and WY was a mixture of public and private; local land management emphasized livestock production. Wolves were managed as a non-essential, experimental population in these areas and removal following livestock conflicts was the primary cause of mortality. YNP wolves were managed as a non-essential, experimental population, but lands within YNP are protected and relatively undeveloped; human-caused mortality was low compared to deaths caused by intraspecific conflicts (Mitchell et al. 2008).

METHODS

Wolf Monitoring

In 2006 and 2007, Idaho, Montana, and Wyoming relied primarily on federal funding to monitor radiocollared packs on the ground and from aircraft at routine intervals throughout the calendar year, at levels of intensity consistent with previous years. On average 30% of the adult-sized wolves in the population were monitored by radio telemetry 2-4 times a month. Some uncollared packs were monitored by ground tracking. Breeding success was documented through observations of pups present in a pack during aerial and ground observations of dens in spring (Montana and Wyoming), visitation of den and rendezvous sites (Idaho) and monitoring of pack composition during fall months (all states; Mitchell et al. 2008). At the end of each calendar year, all available information was used to assess pack size and whether each pack satisfied the successful breeding pair criterion set by USFWS (USFWS et al. 2007, 2008).

Prediction of Successful Breeding Pairs

Because some of the same packs were observed over multiple years, we assessed lack of independence in our pack size and breeding pair data, blocked by individual packs, from 1979 through 2007 by calculating extra binomial variation (i.e. the dispersion parameter) for our data set; a ratio >1 can indicate a lack of independence among observations (SAS Institute 2000). Strict independence of the data collected and presented in this analysis is not requisite for the temporal validation we conducted. In reality, packs survive in this population for multiple years, and thus $\hat{\beta}$ needs to perform accurately given this fact. To predict the number of successful breeding pairs in each state for 2006 and 2007, we first assigned each wolf pack observed in each state in 2006 and 2007 to 1 of the 6 analysis areas defined by Mitchell et al. (2008). For each pack, we used the $\hat{\beta}$ model specific to the analysis area to which it was assigned to predict the probability it contained a breeding pair with lower and upper 95% confidence limits. We summed these probabilities and their confidence limits across packs within each state in each year

(Mitchell et al. 2008) to predict the number of successful breeding pairs ($\hat{\beta}$, with 95% prediction interval, $\hat{\beta}_L$ and $\hat{\beta}_U$) present in Idaho, Montana, and Wyoming in 2006 and 2007. To make these predictions, we used the predictors presented in Mitchell et al. (2008):

$$\hat{\beta} = \sum_{i=4}^{k} (n_i * \hat{P}_i),$$

$$\hat{\beta}_L = \sum_{i=4}^{k} (n_i * \hat{P}_{iL}),$$

$$\hat{\beta}_U = \sum_{i=4}^{k} (n_i * \hat{P}_{iU}),$$

where the summation is over pack sizes i = 4 to k of the number of known packs of size i, n_i, $\hat{P}_i$ is the predicted probability a pack of size i containing a successful breeding pair from the logistic models developed in Mitchell et al. (2008) independently for each of 6 analysis areas, and $\hat{P}_{iL}$ and $\hat{P}_{iU}$ are the back-transformed lower and upper confidence bounds on $\hat{P}_i$ (Neter et al. 1996: 603-604). In 2007, data on pack size were missing for 10 packs in Idaho, 3 packs in Montana (1 from NWMT, 2 from SWMT-GYEPA), and 5 packs in Wyoming (all from outside YNP), comprising 9% of total packs for that year. For these packs, we substituted average pack size, rounded to the nearest integer, calculated for each state: Idaho = 6.45 (3.43 SD), Montana = 5.73 (2.91 SD), Wyoming = 10.19 (5.02 SD). We assumed mean pack size would be accurate estimates of expected size for those packs.

We summed the number of successful breeding pairs observed through monitoring within each state and each year to represent the minimum number known of successful breeding pairs, β_{MIN}, present in Idaho, Montana, and Wyoming in 2006 and 2007. To assess accuracy of $\hat{\beta}$, we compared the predicted number of successful breeding pairs, $\hat{\beta}$ with upper and lower 95% prediction intervals ($\hat{\beta}_L$ and $\hat{\beta}_U$, respectively) in each state to the minimum number known for each year, β_{MIN}. The models presented by Mitchell et al. (2008) used pack size to predict the number of breeding pairs observed through monitoring, i.e., β_{MIN}; for our temporal validation we therefore concluded qualitatively $\hat{\beta}$ was accurate if the prediction intervals for β contained β_{MIN}. Because of the rapid growth of the wolf population in the northern Rockies in recent years, we deemed it likely that monitoring efforts in 2006 and 2007 would fail to detect all breeding pairs. Further, our inclusion of 18 packs of unknown size, and therefore unknown breeding pair status, in our estimation of $\hat{\beta}$ for 2007 meant $\hat{\beta}$ could exceed β_{MIN} even if monitoring detected breeding pairs perfectly among packs of known size. We therefore expected $\hat{\beta}$ to be slightly greater than β_{MIN} for both years.

RESULTS

In 2006, 134 packs comprising 972 wolves and $\beta_{\mathrm{MIN}} = 86$ successful breeding pairs were monitored (USFWS et al. 2007). In 2007, 192 packs comprising 1,192 wolves and $\beta_{\mathrm{MIN}} = 107$ successful breeding pairs were monitored (USFWS et al. 2008). The ratio of deviance to degrees of freedom for our pack size and breeding pair data, blocked by individual packs, from 1979 through 2007 was 0.94, therefore data from 2006 and 2007 were generally independent of data through 2005 used by Mitchell et al. (2008) to build their models. β_{MIN} was included in 95% prediction intervals of β for Idaho, Montana, and

Wyoming for both years, except for Idaho in 2007 (Table 1). Mean range for 95% prediction intervals of β across all states in both years was 12.62 (SD=5.19). $\hat{\beta}$ slightly exceeded β_{MIN} for all states in both years, except for Montana and Wyoming in 2007 when they were approximately equal (Table 1).

DISCUSSION

Model-based predictors derived statistically from historical data are rarely evaluated before being used to predict parameters from future data (Harrell et al. 1996; Justice et al. 1999; Altman and Royston 2000). Goodness of fit of a statistical model to the data used to build it is no guarantee that the model will predict future population parameters accurately; variation in processes and contributors to uncertainty between past and future circumstances can result in model-based predictions that vary widely from reality. Consequences for such error can be significant when predictions are used to assess population status for species of particular biological or regulatory importance.

As part of delisting of wolves in NRM (USFWS 2009), Idaho, Montana, and Wyoming will be required to monitor the number of successful breeding pairs into the future, but likely without the federal funding that supported intensive monitoring prior to delisting. Mitchell et al. (2008) presented models for estimating the number of successful breeding pairs of wolves, β, based on observed pack sizes for 6 analysis areas within the NRM. Their results suggested pack size explained much of the variation in the probability that a pack contained a successful breeding pair within the NRM, with models varying across the analysis areas due to differences in growth rate of wolf populations and human-caused mortality.

We conducted a temporal validation of a model-based β estimator ($\hat{\beta}$) by comparing the number of successful breeding pairs it estimated for the wolf populations of Idaho, Montana, and Wyoming to the minimum number of breeding pairs known through monitoring in 2006 and 2007. We used a model-based β estimator specific to the 6 analysis areas developed by Mitchell et al. (2008) using data that were collected during and prior to 2005. Prediction intervals for $\hat{\beta}$ contained β_{MIN} values for all states in both years, except for Idaho in 2007. As we expected, $\hat{\beta}$ generally exceeded β_{MIN}, except for Wyoming and Montana in 2007, where, and Idaho in 2007 where $\hat{\beta} < \beta_{\mathrm{MIN}}$. We hypothesize this difference between $\hat{\beta}$ and β_{MIN} represents successful breeding pairs unobserved through monitoring because the NRM wolf population continued to grow rapidly whereas monitoring efforts remained relatively constant (USFWS et al. 2007, 2008). We cannot be certain why the confidence interval for $\hat{\beta}$ did not contain β_{MIN} for Idaho in 2007; potentially, this could have been due to under-counting of successful breeding pairs during monitoring, or the true size of the 10 packs for which we used average pack size to estimate successful breeding pair status could have been smaller than the average. Alternatively, if β_{MIN} was in reality close to the true parameter we were trying to predict (β), this discrepancy could simply reflect an aberrant year in the process that generated β, assuming that pack size remained closely related to the probability that a wolf pack contained a successfully breeding pair. The general pattern of $\hat{\beta}_L < \beta_{\mathrm{MIN}} < \hat{\beta}$ across all states in both years (except Idaho in 2007) suggests that $\hat{\beta}$ is a robust predictor for the NRM, accounting for successful breeding pairs present but unobserved through monitoring in 2006 and 2007.

Our use of mean pack size to impute missing data for 18 packs among the 3 states assumed mean pack size was an accurate estimate of expected pack size for those packs. We did not assess how a violation of this assumption would affect our predictions of β. Further, we did not incorporate variability associated with mean pack sizes into our bounds, $\hat{\beta}_L$ and $\hat{\beta}_U$, so both represent underestimates of our uncertainty. For our analyses, we assumed using mean pack size to impute missing data would have negligible effects on $\hat{\beta}$, $\hat{\beta}_L$, and $\hat{\beta}_U$ given the small proportion of packs (9% of packs observed in 2007) for which pack size was unknown, and would result in more accurate predictions than if such packs were excluded from analysis. This assumption is likely to become increasingly questionable in future applications of our predictor, however, because reduced monitoring efforts after delisting will result in larger proportions of packs for which size is unknown. Further development of our predictor to address this problem will require imputation of missing data on pack size and inclusion of associated uncertainties into estimated prediction intervals.

MANAGEMENT IMPLICATIONS

Our results further support the findings of Mitchell et al. (2008) that suggested $\hat{\beta}$ provides an accurate predictor of the number of successful breeding pairs of wolves in the NRM, robust to variation in factors shown to historically influence the relationship between size of a pack and the probability it contains a successful breeding pair. The area-specific nature of $\hat{\beta}$, reflecting different rates in human-caused mortality and population growth across the 6 analysis areas, will allow managers to choose models of $\hat{\beta}$ appropriate to circumstances that could change following delisting (e.g., increased human-caused mortality in NWMT could make the SWMT-CIEPA model most appropriate for packs in NWMT; Mitchell et al. 2008). Provided human-caused mortality or population growth rates do not exceed the range of values encompassed across these models, managers can rely on $\hat{\beta}$ predictions to reliably demonstrate recovery criteria are met following delisting of NRM wolves. In the event circumstances for NRM wolves differ substantially from those influencing data used by Mitchell et al. (2008) to generate their model-based $\hat{\beta}$ predictor (e.g., disease outbreak, severe winter), a modified $\hat{\beta}$ predictor will need to be developed and tested to ensure the models remain robust to the new conditions. Whether future circumstances for NRM wolves are known to change appreciably or not, we recommend periodic evaluation (e.g., every 5 years) of model robustness by comparing predictions of $\hat{\beta}$ to number of breeding pairs observed in intensively monitored sub-populations within the NRM. While packs of unknown size comprise a small proportion of the observations for NRM, we suggest using mean pack size as an expected value for packs of unknown size is likely to provide a more accurate predicition of β than would excluding such packs, provided they are relatively few. As packs of unknown size comprise an increasing proportion of the observations for NRM, rigorous means of imputing unknown pack sizes will be required to ensure that $\hat{\beta}$, $\hat{\beta}_L$, and $\hat{\beta}_U$ remain reliable.

ACKNOWLEDGMENTS

We are very grateful for all the contributions made to wolf monitoring efforts by hundreds of people since 1979. Employees and volunteers affiliated with the U.S. Fish and Wildlife Service, the National Park Service, U.S. Forest Service, the U.S. Bureau of Land Management, USDA Wildlife Services, academic institutions, Nez Perce Tribe, Blackfeet Nation, Confederated Salish and Kootenai Tribe, Turner Endangered Species Fund, Montana Fish, Wildlife & Parks, Idaho Department of Fish and Game, and Wyoming Game and Fish contributed greatly to the data set we used for our analyses. We are also indebted to the citizens and private landowners of the Northern Rockies who reported wolves or wolf sign to agency personnel, which was often the first step in verifying a new wolf pack. We also thank our pilots for their interest, dedication, and years of safe flying. We thank Associate Editor Olivier Gimenez and two anonymous referees for their thoughtful comments that improved this manuscript.

LITERATURE CITED

Altman, D. G., and P. Royston. 2000. What do we mean by validating a prognostic model? Statistics in Medicine 19: 453-473.

Harrell, F. E., Jr., K. L. Lee, and D. B. Mark. 1996. Tutorial in biostatistics; multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in Medicine 15:361-387.

Justice, A. C., K. E. Covinsky, and J. A. Berlin. 1999. Assessing the generalizability of prognostic information. Annals of Internal Medicine 130:515-524.

Mitchell, M. S., D. E. Ausband, C. A. Sime, E. E. Bangs, J. A. Gude, M. D. Jimenez, C. M. Mack, T. J. Meier, M. S. Nadeau, and D. W. Smith. 2008. Estimation of successful breeding pairs for wolves in the U.S. northern Rocky Mountains. Journal of Wildlife Management 72:881-891.

Neter, J., M. H. Kutner, C. J. Nachtscheim, and W. Wasserman. 1996. Applied linear statistical models, fourth edition. W. C. B. McGraw Hill, Boston, MA, USA.

SAS Institute, Inc. 2000. SAS/STAT User's Guide, version 8. SAS Institute Inc., Cary, NC, USA.

U.S. Fish and Wildlife Service. 1994. The reintroduction of gray wolves to Yellowstone National Park and central Idaho. Final Environmental Impact Statement. Denver, Colorado. Appendix 9.

U.S. Fish and Wildlife Service. 2009. Endangered and threatened wildlife and plants; final rule to identify the northern Rocky Mountain population of gray wolf as a distinct population segment and to revise the list of endangered and threatened wildlife. Federal Register 74(62):15123-15188.

U.S. Fish and Wildlife Service, Nez Perce Tribe, National Park Service, and USDA Wildlife Services. 2007. Rocky Mountain Wolf Recovery 2006 Annual Report. USFWS, Ecological Services, 100 N. Park, Suite 320, Helena, Montana, USA.

U.S. Fish and Wildlife Service, Nez Perce Tribe, National Park Service, and USDA Wildlife Services. 2008. Rocky Mountain Wolf Recovery 2007 Annual Report. USFWS, Ecological Services, 100 N. Park, Suite 320, Helena, Montana, USA.

Table 1. Comparison of estimated number of successful breeding pairs of wolves, $\hat{\beta}$, to minimum known number of successful breeding pairs (β_{MIN}) for Idaho, Montana, and Wyoming, 2006 and 2007.

State	Year	# packs	$\bar{x}$ pack size (SD)	β_{MIN}	$\hat{\beta}$	$\hat{\beta}_L$	$\hat{\beta}_U$
Idaho	2006	50	7.66 (2.76)	41	42.07	37.37	44.96
	2007	83	6.51 (3.44)	43	57.31	50.83	61.86
Montana	2006	45	6.29 (2.49)	21	28.19	18.89	35.29
	2007	73	5.73 (2.91)	39	37.88	25.09	46.06
Wyoming	2006	39	7.85 (3.74)	25	28.32	21.55	33.36
	2007	36	10.19 (5.02)	24	28.83	23.55	31.47