

## Review article

### Estimating wolf abundance with unverified methods

Adrian Treves †, Nelson Institute for Environmental Studies, University of Wisconsin-Madison  
Francisco J. Santiago-Ávila †, Project Coyote, The Rewilding Institute

† equal co-authors

Correspondence: AT [atreves@wisc.edu](mailto:atreves@wisc.edu)

#### Abstract

Wildlife abundance can be very difficult to estimate, especially for rare and elusive species, such as wolves. Over nearly a century, wolf scientists have developed methods for estimating abundance across large areas, which involve marked animals being detected again after capture, sometimes supplemented by observations of the associates of those marked animals. Recently, several USA states have departed from those repeatedly validated methods to explore alternatives that are believed to be less expensive for wolf populations estimated >1000 individuals. The new methods sacrifice precision but are believed to retain adequate accuracy and sensitivity to changing conditions for reliable decision-making. We review evidence for the accuracy, precision, sensitivity, and reproducibility of the new ‘scaled occupancy model’ (SOM) applied in Wisconsin. We conclude that the Wisconsin method would systematically overestimate wolf abundance by large (but currently incalculable) margins. Because Wisconsin, similar to other states, not only changed to unverified methods but also implemented widespread wolf killing, shortcomings in their estimates of wolf abundance may have far-reaching consequences for population viability and confidence in state wildlife policy. We discuss findings from Wisconsin alongside similar findings for other jurisdictions’ occupancy models being insensitive to changing conditions. Overall, Wisconsin’s method for estimating wolf abundance shows significant departures from best practices in scientific measurement. Verification will require independent replication and unbiased tests at multiple scales in multiple habitats under different human-induced mortality rates, and rigorous independent review before the new methods are considered reliable.

**Keywords:** *Canis lupus*, accuracy, precision, reproducibility, science-policy interface, sensitivity

#### Introduction

The abundance of wild animals is often central to policy and of interest to many publics. A century has been spent in devising and evaluating methods to estimate wildlife abundance (1). Mammalian carnivore abundances can be particularly difficult to estimate accurately, when they are naturally low-density, when their habits or habitats are inconvenient for humans, or when individuals have overlapping ranges but cannot be distinguished individually. The latter issue of counting known individuals or reliable use of mark-recapture methods has animated the field for some time (2-6). Scientists in many regions have raised concerns over dubious estimates of abundance influencing policy, especially when those estimates are not verified independently or the methods diverge from those validated (7, 8). Such concerns have arisen lately in three

jurisdictions of the USA where estimates of gray wolf (*Canis lupus*) abundance relied on new methods.

Wolf populations in many regions are recolonizing their former range in increasing numbers [European Union: (9); USA: (10)]. Wolf recolonization has triggered sociopolitical clashes over how to respond (11-15). Different worldviews of wolves surfaced as interest groups vied for control of policy-making processes. Persistent points of friction in public policy debates over wolves center on the abundance and geographic distribution of individuals and packs and the role such variables should play in policy.

The number and geographic extent of wolves is thought to serve as a correlate of the benefits and costs of wolves to humans and to ecosystems, and as a measure of the success or failure of policies. For example, decisions over the (de)listing of species under the USA Endangered Species Act largely rely on data describing number of breeding pairs and their rate of change over time. Wolves in the USA are still listed as threatened or endangered in most regions. Therefore, recent actions by some jurisdictions to change methods for estimating wolf abundance and geographic distribution have raised concerns over accuracy.

In particular, three states in the USA have recently changed methods for estimating wolves across large areas, as both the cost of the task has grown and the sociopolitical scrutiny of policy has intensified. In 2021, the Fish, Wildlife & Parks agency of Montana, USA (MT) changed from monitoring intensive efforts and a patch occupancy model (POM; (16, 17) to predicting population size by combining predictions on area occupied with predictive models of home range size and pack size (called an 'integrated population occupancy model' iPOM (18). In 2021, the Fish & Game agency of Idaho (ID) changed to space-to-event models with stationary cameras (19, 20). In 2021, the Department of Natural Resources of Wisconsin (WI) transformed its efforts to count every wolf into an effort to estimate statewide population size with a scaled occupancy model ('SOM') that combines estimates of pack size, pack ranging area, and occupied range, similar to MT (21, 22).

Policies that liberalized killing of wolves (expanded methods, timing, and participation in hunting, trapping, and hounding) in those same states raised concerns about the sensitivity of the new methods to changes in widespread human-caused mortality (7). The same three jurisdictions substantially expanded both the methods and opportunities for private individuals to pursue or kill wolves, in ID (23), MT (24), and WI (25, 26). As a result, concerns arose that the new methods for estimating wolf abundance are inadequate to detect population changes (7). If the concerns are well-founded, then states would not be safeguarding populations from excessive exploitation and risk of extinction. Therefore, reliable methods to estimate wolf populations are swiftly needed not only to continue to inform policy, but especially to address the consequences of regulatory changes that likely raised wolf mortality and can hinder or reverse wolf population growth (27-31). The reliability of new methods is particularly important to federal and state policy in those three jurisdictions and to their claims that wolf policy is science-based. Independent verification of state methods may raise public confidence in the actions and statements of public

trustees whose primary legal duty is to preserve wildlife for future generations, e.g., *Hughes v Oklahoma* 1979 (32).

### **Attributes of valid methods for estimating abundance of wild animals**

Reliability of measurement methods and estimation techniques in quantitative sciences can be conceptualized using the criteria of *precision*, *accuracy*, *sensitivity*, to changing conditions, and *reproducibility*. *Precision* increases as the margin of error ('confidence interval' or 'credible interval') around the estimate decreases. Therefore, precise estimates of wildlife abundance are more certain or credible, with narrower clustering of repeated measures, and thereby produce more confidence in policy.

*Accuracy* refers to measurements that are neither systematically lower nor higher than the actual value and errors are random with regard to the actual value, i.e., accurate measurements are unbiased. Measurements that are highly precise but inaccurate provide false confidence about a poor approximation of reality. Likewise, measurements that are accurate but imprecise give low confidence about a good approximation to reality and can obscure trends toward increase or decrease.

An accurate and precise estimate in one period may be reliable under stable conditions but generate biased (inaccurate) estimates or lose precision (credibility) when conditions change the range of input variables. This attribute of measurement methods we call *sensitivity*.

*Sensitivity* is achieved when methods of measurement adjust to changes across periods (i.e., generalizable). For example, a method for estimating occupancy and abundance that was designed outside of hunting seasons may not generalize to a hunted population. Insensitive measurements are not generalizable or cannot be confidently extrapolated for policy-making that reliably anticipates future conditions.

Finally, *reproducibility* refers to replication via independent attempts. No matter how accurate, precise, or sensitive a measurement is claimed, if independent efforts to replicate those attributes fail despite good faith efforts, the method is fundamentally unscientific. Therefore, policy-making on that basis is likely to fail in unpredictable ways.

### **Background to wolf abundance in USA and Wisconsin policy**

As USA states try to save time and effort in estimating their wolf abundances while imposing more killing on those populations, proponents of the new methods should invite independent review and demonstrate accuracy, precision, sensitivity, and reproducibility of the methods through the usual processes of scientific debate, peer review, transparency, replication, correction, retraction, revision, and persuasion until the wolf science community comes to consensus. That was the process for the older methods. To estimate wolf abundance by meeting the criteria above, the scientific community has come to consensus over a century or so that one needs a large sample of marked animals detected again over a year or more with methods that account for migrants, variability in pack structure and territory sizes, and ideally some understanding of birth and death rates (33-37).

We offer the following discussion of concerns about the new Wisconsin method in hopes of informing public policy debate, whether in the USA or abroad, whether federal, state, tribal, or private. We have a particular audience in mind in the USA federal government because it collects information on these and other jurisdictions' wolf monitoring programs; wolves have been periodically protected, and still are in some areas such as Wisconsin, under the USA Endangered Species Act (ESA). The ESA requires the Secretaries of Interior and Commerce to base determinations "solely on the basis of the best scientific and commercial data available" (16 USC. §§ 1531 Sec 4(b)(1)(A) and Sec.7 "Interagency cooperation"), including those to relist species. Moreover, monitoring and population estimation methods are an essential part of, or pre-condition to, each of the five factors the USFWS should analyze when making those determinations. For examples of two factors, population estimates are essential for assessing "overutilization" or for assessing the (in)adequacy of regulatory mechanisms in states hosting wolves (16 USC. §§ 1531 Sec 4(a)(1) "Determination of Endangered Species and Threatened Species"). We also note the many invested interest groups including state and tribal governments, litigants, lobbyists, and the public who continue to pay careful attention to rhetoric and to science in wolf policy. Setting aside the important (but rare) debate over whether the number of wolves (census population size) is a more meaningful or useful measure for policy decisions than the number of packs, geographic distribution, net benefits-costs, or estimates of viability (e.g., effective population size), we focus here on the reliability of the science brought to recent changes to methods for estimating abundances.

Before proceeding to the case of Wisconsin's new method for estimating wolf abundance and geographic distribution, we have a disclaimer and two acknowledgments. We acknowledge that counting wolves and many other rare and elusive wildlife is very difficult, time-consuming, expensive, and the eventual estimates are often disputed. Indeed, mitigating costs, difficulty, and controversy in part motivated the recent methodological shifts for estimating wolves. As publicly funded scientists, we see it as our duty to evaluate the methods, and their assumptions and application, for the broadest public to judge policy (38). Second, we recognize that government scientists in all jurisdictions are increasingly pressured by political appointees to deliver answers that the politicians like. None of our writing should be construed as a personal criticism of state agency scientists. Our disclaimer is that we will not compare the current methods to the previous method. However, in every case when we identify a current problem we point to a feasible solution. We hope our comments are viewed in the spirit intended, as constructive criticism.

Wisconsin's 'scaled occupancy model' (SOM) applied in 2022: The new method adopted by WI's Department of Natural Resources WDNR in 2022 (21) takes the raw data from the traditional "territory mapping method" and replaces the last step of the population estimation process with new steps. The traditional method involved primarily snow-track surveys, supplemented by aerial telemetry, and observation of a small number of collared wolves supplemented by summer howling surveys to detect litters of pups (39-41). The final step of the traditional method was a public process (from 1996 approximately to 2012) of counting the number of wolf packs; physically placing a marker on a large map of the state in public; and

writing the estimated number of wolves on each pack's location for summing the total. From 2012–2017 it was somewhat less public (38). Later the traditional method was supplemented by early versions of the SOM and then supplanted in 2022.

The WDNR SOM (21) took the data from the traditional counting method (snow-track surveys and telemetry) and replaced the population estimation process described above with a model-assisted estimate. In essence, the SOM is a sampling method while the traditional method attempted a complete enumeration analogous to the USA census. The new steps in the SOM treat evidence of wolves as the basis for probable pack occupancy across a large number of relatively small survey blocks, then apply an estimate of home range size and pack size to extrapolate the statewide population across estimated occupied range. These new steps rely on the predictive model called the 'scaled occupancy model' first recommended for WI wolves in (22).

The WDNR deserves praise for two aspects of their transition to the use of the SOM. First, they have continued to use actual field snow tracking observations collected by many community volunteers to supplement state, tribal, and federal staff efforts. Second, the WDNR used both methods side by side for several years, so that scientists and the public could see how the SOM outputs related to the traditional census. Although the raw data have not been provided despite our repeated requests, the final outputs and their uncertainty (precision) can be examined for several years before the 2022 implementation of the SOM alone in 2022 (21). That period of simultaneous traditional method and SOM were therefore conducted under similar management and ecological conditions. However, the decision to adapt (22) to the situation in 2022 (21), carries with it both the typical scientific assumption that a model from the past can predict the present and future and also a secondary, more tenuous assumption. The tenuous assumption is that the model in (22) can generate reliable estimates when conditions are dramatically different in input conditions (sensitivity). For example, (22) was developed in the absence of legal wolf-killing and compared to the traditional method entirely under conditions without wolf-hunting or state-sanctioned killing of suspected predators of domestic animals (22). The policy change of the 2021 wolf-hunt (25), the first since 2014, and the novelty of such a hunt during breeding season raises questions about the relevance of (22) to any period with legal wolf-killing. We examine that model's relevance after describing the WDNR SOM of 2022 (21).

The SOM has three components, each estimated separately: an estimate of the area occupied by wolves, an estimate of territory size, and an estimate of pack size. To understand the SOM, one must scrutinize the methods for estimating each of the three components. Below, we focus on precision and sensitivity of the measurement methods and spend less time discussing accuracy because we still await WDNR data >9 months after our first of two requests.

### **Input Data**

Input data from winter track surveys: In 2022, the WDNR reported core range of wolves and evidence of wolves based on winter track surveys, telemetry, and other observations less systematically described (21). For winter track surveys, community volunteers and government

agents traveled survey blocks of irregular size and shape. They traveled these blocks 0-13 times during the two winters of 2021 and 2022 with one modification. Because "...of the Feb 2021 wolf-hunt, only 14,000 km of **pre-hunt survey effort were considered in the model.**" (emphasis added) (21). Importantly, the boldface phrase emphasizes how, for the first time ever, two winters' worth of field data were combined for the September 2022 estimate of the statewide wolf population. Also, estimates of wolf presence were used over four years to designate core wolf range in 2022 as explained next.

Input data for area occupied by wolves: Winter tracking surveys provide inputs for the area putatively occupied by wolves. The description of this method differs in two documents (21) and its Addendum as follows: The Addendum states, "winter tracking blocks with 'confirmed pack activity' during **at least one of the previous four years**" (emphasis added, p.3), but later defined it "as winter tracking blocks with confirmed pack activity **during the previous four years.**" [emphasis added, p.26 of the Addendum to (21)]. The two sources create confusion that cannot be resolved simply by a query. Rather the WDNR should resolve the confusion by publishing a correction. Until then, the method cannot be reproduced (a hallmark of science). Subsequent statements do not clarify: "Four years was identified as the number of years which allowed the core range to respond to possible expansions and contractions of wolf range, while minimizing inclusion or exclusion based on transient wolf movements or imperfect detection of wolves in pack-occupied areas." p. 26, (21). We fail to see any scientific evidence for those statements.

In sum, the data input to the SOM comes from previous years (two winters of snow track surveys and four years of pack-occupied area estimates). This may help explain unusual claims by WDNR. First, (21) claimed that the occupied area increased to 28824 mi<sup>2</sup> when compared to the pre-hunt occupied area reported in the 2020-2021 population report (28493 mi<sup>2</sup>). That is an unusual claim after the high wolf death toll in 2021, estimated at more than 27-33% of the population 12 months earlier (25). Second, in 2020, when the census method was compared to the SOM method using the same data, the SOM estimate produced a much higher estimate of the statewide number of wolf packs whose lower bounds did not include the traditional method with its narrow bounds (high precision). Therefore, prior years in which both traditional and SOM method were used would lead us to expect the SOM to over-estimate the number of packs in the state. There is no apparent effort in (21) to address or remedy the issue. Several unjustified steps in handling the input data may explain this apparent range expansion and the surprising stability in number of packs after the February 2021 wolf-hunt (described below).

### **Data handling and analysis**

An area was included in 'core wolf range' if either of the following criteria was met: "Tracks from at least two wolves were observed within a block during a single tracking event" or "Single wolf tracks were observed in a grid during separate surveys within a tracking season" (p. 26 and 28 respectively). Because single tracks in the snow do not confirm the presence of a pack, this step of the analysis can inflate the estimated 'core wolf range' that will be later multiplied by pack size.

Also, 25 survey blocks (of 156; 16%) were surveyed 0-1 times in winter 2021-2022. The tracking data used from 2020-2021 is equally or more concerning, as there were no surveys done on 21 blocks (of 154 blocks, 16%) and only 1 survey on 50 blocks (32%). Surveying a block once or twice makes it impossible to verify or replicate the first count and impossible to rule out double-counting the same wolf laying down two sets of tracks at different times. It also departs from decades of precedent by the WDNR (41-44). A departure from past methods is seen in the use of snow track survey data collected with fewer than 3 surveys per census block. Indeed, by estimating area occupied for survey blocks surveyed <3 times [62 of 154 survey blocks, 40%; Fig. 5, p.18 in (21)] the WDNR violates its own quality control rules established in the early 2000s (44-46). We had previously warned that changing methods inflated the variability (imprecision) of population estimates after methods of counting wolves changed 3-4 times from 1979-2012 (47).

Moreover, tracks in the snow can be of non-wolves, transient lone wolves or packs that have disappeared after the first year they were detected. Wolves that dispersed since being counted or spend most of their time in tribal reservations, Michigan, or Minnesota, might also have left a track that resulted in a survey block being included in 'core wolf range', which was not clearly defined in (21). For example, numerous wolves made long-lasting, long-distance extraterritorial movements (48), which might have produced tracks in areas unoccupied by wolf packs. This concern is only heightened by the potential proliferation of false positives (identifying a pack where none exists) due to the social disruption and pack disbandment that may be caused by wolf-killing (49-53). Many packs could have disappeared given increases in human-mortality throughout wolf range during and after the February 2021 wolf hunt that killed 218 wolves legally and >100 illegally (25). Four to five dozen surviving wolves also faced killing by USDA-WS in summer 2021 (54).

To illustrate our concern with inflating occupied range with spurious packs, we call attention to survey block 167 in the winter 2020-2021 population report. Survey block 167 appeared in the report for 2020-2021 (55) but not in the report for 2021-2022 (21), with no explanation for its omission. No surveys were conducted there in winter 2021-2022 [Figure 5, p. 18 in (21)], yet it is included in pack-occupied wolf range and core range [Figures 6, 7, and 13 in (21)]. The latter figures show occupancy probabilities, wolf density estimates, and included and excluded survey blocks respectively. Indeed, there is a probability of occupancy of survey block 167 in Figure 6 despite no snow-track survey there, no anecdotal observation of wolves there [Figure 1, p.14 in (21)], and no verified complaints of domesticated animal loss there in 2021-2022 [Figure 2, p.15 in (21)]. This is direct evidence that (21) used data from prior years to estimate wolf numbers. Survey block 167 is only unique in its obvious position outside the main population range, making it obvious on visual inspection. Currently, we have no way of knowing how many other survey blocks were populated with wolf tracks from years before the February 2021 wolf-hunt.

Finally, for occupied range, we have concerns about neighboring and included jurisdictions whose wolves should not legally be included in the statewide count (56). A basic principle of spatial models such as the SOM is to accommodate the boundaries with unoccupied or

unstudied areas, such as Michigan, Minnesota, large bodies of water such as Lake Superior, and tribal reservations such as Menominee County. Exclusion of these regions and care in handling adjacent areas would avoid simulating wolves that do not exist or double-counting wolves counted by neighboring jurisdictions. However, (21, 22) are not transparent on these issues. If a wolf pack territory were simulated to overlap any of these neighboring areas, the statewide count would overestimate the wolf population size by counting an area as occupied by a pack full-time when it was in fact unoccupied water or claimed by neighboring jurisdictions. The SOM grid cells neighboring those problematic areas should be deleted to avoid the state claiming wolves that spend most of their time in non-Wisconsin areas, especially tribal reservations. In sum, we have numerous concerns about overestimation bias and arbitrary and capricious inclusion and exclusion criteria.

The extent of the overestimate is impossible for us to disentangle without greater transparency from the WDNR. One dubious result is that the number of packs remain virtually unchanged between years: 292 packs in 2020-2021, and 288 packs in 2021-2022. We surmise the introduction of data from winter 2020-2021 (and earlier years) raises the likelihood that (21) populates the state with simulated wolves. Due to lack of transparency about the number of data points collected prior to December 2021, inclusion or exclusion of other jurisdictions' wolves, and how the WDNR handled surveys that counted only one wolf (in the majority of efforts in that block), we cannot correct how much their over-estimation bias inflates the state wolf population.

### **Occupancy model**

The inputs from above are survey blocks ostensibly occupied by wolf packs. To assign probability of occupancy given the input data, a land cover model was built, based on forest cover, proportion of agricultural and developed land (2016 National Land Cover Data, NLCD), and road density (22). A major concern with the exclusive use of such variables is that none of them change annually or even considerably over 5 years or longer. Hence, any changes in conditions beyond those captured by the model but that affect occupancy, such as increases in human-caused mortality, would not affect the occupancy estimates (i.e., the model is insensitive). Without ground-truthing, the model used to estimate occupancy is also speculative. Any event that changed the land cover or the presence of wolves since the land cover model was produced would alter the probability that there are packs there now. In particular, the February 2021 wolf-hunt would have changed occupancy probabilities, as would any future increased use of lethal methods. That means the assumed probability of wolves occupying a given grid cell is of *potential* (or simulated) packs, not counted packs.

The above oversight is troubling given the many studies showing that human-induced killing affects wolf population, pack, and individual responses, sometimes beyond their numerical effects (28, 29, 33, 34, 36, 37, 49-51, 57-61), and increased risk of illegal killings (30, 31, 62, 63). Liberalizing lethal methods is a pervasive condition that structures wolf population dynamics, whereas the (22) predictive model had underlying assumptions of full protections and saturation of range with packs, and therefore different population dynamics that are unaccounted for by the 2022 data. Although such disruption of wolf dynamics and their effect on



occupancy estimates are not addressed in their article, Stauffer et al. (22) do note that “a situation with highly dispersed packs and extensive interstitial space is more subject to positive bias than a situation where packs are clustered.” (p.1418). Given the above, assigning occupancy probabilities based on forest, agriculture, and roads without considering public hunts as a critical model variable can only produce overestimates of occupancy probability, never underestimates.

Moreover, Stauffer et al. (22) was not validated with data following wolf-hunts in WI. Rather, the model was validated on data from winters 2016-2017 and 2018-2019, which came several years after the wolf-hunts of 2012-2014, despite the WDNR authors having had access to all the requisite data (22). We are unclear why they chose to validate their model with the small subset of all the data at their disposal. This is an unjustified assumption and step in that paper. Therefore, it is unknown whether the model in (22) is sensitive to changes in wolf populations after a wolf-hunt, especially an unprecedented hunt in timing, methods, and proportion of the wolf population estimated to have been killed legally and illegally (25). Indeed, the model in (22) was exclusively validated with data of periods during which wolves enjoyed full ESA protections (classified as ‘endangered’), and hence we suggest this is an underlying and substantial model assumption that has remained unidentified but that challenges the model’s application to liberalized killing anywhere at any time but for the two years they hand-picked for validation. The assumptions of (22) are numerous: habitat saturation, static land cover variables, little overlap in wolf pack territories with neighbors and with independent jurisdictions, no legal wolf-killing (except for human health and safety). They also assume the average territory size and average pack sizes represent accurately the entire state’s wolf packs, and normal distributions around those averages. Some of those assumptions were not made clear in (22) despite peer review, but others were transparent. For example, (22) stipulated that “If the number of occupied sites adequately describes the distribution of home ranges (i.e.,  $\psi$  is appropriately defined and interpreted, and is unbiased), then population abundance can be derived...” p. 411, (22). The WDNR has not addressed such assumptions in (21) and we show the numerous sources of over-estimation bias that make us skeptical. Therefore, (21) has misapplied a model to a new set of conditions, without validation. The claim that they use a peer-reviewed method is not accurate with regard to its application. This is analogous to treating an illness with a therapy validated for a different health condition dissimilar from the former.

### **Wolf pack home range size**

Next, the pack-occupied area with some unspecified probability of occupancy was divided into estimated wolf pack territories statewide. These wolf pack areas are variously referred to as home ranges or territories because we lack systematic information on whether the ranges were defended as territories. The WDNR estimated pack home ranges from a small number of GPS-collared wolves monitored by telemetry (21). (21) relied on fewer than two dozen GPS collared wolves (<3% of the population by all estimates) from December 2021 to April 2022 to estimate home range size and did not reveal which zones those wolves ranged within. This is a small sample compared to past years, e.g., (41) reported 13% of the population on average. Moreover, the agency did not report the number of GPS data points (e.g., number of locations per year) used in (21). This small sample raises concerns over the estimates themselves

because such estimates are dependent on the number of locations, with small sample sizes leading to underestimation of territory sizes, and thus overestimation of the number of territories (64).

Regarding how the GPS locations were transformed into territory sizes, the WDNR Addendum describes a highly subjective process in 6 steps [p.28-29, (21)]. Because an individual analyst's subjective judgments were used for an unstated proportion of the ranges, we doubt the process could be replicated by others (irreproducible). Moreover, the model provides no measures of sensitivity to changes in such an estimate, overall or by zone. Namely, (21) employs a single value for statewide average home range size (171 km<sup>2</sup>), despite their own data revealing that home ranges differ markedly by zone and have changed over time (41). One step to avoid overestimation bias if one does not have unbiased, reliable zone-specific home range estimates is to use the median from all wolves that were observed in the same manner (e.g., via GPS collars). We recommend more effort at collaring if the state retains the SOM approach.

### **Wolf pack size**

Next, the estimated statewide number of territories were multiplied by estimated pack size for a total statewide count (perhaps omitting tribal reservations). Home range shape, as well as grid placement relative to where the actual territory boundaries lie (and their spatial relationship) could materially affect the number of estimated territories within state lines: Stauffer et al.'s (22) best simulation exercise(s) with saturated territories (the assumption made for Wisconsin core wolf range) still resulted in an overestimation in *number of territories* of 1.38-1.9 times the true value [see Figure 7, (22)]. That assumption of range saturation was made for two or more winters after the last fall wolf-hunt, during periods of full protections, and still resulted in considerable overestimation. Yet, (21) applied that assumption to a period of one year after an unprecedented February wolf-hunt, and we cannot find any evidence of attempts to acknowledge or correct for such overestimation.

Moreover, (21) uses a zone-specific average pack size ranging from 4.13-2.7 adult wolves. It is a well-known statistical principle that an average is pulled high or low by extreme values and in this case the range of values for pack sizes is bounded by 2 on the left and 12-13 on the right historically (40, 41). Therefore, extreme pack sizes and inflated counts in a particular survey block can pull the zone-specific average up artificially. Indeed, the median is a safer measure because it is defined as the point estimate at which half of the wolf packs are smaller and half are larger. We recommend the WDNR present all pack size counts and how that pack was counted in each survey block. This would enable independent review and replication of their estimates. Previous work recommended methods and communications that would enhance transparency (26, 47).<sup>1</sup>

### **Bias and uncertainty**

---

<sup>1</sup>Note that our recommendation does not provide locations of wolf packs per se, just the count per survey block and the basis for that count (effort, by whom, collar, range of counts per survey attempt). Hence, there should be no concerns over concealing pack locations to avoid illegal killing.

Our conclusion of over-estimation bias should be considered in light of the very low probability of an under-estimation bias. We can imagine an unlikely set of conditions under which the (21) SOM approach would under-estimate the state wolf population. First, the 2022 territory size would have to be a significant over-estimate. If pack ranges were significantly smaller than 171.5 km<sup>2</sup>, then the state could fit many more (smaller) packs into the state. Although there is a low probability that the few GPS-collared wolves had extremely large home ranges, resulting in fewer packs being estimated statewide by the WDNR SOM method, the effects would be unclear because larger packs use larger ranges. Hence, the number of territories and the sizes of the packs that use those territories tend to be inversely related (34), so tweaks to pack size and range size tend to cancel each other out when it comes to estimating statewide number of wolves. By contrast, the risk of double-counting a pack would create a more consistent over-estimation bias. Every time a pack is mistakenly double-counted, that inflates the statewide count, so imagining an under-estimate of the statewide total requires zero double-counting. Avoiding double-counting would depend on survey blocks having been chosen to encompass tracks of only one pack, zero tracks of other packs, and no loners or transients. Double-counting is practically guaranteed by the (21) method of incorporating data from prior years to define pack-occupied range. Furthermore, the occupancy model designates its area of analysis after data are collected on any verified wolf sightings across the state, so it seems implausible that the actual occupied wolf range exceeds the bounds of analysis (indeed Figure 6, p.9 suggests the converse, margins have 0-20% probability of occupancy). Moreover, the very small packs that historically occur on the periphery of the wolf population would not contribute meaningfully to the entire state estimate. Therefore, an under-estimate of the state wolf population seems to depend on an extremely rare set of conditions and errors in data collection for which we have no evidence. Given how difficult it is to argue for an under-estimate, the reasonable conclusion is that our concerns about over-estimation deserve the focus of independent researchers and reconsideration by the WDNR.

The three prior estimates (area occupied, territory size, and pack size) were not fixed values themselves, but rather varied across space and over time in response to ecological conditions, including human influences that permeate wolf range, such as hunting. Because each component of the SOM is estimated separately, they each carry their own uncertainty (i.e., precision), which compound in the final population estimate to broaden the uncertainty. The result is a broad range of uncertainty around single values and a broad range of likely values. The meaning of a broad range of likely values has been lost in (21) and its associated greensheet, because both documents identify a single point estimate for the state population size. This so-called “mode” does not seem to us to deserve the attention it garnered from the Natural Resources Board 28 September 2022, time stamp 2 hours). In statistical terms, the mode is the most frequently measured value. Yet, points around the mode may be near or equally credible in a model such as (22). We are dubious this mode can be defended on statistical grounds, so we call for the WDNR to publish the probability distributions and revise its confident claim about the statewide wolf population estimate to a range of values. Also, (21) presents little or no data on breeding success, so the estimate of abundance in 2022 is based on presumed breeding in summer 2021 (26) without systematic evidence for pup survival to winter 2021-2022. We demonstrated how WDNR might communicate uncertainty about

abundance in the latter paper, therefore we do not repeat the many recommendations on transparency and even-handed explanations of scientific uncertainty.

### **Recommendations**

We recommend future efforts start with repeated, independent counts of two or more wolves as the only evidence for pack-occupied areas, while also noting the possible overestimation bias inherent to that assumption (not all pairs of wolves are a pack, some pairs will cross multiple survey blocks, etc.). Then, ground-truthing should estimate how often that assumption is violated in each of the six management zones and deduct a proportion of packs accordingly.

We also recommend the use of number of packs as more informative than counts of individual wolves for projecting population dynamics. We also consider packs more informative for evaluating the benefits and costs of wolf coexistence with other organisms such as deer, pathogens, domesticated animals, and humans. The (21) is not a conservative approach given it is estimating abundance of individual wolves with a coarse and outdated occupancy model that does not consider increases in mortality. Moreover, (21) repeatedly failed to assess assumptions of the published model. A conservative approach would reduce the risk of falling below legal and social thresholds (26) and lead to under-estimation bias, not the current over-estimation bias.

Further, we recommend the WDNR continue its practice of enumerating wolf packs on tribal reservations separately, rather than treating them as the statewide total as was done in (22) and the Natural Resource Board meeting of September 2022 (see [www.youtube.com/watch?v=V-Pp-skgBVk](https://www.youtube.com/watch?v=V-Pp-skgBVk), time stamp 2 hours, accessed 1 May 2023). We cannot estimate the number of wolves that are not under state authority yet were counted in the state wolf population estimate (21) because the data on locations have not been presented transparently.

We recognize that agency staff often face intense scrutiny and probably undue and uncomfortable political pressure to work quickly towards political goals. We call for the insulation of agency scientists from political appointees when – and this is the critical step – the agency scientists are asked a scientific question such as ‘how many wolves live in Wisconsin?’ That question is purely scientific and has no relationship to ‘how many wolves should live in Wisconsin?’ or ‘ought we to hunt wolves?’, which are value-based questions they are not trained to address (65, 66). The public and politicians should transparently debate the value-based questions (65, 66). Similarly, the politicians should steer clear of the scientific question because they are not trained to set aside their personal preferences when evaluating fact claims (67). Likewise, scientists should not treat value-based debates as if science resolved those. Scientists can respond as do other members of the public when addressing values. They should not treat value-based assumptions and policy decisions as if their scientific expertise gave their opinions more weight, nor treat such issues as if science resolved the issues (67). Nor should scientists who participate in hunting mislead the public or themselves that they are more objective, impartial, or unbiased than scientists who do not participate in hunting (47, 66, 68, 69). Science respects no authority or expertise.

Finally, results of predictive models developed and validated under particular conditions are limited to those conditions (7). Scientific confidence in the models declines as the extrapolation to novel conditions expands and as the assumptions of the original model are violated (7). For models like (22) that multiply three different estimates or models together, the uncertainties multiply (7). We have shown irreproducible input values and several, large biases, insensitivities and imprecisions in how the WDNR went about relating observations of tracks in the snow to the pack-occupied range and pack size in (21). Therefore, we conclude that (21) is unreliable because it systematically overestimates the statewide wolf abundance, particularly when hunting seasons and killing are in effect or recently passed. Moreover, given the WDNR is required by law to hold a wolf hunt if the wolf is not federally listed, future agency policy would clearly violate the model assumption of full protections underlying the modeling methods (21, 22).

### **Acknowledgments**

We thank P. David and M. Lagally for scientific input. No financial support was provided beyond our employers' support of our time.

**Funding:** There are no funding sources to declare.

**Author contributions:** Both authors contributed equally to analysis and writing. AT led on submission and revision.

**Conflicts of interests:** The authors declare no conflicts of interest. Full disclosures of potentially competing interests for both authors are presented at [http://faculty.nelson.wisc.edu/treves/archive\\_BAS/funding.pdf](http://faculty.nelson.wisc.edu/treves/archive_BAS/funding.pdf) (AT) and [http://faculty.nelson.wisc.edu/treves/archive\\_BAS/Santiago-Avila\\_CV\\_2023.pdf](http://faculty.nelson.wisc.edu/treves/archive_BAS/Santiago-Avila_CV_2023.pdf) (FJS-A).

**Data availability statement:** We stored documents used for this review -- state population reports for 2020-2022 and the peer-reviewed paper on which the WDNR's SOM method was based (21, 22, 55) with the addendum to (21), and its 'greensheet' (a summary required by the state Natural Resources Board) -- on an institutional server at [http://faculty.nelson.wisc.edu/treves/archive\\_BAS/WDNR\\_SOM\\_Critique/Source\\_documents\\_archive.zip](http://faculty.nelson.wisc.edu/treves/archive_BAS/WDNR_SOM_Critique/Source_documents_archive.zip).

**Institutional review board statement:** Not applicable

**Informed consent statement:** Not applicable

### **References**

1. Leopold A. Game Management. New York and Madison, Wisconsin: Scribner 1933 , University of Wisconsin Press 1986; 1933 reprinted 1986.
2. Elliot NB, Gopalaswamy AM. Toward accurate and precise estimates of lion density. *Conserv Biol.* 2017;31(4):934-43.
3. Balme G, Slotow R, Hunter LTB. Edge effects and the impact of non-protected areas in carnivore conservation: leopards in the Phinda-Mkhuze Complex, South Africa. *Anim Conserv.* 2010;13(3): 315-23.
4. Balme G, Hunter LTB, Slotow R. Evaluating Methods for Counting Cryptic Carnivores *J Wildl Manage.* 2009;73(3):431–43.

5. Davis RS, Gentle LK, Stone EL, Uzal A, Yarnell RW. A review of spotted hyaena population estimates highlights the need for greater utilisation of spatial capture-recapture methods. *Journal of Vertebrate Biology*. 2022;71(22017):22017-1.
6. Karanth KU, Nichols JD. Estimation of tiger densities in India using photographic captures and recaptures. *Ecology*. 1998;79:2852-62.
7. Creel S. Methods to estimate population sizes of wolves in Idaho and Montana. Comment on "Endangered and Threatened Wildlife and Plants; 90-Day Finding for Two Petitions To List the Gray Wolf in the Western United States". *Federal Register*. 2021;86:51857.
8. Karanth KU, Gopalaswamy AM, Kumar NS, Delampady M, Nichols JD, Seidensticker J, et al. Counting India's wild tigers reliably. *Science* 2011;332(6031):791
9. Chapron G, Kaczensky P, Linnell JDC, von Arx M, Huber D, Andr  n H, et al. Recovery of large carnivores in Europe's modern human-dominated landscapes. *Science*. 2014;346(6216):1517.
10. USFWS D. Endangered and Threatened Wildlife and Plants; Removing the Gray Wolf (*Canis lupus*) From the List of Endangered and Threatened Wildlife (Final rule). *Federal Register*. 2020;85(213):69778-895
11. Bruskotter JT, Toman E,ENZLER SA, Schmidt RH. Are Gray Wolves Endangered in the Northern Rocky Mountains? A Role for Social Science in Endangered Species Listing Determinations. *Bioscience*. 2010;60(11):941-8.
12. Chapron G, L  pez Bao JV, Kjellander P, Karlsson J. Misuse of Scientific Data in Wolf Policy. *Science*. 2013;339:1521.
13. Chapron G, Lopez-Bao JV. Conserving Carnivores: Politics in Play. *Science*. 2014;343(14):1199-200.
14. Nie M. The Sociopolitical Dimensions of Wolf Management and Restoration in the United States. *Hum Ecol Rev*. 2001;8(1):1-11.
15. Nie M. Beyond Wolves: The Politics of Wolf Recovery and Management. Minneapolis, MN, USA: University of Minnesota Press; 2003. 253 p.
16. Miller DA, Nichols JD, Gude JA, Rich LN, Podr  zny KM, Hines JE, et al. Determining occurrence dynamics when false positives occur: estimating the range dynamics of wolves from public survey data. *PLoS one*. 2013;8(6):e65808.
17. Rich LN, Mitchell MS, Gude JA, Sime CA. Anthropogenic mortality, intraspecific competition, and prey availability influence territory sizes of wolves in Montana. *J Mammal*. 2012;93(3):722-31.
18. Sells SN, Podr  zny KM, Nowak JJ, Smucker TD, Parks TW, Boyd DK, et al. Integrating basic and applied research to estimate carnivore abundance. *Ecol Appl*. 2022;32(8):e2714.
19. Ausband DE, Lukacs PM, Hurley M, Roberts S, Strickfaden K, Moeller AK. Estimating wolf abundance from cameras. *Ecosphere*. 2022;13(2):e3933.
20. Moeller AK, Lukacs PM, Horne JS. Three novel methods to estimate abundance of unmarked animals using remote cameras. *Ecosphere*. 2018;9(8):2331.
21. WDNR. Wisconsin Gray Wolf Monitoring Report 15 April 2021 through 14 April 2022. In: WI Department of Natural Resources BoWM, editor. Madison, WI, USA.: WI Department of Natural Resources, Bureau of Wildlife Management; 2022.
22. Stauffer GE, Roberts NM, MacFarland DM, Van Deelen TR. Scaling Occupancy Estimates up to Abundance for Wolves. *The Journal of Wildlife Management*. 2021;85(7):1410-22.
23. Ridler K. Bill to kill up to 90% of Idaho wolves signed by governor. *The Seattle Times*. 2021 7 May 2021.
24. Daly N. Montana has made killing wolves easier. Some hunters are pushing back. *National Geographic*. 2021.
25. Treves A, Santiago-  vila FJ, Putrevu K. Quantifying the effects of delisting wolves after the first state began lethal management. *PeerJ*. 2021; 9: e11666.

26. Treves A, Louchouart NX. Uncertainty and precaution in hunting wolves twice in a year. *PLoS One*. 2022; 17(3):e0259604.
27. Chapron G, Treves A. Reply to comments by Olson et al. 2017 and Stien 2017. *Proceedings of the Royal Society B*. 2017; 284(1867):20171743.
28. Creel S, Rotella JJ. Meta-analysis of relationships between human off-take, total mortality and population dynamics of gray wolves (*Canis lupus*). *PLoS ONE*. 2010;5(9):1-7.
29. Creel S, Becker M, Christianson D, Dröge E, Hammerschlag N, Hayward MW, et al. Questionable policy for large carnivore hunting. *Science*. 2015;350(6267):1473-5.
30. Santiago-Ávila FJ, Agan S, W., Hinton JW, Treves A. Evaluating how management policies affect red wolf mortality and disappearance. *Royal Society Open Science*. 2022; 9:210400. .
31. Louchouart NX, Santiago-Ávila FJ, Parsons DR, Treves A. Evaluating how lethal management affects poaching of Mexican wolves (registered report). *Open Science*. 2021;8:200330.
32. *Hughes v Oklahoma*. U.S. Supreme Court 441 U.S. 322; 1979.
33. Fuller TK. Population dynamics of wolves in north central Minnesota. *Wildlife Monographs*. 1989;105:3-41.
34. Fuller TK, Mech LD, Cochrane JF. Wolf population dynamics. In: Mech LD, Boitani L, editors. *Wolves: Behavior, ecology, and conservation*. Chicago: University of Chicago Press; 2003. p. 161-91.
35. Vucetich JA, Peterson RO. Wolf and moose dynamics on Isle Royale. In: Wydeven AP, Van Deelen TR, Heske EJ, editors. *Recovery of gray wolves in the Great Lakes region of the United States: An endangered species success story*. New York: Springer; 2009. p. 35-48.
36. Vucetich JA. Appendix: The influence of anthropogenic mortality on wolf population dynamics with special reference to Creel and Rotella (2010) and Gude et al. (2011) in the Final peer review of four documents amending and clarifying the Wyoming gray wolf management plan. *Federal Register*. 2012;50:78-95.
37. Adams LG, Stephenson RO, Dale BW, Ahgook RT, Demma DJ. Population dynamics and harvest characteristics of wolves in the Central Brooks Range, Alaska *Wildlife Monographs*. 2008;170:1-25.
38. Treves A, Chapron G, López-Bao JV, Shoemaker C, Goeckner A, Bruskotter JT. Predators and the public trust. *Biological Reviews*. 2017; 92:248-70.
39. Wydeven AP, Mladenoff DJ, Sickley TA, Kohn BE, Thiel RP, Hansen JL. Road density as a factor in habitat selection by wolves and other carnivores in the Great Lakes Region. *Endangered Species Update*. 2001;18:110-4.
40. Wydeven AP, Treves A, Brost B, Wiedenhoef JE. Characteristics of wolf packs in Wisconsin: Identification of traits influencing depredation. In: Fascione N, Delach A, Smith ME, editors. *People and Predators: From Conflict to Coexistence*. Washington, D. C.: Island Press; 2004. p. 28-50.
41. Wydeven AP, Wiedenhoef J, Schultz RN, Thiel RP, Jurewicz RR, Kohn B, et al. History, population growth and management of wolves in Wisconsin. In: Wydeven AP, Van Deelen TR, Heske EJ, editors. *Recovery of Gray Wolves in the Great Lakes Region of the United States: an Endangered Species Success Story*. New York: Springer; 2009. p. 87-106.
42. Wydeven AP, Schultz RN, Thiel RP. Monitoring a recovering gray wolf population in Wisconsin, 1979-1995. In: Carbyn LN, Fritts SH, Seip DR, editors. *Ecology and Conservation of Wolves in a Changing World*. Edmonton, Alberta, Canada.: Canadian Circumpolar Institute; 1995. p. 147-56.
43. Wydeven AP, Wiedenhoef JE. Status of the timber wolf in Wisconsin: Performance report 1 July 2002 through 30 June 2003. In: Resources WE, editor. *Report #127: Wisconsin Department of Natural Resources*; 2003.

44. Wiedenhoeft JE, Boles SR, Wydeven AP, editors. A Volunteer Carnivore Tracking Program and its Potential Use in Monitoring the Timber Wolf (*Canis lupus*) Population in Northern and Central Wisconsin. World Wolf Congress 2003: Bridging Science and Community; 2003; Banff, Alberta, Canada.
45. Wydeven AP, Wiedenhoeft JE, Schultz RL, Thiel RP, Boles SH, Heilhecker E, et al. Progress Report of Wolf Population Monitoring in Wisconsin for the Period October - March 2004. In: Resources WDoN, editor. Park Falls, Wisconsin PUB-ER- 2004: Wisconsin Department of Natural Resources; 2004.
46. Wydeven AP, Wiedenhoeft JE, Schultz RL, Thiel RP, Boles SH, Heilhecker E. Progress Report of Wolf Population Monitoring in Wisconsin for the Period October 2005 - March 2006. In: Resources WDoN, editor. Park Falls, Wisconsin PUB-ER- 2006: Wisconsin Department of Natural Resources; 2006.
47. Treves A, Paquet PC, Artelle KA, Cornman AM, Krofel M, Darimont CT. Transparency about values and assertions of fact in natural resource management. *Frontiers in Conservation Science: Human-Wildlife Dynamics*. 2021;2:e631998.
48. Treves A, Martin KA, Wiedenhoeft JE, Wydeven AP. Dispersal of gray wolves in the Great Lakes region. In: Wydeven AP, Van Deelen TR, Heske EJ, editors. *Recovery of Gray Wolves in the Great Lakes Region of the United States: An Endangered Species Success Story*. New York: Springer; 2009. p. 191-204.
49. Borg BL, Brainerd SM, Meier TJ, Prugh LR. Impacts of breeder loss on social structure, reproduction and population growth in a social canid. *Journal of Animal Ecology* 2015;84(1):177-87.
50. Borg BL, Arthur SM, Broman NA, Cassidy KA, McIntyre R, Smith DW, et al. Implications of Harvest on the Boundaries of Protected Areas for Large Carnivore Viewing Opportunities. *PLoS ONE*. 2016;11(4):e0153808. doi:10.1371/journal.pone.
51. Brainerd SM, Henrik Andrén, Edward E. Bangs, Elizabeth H. Bradley, Joseph A. Fontaine, Wayne Hall, et al. The effects of breeder loss on wolves. *Journal of Wildlife Management*. 2008;72(1):89-98.
52. Haber GC. Biological, conservation, and ethical implications of exploiting and controlling wolves. *Conserv Biol*. 1996;10:1068-81.
53. Santiago-Avila FJ, Cornman AM, Treves A. Killing wolves to prevent predation on livestock may protect one farm but harm neighbors. *PLoS One*. 2018;13(1):e0189729
54. Natural Resources Board. Request approval of the fall 2021 wolf season harvest quota. In: Resources WDoN, editor. Madison, WI: Wisconsin Department of Natural Resources; 2021.
55. WDNR. Wisconsin gray wolf monitoring report 15 April 2020 through 14 April 2021. In: WI Department of Natural Resources BoWM, editor. Madison, WI, USA2021.
56. Sanders JD. Wolves, lone and pack: Ojibwe treaty rights and the Wisconsin wolf hunt. *Wisconsin Law Review*. 2013;2013:1263-94.
57. Ausband DE, Stansbury CR, Stenglein J, Struthers JL, Waits LP. Recruitment in a social carnivore before and after harvest. *Anim Conserv*. 2015;18(5): 415-23.
58. Ausband DE. Gray wolf harvest in Idaho. *Wildl Soc Bull*. 2016;40(3):500-5.
59. Barber-Meyer SM, Wheeldon TJ, Mech LD. The importance of wilderness to wolf (*Canis lupus*) survival and cause-specific mortality over 50 years. *Biol Conserv*. 2021;258:109145.
60. Hinton JW, Brzeski KE, Rabon DR, Chamberlain MJ. Effects of anthropogenic mortality on critically endangered red wolf *Canis rufus* breeding pairs: implications for red wolf recovery. *Oryx*. 2016; doi:10.1017/S0030605315000770.
61. Sparkman AM, Waits LP, Murray DL. Social and demographic effects of anthropogenic mortality: A test of the compensatory mortality hypothesis in the red wolf. *PLoS ONE*. 2011;6(6):1-9.
62. Agan SW, Treves A, Willey L. Wild red wolf *Canis rufus* poaching risk. *bioRxiv*. 2020.



63. Santiago-Ávila FJ, Treves A. Poaching of protected wolves fluctuated seasonally and with non-wolf hunting. *Scientific Reports*. 2022;12:e1738
64. Laver PN, Kelly MJ. A critical review of home range studies. *The Journal of Wildlife Management*. 2008;72(1): 290-8.\.
65. Santiago-Ávila FJ. Muddled Facts and Values: Positivism, Egoism, and Anthropocentrism in the Anthropocene. *Animals and Society*. 2020.
66. Treves A, Santiago-Ávila FJ. Myths and assumptions about human-wildlife conflict and coexistence. *Conserv Biol*. 2020;34(4):811–8.
67. Oreskes N. *Why Trust Science?* Princeton, NJ: Princeton University Press; 2019.
68. Karns GR, Heeren A, Toman EL, Wilson RS, Szarek HK, J.T. B. Should Grizzly Bears Be Hunted or Protected? Social and Organizational Affiliations Influence Scientific Judgments. *Canadian Wildlife Biology & Management*. 2018;7(1):18-30.
69. Treves A. Scientific ethics and the illusion of naïve objectivity. *Front Ecol Environ*. 2019;7:361.