Drivers of taxonomic bias in conservation research: A global analysis of terrestrial mammals

Authors: Janisson W. dos Santos[1*], Ricardo A. Correia[1,2,3], Ana C. M. Malhado[1], João V. Campos-Silva[1,4], Davi Teles[1], Paul Jepson[5], Richard J. Ladle[1]

[1]Institute of Biological Science and Health, Federal University of Alagoas, Campus A. C. Simões, Maceió, Alagoas, Brazil.

[2] Helsinki Lab of Interdisciplinary Conservation Science, Department of Geosciences and Geography, University of Helsinki, Finland.

[3]Helsinki Institute for Sustainability Science, University of Helsinki, Finland.

[4]Faculty of Environmental Sciences and Natural Resource Management, Norwegian. University of Life Sciences, Ås, Norway.

[5]Ecosulis, Paul Street, London, England.

[*]Corresponding author: janisson.santos@icbs.ufal.br

Abstract

Scientific knowledge of species and the ecosystems they inhabit is the cornerstone of modern conservation. However, research effort is not spread evenly among taxa (taxonomic bias), which may constrain capacity to identify conservation risk and to implement effective responses. Addressing such biases requires an understanding of factors that promote or constrain the use of a particular species in research projects. To this end, we quantified conservation science knowledge of the world's extant non-marine mammal species (n=4,108) based on the number of published documents in journals indexed on Clarivate Analytics' *Web of Science*™. We use an innovative hurdle model approach to assess the relative importance of several ecological, biogeographical and cultural factors for explaining variation in research production between species. The most important variable explaining the presence/absence of conservation research was scientific capacity of countries within the range of the species, followed by body mass and years since the taxonomic description. Research volume (more than one document) was strongly associated number of years since the data describing on that species, followed by scientific capacity within the range of species, high body mass and invasiveness. The threat-status was weakly associated to explain the presence/absence and research volume in conservation research. These results can be interpreted as a consequence of the dynamic interplay between the perceived need for conservation research about a species and its appropriateness as a target of research. As anticipated, the scientific capacity of the countries where a species is found is a strong driver of conservation research bias, reflecting the high variation in conservation research funding and human resources between countries. Our study suggests that this bias could be most effectively reduced by a combination of investing in pioneering research, targeted funding and supporting research in countries with low scientific capacity and high biodiversity.

Key-words: Taxonomic Bias, Knowledge Production, Biodiversity Conservation, Mammals

Introduction

'Saving' species from extinction is a central objective of the global conservation movement and a focal point for conservation actions (Adams, 2013). Success in this endeavour requires at least three general conditions to be fulfilled: i) species need to be described and identified as being at risk of extinction, through processes such as the IUCN's Red List assessment (Rodrigues *et al.*, 2006); ii) there should be sufficient biological, ecological and cultural knowledge of the species to support the design and implementation of appropriate conservation interventions (Cooke *et al.*, 2017; Murray *et al.*, 2015; Sutherland *et al.*, 2004), and; iii) conservation groups with the technical capacity, financial resources and willingness to intervene should be present within the geographic region occupied by the species (Ladle and Jepson, 2008). Scientific knowledge is central to the first two conditions and is often a fundamental component of effective conservation actions (Sutherland *et al*., 2004). Of course, more knowledge does not always lead to better conservation or swifter action, but *ceteris paribus* adequately studied species are more likely to be the recipients of effective conservation actions.

It is well known that scientific knowledge of species is extremely patchy, both taxonomically and spatially (Clark and May, 2002; Fleming and Bateman, 2016; Murray *et al.*, 2015; Meyer *et al*., 2015) with potentially serious consequences for conservation. For example, even if a species is identified as being threatened, a lack of scientific knowledge can seriously impede the development of effective conservation interventions. The importance of scientific knowledge is reflected in Aichi Target 19, that identifies the improvement of "scientific knowledge about biodiversity and its applicability in decision-making" as a key enabling condition for the development of the Strategic Plan for Biodiversity (Marques *et al*., 2014). Scientific knowledge also has a broader role in conservation, helping the public to understand the need for protection and why certain policies (e.g. eradication of invasive species) are favoured over others (Dreyfus, 1995).

The reasons for the extreme patchiness of scientific knowledge about species are complex, reflecting factors such as an unequal allocation of resources, spatial and temporal variation in research capacity, and the intrinsic characteristics of a species that makes it an convenient target for a particular type of research project (Clark and May, 2002). In this context, 'appropriateness' reflects both the extent of existing knowledge (both generally and specific to the individual/research group) and the difficulty of collecting new data. This latter characteristic is, in turn, dependent upon the ecological characteristics of the species and their geographical distribution.

Some of the factors that could influence whether a given species is the subject of research (e.g. cultural preferences, availability of local research funding, research history, etc.) vary enormously in time and space and are therefore difficult to systematically assess at a global level. However, other factors (e.g. country level research capacity, species range size, etc.) should be more temporally stable and, for that reason, are likely to be the main drivers of the observed systematic taxonomic biases in conservation research (Clark and May, 2002). For example, all things being equal, we would predict an endemic species in a country with low scientific capacity to be less studied (be the subject of fewer scientific articles) than an ecologically equivalent (e.g. in terms of body size, range size, habitat, etc.) endemic species in a more scientifically developed country. This is because: i) scientists tend to study species within the country where they work due to a combination of funding priorities, cost and practical convenience, and; ii) countries with low scientific capacity probably have fewer qualified scientists and less resources available for research. Thus, we would predict a strong influence of geography on taxonomic bias in conservation research effort (Fisher *et al*., 2011; Meyer *et al*., 2015; Hortal *et al*., 2016).

Another group of systematic biases is associated with the 'researchability' of a species, defined here as any characteristic of the species that potentially increases the costs of data collection or which impedes or reduces the feasibility of a research project. For field-based conservation research this includes any characteristics that make a species more difficult to observe, such as small body size, nocturnal activity patterns (Chetana and Ganesh, 2007), elusiveness (Lampa *et al*., 2015) or cryptic coloration (Vine *et al*., 2009). Such issues may be particularly problematic for academics whose career advancement strongly depends on their publication records or students who need to meet dissertation requirements (Caro, 2007), and could conceivably act as a disincentive to choose certain species as the subjects of a research project. Moreover, researchability may also be influenced by geographical factors such as range size or remoteness (Ladle *et al*., 2011) since these can considerably increase research costs and feasibility (depending on resources and technical equipment requirements). The importance of some of these systematic biases has been well studied in relation to the collection of biological samples, whose distribution is often highly correlated with the presence of roads or proximity to research centres (e.g. Reddy and Dávalos, 2003; Kuper *et al*., 2006; Stropp *et al*., 2016). More recently, a regional scale bibliometric analysis of Australian birds showed that there were significantly more publications on species with larger body sizes, larger ranges, higher relative abundance, and which can be found in urban environments (Yarwood *et al*., 2019).

Finally, given that conservation science is both globalized and mission-orientated (Jepson and Canney, 2003), we would also predict that conservation scientists around the world would also respond to conservation need (as indicated by global conservation priority classifications). At a species level, the most commonly used prioritization system is the IUCN Red List of endangered species (Rodrigues *et al*., 2006) which classifies species into endangerment categories based on a combination of demographic and geographic characteristics. We might, therefore, predict that individual researchers and funding agencies (national and international) might respond to this categorization by prioritizing research on endangered species (Rodrigues *et al*., 2006). It should be noted that endangered species may also be among the least 'researchable', since they are by definition often difficult to locate, observe and study (Pawar, 2003). These conflicting drivers may explain why a recent bibliometric analysis of felids and canids failed to find any influence of conservation status on the volume of published conservation research (Tensen, 2018).

In summary, it is clear that various factors influence taxonomic bias in research and that perceived conservation need may not always be the overriding priority when a conservation researcher chooses to work on a particular species, leading to a potential mismatch between what species are actually being studied and what species we most need to know about. Here, we explore this issue by developing the first quantitative model of global conservation science knowledge for non-marine terrestrial mammal species. We chose terrestrial mammals because they are large and highly culturally visible taxon whose species vary considerably in ecological and biogeographical attributes. Moreover, research on mammal conservation has received more attention from researchers in comparison to other vertebrate groups, although this attention is not evenly distributed among taxa (Clark and May, 2002). Specifically, we use our model to quantify the relative importance of factors associated with conservation need (e.g. threat status, endemism) and the more prosaic and pragmatic factors that make some species easy and cheaper to research (e.g. large range size, diurnal behaviour, etc.).

Materials and Methods

Database

We originally considered all non-marine mammal species present in the IUCN Red List (version 2017.1). For each of the 5,346 mammal species on this list, we aimed to collect information on the currently accepted scientific names and any synonyms to guarantee the adequate retrieval of information available in digital databases (Correia *et al.*, 2017; Correia *et al*. 2018; Remsen, 2016). However, we were unable to identify one or more explanatory variables (see below) for 1,238 species, 734 of which are classified as Data-Deficient by the IUCN Red List. These species were consequently excluded from our final analysis, which considered a total of 4,108 non-marine mammal species.

Dependent Variable

We quantified the conservation science research effort for each mammal species in our database through the number of scientific publications (including research articles, reviews, notes, book chapters, and other peer reviewed documents) indexed by Clarivate Analytics' Web of Science (WoS) platform. We attributed published documents to species by searching for scientific names and any known scientific synonyms (e.g. "*Mus musculus*" OR *"Mus domesticus"*) in a topic search (covering titles, abstracts and keywords). We searched using scientific names because we reasoned that if a species name appears in the title, abstract or keywords then it is likely that the publication contains significant information about that species. Clearly, a proportion of published documents will mention the species name and little relevant information while other documents may be relevant, but not be captured by our search terms. While this reduces the precision of the results, our search method is replicable and should be taxonomically unbiased, allowing for the identification of broad-scale patterns.

Data were collected between January and April of 2017 and the number of documents published between 1945 and 2016 returned by each search were recorded. After this, we filtered results for WoS's "Biodiversity Conservation" topic (excluding documents that did not appear in conservation-themed journals), and used this as our metric of conservation-relevant knowledge.

It should also be noted that the published documents in our study represent only a proportion of the research conducted for any given project, and that many research projects may never generate a peer-reviewed publication. Of course, there are many reasons that a conservation scientist may not publish, including: i) insufficient evidence (e.g. observations of a rare species) to construct a convincing narrative; ii) lack of significant results; iii) research that is too local or descriptive to be easily published, and; iv) lack of capacity and/or interest on the part of the project team. Some of this information ends up in non-peer reviewed scientific products such as undergraduate theses and expedition reports, and much of it ends up in the file drawers and computers of scientists. Many of the above factors are more likely to be associated with a rare/threatened species potentially pushing conservation scientists to choose study species that have greater potential for generating a publication (Caro, 2007).

Explanatory variables

To better understand the factors influencing variability in conservation research between species, we identified two main factors that could influence taxonomic bias in conservation research on mammals:

Conservation need: researchers may respond to perceived conservation need, such as species identified as at risk of extinction (Rodrigues *et al.*, 2006), threats to the existence of other species (Clavero and García-Berthou, 2005), or the intrinsic value of their evolutionary distinctiveness (Isaac *et al.*, 2007; Jetz *et al.*, 2014);

Researchability: Some mammal species are easier than others to find, observe, manipulate and write about due to: i) intrinsic characteristics such as body size, diurnality, habitat use and population density (Ladle *et al.*, 2011) and; ii) geographic factors that are extrinsic to the species, such as the overlap between the distribution of scientists and that of the species they study (Meyer *et al*., 2015), i.e. a species may have intrinsic characteristics that facilitate research, but there may be limited local capacity to take advantage of this. At an international level we would predict that species in countries with high conservation science capacity would be more studied than those distributed in countries with lower capacity (Fisher *et al*., 2011). Finally, science is iterative, and we would therefore expect that *a priori* knowledge of a species (e.g. volume of historical research) will facilitate the development of innovative science which may be more easily published in peer-reviewed journals.

We represented these factors in our model with the following proxy variables (See Table 1 for details): i) **Conservation need**: conservation threat status (Baillie *et al.*, 2004); introduced species; evolutionary distinctiveness, which is a measure of species exclusivity; ii) **Researchability:** range size (log10 + 1), environmental science capacity within the countries where the species is present, nocturnal habit, body mass (log10 + 1), years since species description (Table 1).

Data analysis

We explored the relationship between the different explanatory variables and research productivity at the species level using a hurdle model analysis for zero-inflated count data (Zeileis *et al.*, 2008). Hurdle models are two-component models composed of a zero-hurdle component (henceforth Zero-hurdle model) that models the probability of counts being zero or not, and a truncated count component (henceforth Count model) that is applied to positive counts (i.e. those > 0). This modelling approach was chosen due to the high number of species without any recorded study. This approach is not only more adequate to model zero-inflated data than standard Generalized Linear Models, it also allows for modelling the effect of each explanatory variable on both the presence or absence of research on mammals and the amount of research for each species with at least one scientific product.

Due to the large number of variables than can plausibly influence a scientist's decision to work on a particular species, it is unlikely that a single model can accurately represent such a complex decision-making process. We therefore decided to adopt a multi-model inference approach, which allows us to calculate a weighted-average estimate of the effect of each explanatory variable based on the most plausible hypothesis explaining the decision process (Burnham and Anderson, 2004; Burnham *et al.*, 2011). Hence, we calculated all possible model combinations considering our set of explanatory variables and identified the set of most plausible models according to AIC corrected for small sample size (AICc) and considered all models with a ΔAICc ≤ 4 in relation to the best model (Table S1) for a conditional-model averaging process. Each continuous variable was standardized before inclusion in the models (Schielzeth, 2010), so that their relative effect size could be considered a measure of relative importance explaining species-level scientific interest.

All model assumptions were tested prior to analysis (Zuur *et al.*, 2010) and variable multicollinearity was assessed; we found no evidence that assumptions were not met and no evidence of strong correlation (Spearman’s correlation; $r \leq |0.7|$) between variables. Hurdle regression models were implemented using the function ‘hurdle’ of the package ‘pscl’ and every model combination examined with the ‘MuMIn’ package (Barton, 2009) within the R platform (R Core Team, 2013).

Results

Our searches on WoS for the scientific names and synonyms of 4,108 non-marine mammal species resulted in a total of 95,420 published documents in journals in the Biodiversity Conservation area. Approximately 20% of these documents were associated with the 10 most-researched mammal species; *Sus scrofa* (wild boar), *Odocoileus virginianus* (white-tailed deer), *Cervus elaphus* (red deer), *Canis lupus* (grey wolf), *Vulpes vulpes* (red fox), *Alces alces* (moose), *Loxodonta africana* (African elephant), *Odocoileus hemionus* (mule deer), *Rangifer tarandus* (reindeer) and *Ursus arctos* (brown bear), respectively (Figure 1).

In contrast, almost 76% of the studied species were associated with 10 documents or less, representing about 8% of all documents. That is, almost a quarter of species studied were associated with about 92% of all documents. Moreover, approximately a quarter of species were not have any document in the WoS database. At the order level, approximately 99% of published documents were associated with species belonging to less than half of extant mammalian orders (Fig. S1). Species in the three most studied orders, *Cetartiodactyla*, *Carnivora* and *Rodentia* were associated with 70% of all documents. Note, some documents on *Sus scrofa* relate to work on domestic pigs (*Sus scrofa domesticus)*, since these occasionally relevant in conservation-related studies.

Our hurdle analysis clearly shows that even though we focused on conservation-related articles, variables representing ‘researchability’ were the most important determinants of whether a mammal species had any associated articles in our database. Specifically, scientific capacity of countries within the range of a species was the most important variable explaining the presence/absence of conservation research. Body mass and years since taxonomic description were also associated with species with one or more associated document. ‘Conservation need’ as measured by threat-status was only weakly associated with research effort, while evolutionary distinctiveness and nocturnality had no relationship with presence/absence of published research (Fig. 2). The results of the most parsimonious hurdle models reinforce the findings that threat status, evolutionary distinctiveness and nocturnality have a negligible influence on whether a species has been the subject of published research (Table 2). Invasiveness was not included in the Zero-hurdle part of the analysis because all species with this characteristic were associated with at least one published document in the database.

For species that had one or more associated scientific documents, the average of most parsimonious models (Table S1) indicates that all variables have, to a greater or lesser degree, a significant influence. The most important variable explaining the volume of scientific documents (more than one document) was the number of years since the data describing on that species (Fig. 2, Table 2). Scientific capacity of range countries, high body mass and invasiveness also had a strong positive association with the number of scientific documents. Nocturnality, threat status and range size were weakly associated with research volume and evolutionary distinctiveness had a negative association.

Discussion

Our most general finding is that conservation research on mammals shows dramatic taxonomic biases, broadly confirming the conclusions of previous studies (Clark and May, 2002; Donaldson *et al.*, 2016; Fazey *et al.*, 2005a; Tensen, 2018). More than a quarter of species in our database were associated with few or no published documents on WoS. While this is probably an accurate and relatively unbiased reflection of the relative taxonomic distribution of conservation research on mammals, it is important to acknowledge that our metric does not capture all conservation knowledge. There is a wealth of information in the grey literature and in non-text sources, although we would argue that, *ceteris paribus*, there is likely to be a strong correlation between the volume of published and unpublished literature about a given species (De Lima *et al*., 2011). Similarly, recent studies have shown strong and consistent correlations between the frequency of use of species vernacular and scientific names on the internet, in newspapers and on social media networks (Jarić *et al*., 2016; Correia *et al*., 2017), even though the latter are mainly restricted to technical documents.

The reasons for such a highly skewed distribution of conservation research are undoubtedly both complex and interacting. Scientists might be actively avoiding working on rare and understudied species. Limited resources (Wilson *et al*., 2006) and pressure to publish could encourage risk-averse behaviour of conservation scientists and funders, who may be unwilling to invest in the development of new study systems. For example, Tim Caro recently observed a growing tendency of graduate students studying animal behavior to work on common species that are considered to be ecologically similar to a species of conservation concern (Caro, 2017). Caro attributes this trend to the fact that rare species are “difficult to locate and result in small sample sizes” (Caro, 2017), which presumably leads to poorly substantiated studies that are difficult to publish. Such risk-averseness may have contributed to the large number of studies on introduced species (which are often abundant and easy to study) in our database. More broadly, there may often be a conflict between what needs to be studied (because it is endangered) and the career aspirations of the researcher who may need to publish in prestigious journals.

Another factor that could potentially increase taxonomic bias is geographic biases in research capacity. Indeed, environmental science capacity of countries within the range of a species was strongly associated with research effort for both components of our Hurdle model. This is most simply explained as a consequence of conveniently located study populations overlapping with a qualified ‘corpus’ of conservation researchers (Fazey *et al.*, 2005b; Meyer *et al.*, 2015; Ibáñez-Álamo *et al.*, 2017). Such a consequence inevitably leads to a mismatch between conservation research effort and conservation research need which is higher in the world's most biodiverse countries in the global south (Fisher *et al.*, 2011). This finding parallels several studies that have shown a strong geographic correlation between the presence of a research centres and a high density of biological records and conservation research (e.g. Amano and Sutherland, 2013; Engemann *et al.*, 2015; Ibáñez-Álamo *et al.*, 2017; Schulman *et al.*, 2007; Lessa *et al.*, 2019; Correia *et al.*, 2019). Such geographical biases in research are likely to be reduced in the future if few research capacity countries invest greater amounts of resources in science (Fazey *et al.*, 2005b) and consequently insert more conservation qualified researchers in areas with low research capacity However, it is unlikely that such biases will ever be eliminated given our finding that the number of years since the first published study was strongly correlated with research volume. This result reflects the iterative nature of scientific research, with previous studies providing context, baselines and inspiration for future studies (dos Santos *et al.*, 2015). In other words, the more a species is researched, the more it will be researched.

Body size was also strongly associated with both presence and volume of conservation research. That larger species are frequently more studied has previously been noted (Ibáñez-Álamo *et al.*, 2017; Tensen, 2018), and may be related to their higher cultural profile (Frynta *et al.*, 2013; Jepson and Barua, 2015; Macdonald *et al.*, 2015; Ladle *et al.*, 2019), and that they are more likely to be hunted, have lower population densities, slower life histories and, consequently, to be at greater risk of extinction (Schipper *et al.*, 2008). Additionally, large species are often more conspicuous and may be easier to study *in situ*. They also appear to attract more attention to both scientists and citizens, and thus can be used to mobilize resources for research and conservation (Brodie, 2009; Frynta *et al*., 2013).

Another of our results was the strong association between the time since a species was scientifically named and conservation research volume. This may be related to the contrasting biocultural traits of the first mammals to be described in comparison to more recently discovered species. The former tended to be from Europe where most of the early taxonomists lived and worked, or were sufficiently impressive or culturally important to have come to the attention of these taxonomists.

From a conservation perspective, the association between threat category and presence and volume of scientific documents suggests that conservation science research is responding, albeit weakly, conservation need. This is especially encouraging given that endangered species will frequently be more difficult to study due to low densities and population sizes, and because their study may entail additional bureaucratic hurdles (Berenbaum, 2008; Strier and Mendes, 2009). Our results suggest the act of listing (e.g. IUCN Red Lists, EDGE or CITES appendices) may provide scientists with additional justifications for engaging in new research projects on a species.

The above result is at variance with a recent study on European birds that concluded that "research effort was not well targeted with respect to either European or global threat status" (Murray *et al.*, 2015, p. 193). Likewise, Amori and Gippoliti (2000) analyzed the scientific articles present in four important international conservation journals (*Oryx*, *Conservation Biology*, *Biological Conservation* and *Biodiversity and Conservation*) and concluded that there was a lower research effort associated with more threatened species of mammals. A study on British breeding birds also found that species with declining range size were less studied based on ecology publications (McKenzie and Robertson, 2015). For Canidae (Tensen, 2018) and Felidae (Brodie, 2013) families, threat status also had no significance in relation to other variables in the search allocation effect, such as body mass. However, the conservation-focused research appears to target endangered island endemic bats, although there was no greater research attention with the increased risk of extinction of these species (Conenna *et al*., 2017). These discrepancies are potentially caused by the smaller taxonomic or geographic scale of some of the studies and the different ways of measuring research effort.

It is important to reiterate that there are a number of factors that may significantly influence conservation research on mammals, but were not included in our model because they are either: i) locally important, but are expected to have little influence at a global level, or ii) are difficult to systematically quantify. A possible example of the former is national level funding priorities that target certain endangered or iconic species. An example of the latter are traits associated with species charisma (Lorimer, 2007) or aesthetic appeal (Lišková and Frynta, 2013). Species with such traits often benefit from increased public interest, making them excellent candidates for flagship species or as the subject of conservation fundraising campaigns (Clucas *et al*., 2008; Jepson and Barua, 2015). Interestingly, charismatic species may also be highly threatened, possibly because the public are so familiar with representations of these species that they assume that they must have healthy populations (Courchamp *et al*., 2018). However, aesthetic appeal cannot be easily quantified at scale, although this may soon change with the development of increasingly sophisticated tools to quantify different dimensions of human interest in wild species and nature (*cf*. Ladle *et al*., 2016).

Species charisma is not the only driver of human interest in non-human species, and another factor that could influence research effort is their degree of similarity (physical or otherwise) with humans. Such anthropomorphism, in addition to promoting empathy with non-human species (Chan, 2012) could also act to encourage research. Moreover, while anthropomorphism itself is hard to systematically quantify, a recent social survey found that empathy towards a variety of non-human species was inversely related to evolutionary divergence times from the human lineage (Miralles *et al*., 2019), potentially opening a path to incorporate a broad proxy of anthropomorphism/empathy into macroscale studies of human interest in nature. It should noted, however, that while charisma and anthropomorphic traits clearly relate to human interest, their impact on research may be much less marked. This is supported by a recent study by Troudet *et al*. (2017) who showed that societal preferences (as measured by internet searches) were a much better predictor of taxonomic bias in biodiversity information (measured by GBIF records) than was research effort.

Finally, the incremental nature of scientific research (De Silva and Vance, 2017) may mean that a species that has already been well researched becomes a `better` subject for future research. Such positive feedback could, over time, act to increase inequalities between species in terms of research effort and publications. If such an effect is operating, it places exceedingly high value on pioneer research, which can form the basis for future, more sophisticated research. Interestingly, there is good evidence that pioneer research also boosts research effort in geographic regions (Dos Santos *et al.*, 2015) and in protected areas (Correia *et al.*, 2016).

Conclusions

Most conservation scientists would agree that choice of research organism is of fundamental importance, influencing research and conservation outcomes, societal relevance, future funding opportunities and even personal motivation and job satisfaction. Nevertheless, such choices are also strongly constrained by professional requirements for high impact research, accessing existing funding streams and practical considerations such as access to conveniently situated field sites. Not only does this lead to the well-known pattern of taxonomic bias in conservation research (Clark and May, 2002), it strongly suggests that such bias is structural and will not be easily remedied. Well-studied species will continue to be the best models for sophisticated research requiring international journals. Thus, additional incentives are required for species that are poorly researched and largely ignored by researchers. Our research indicates that these species are typically small, present in countries with low scientific capacity, have restricted geographic distributions, have not been introduced elsewhere, and have often been described recently and are evolutionarily distinct. This highlights the importance of increasing dedicated incentives to work on poorly known species (e.g. dedicated funding streams, sympathetic journal editors, changes in evaluation systems for researchers, etc). Such incentives have added importance given that new species discoveries and taxonomic revisions are likely to add to the global total of poorly known species and gradually fill the knowledge gaps over time (Hortal *et al.*, 2015). In addition to ensuring dedicated funding streams for poorly known taxa, it will also be important invest in ecological surveying and taxonomy which, while unlikely to generate many high impact publications, will produce invaluable baseline data for conservation decision making and provide a start point for future studies.

Acknowledgments

The authors would like to thank Bruno Umbelino and Thainá Lessa for discussions. This work was funded by the Brazilian National Council for Scientific and Technological Development CNPq with research grants to PJ-RJL (#400325/2014-4 and #448966/2014-0). JWS´s dissertation was funded by Coordination for the Improvement of Higher Education Personnel CAPES-Brazil. RAC was initially supported by a CNPq post-doctoral grant (#158841/2015-8) and is currently funded by post-doctoral scholarship from FCT – Fundação para a Ciência e Tecnologia (SFRH/BPD/118635/2016). RJL and ACMM are funded by CNPq (#309879/2019-1; #309980/2018-6). JVC-S is funded by a CAPES post-doctoral grant (Ref. 1666302).

Author Contributions: Conceived and designed the study: RJL ACCM JWS RAC JVC-S. Collected the dataset: JWS. Analysed the data: JWS JVC-S DT RAC. Wrote the article: JWS RJL RAC ACMM JVC-S DT PJ.

References

Adams, W. B. (2013). *Against extinction: The story of conservation*. Oxford: Routledge.

Amano, T., Sutherland, W. J. (2013). Four barriers to the global understanding of biodiversity conservation: Wealth, language, geographical location and security. *Proceedings of the Royal Society B* 280, 20122649.

Amori, G., Gippoliti, S. (2000). What do mammalogists want to save? Ten years of mammalian conservation biology. *Biodiversity & Conservation* 9, 785.

Baillie, J., Hilton-Taylor, C., Stuart, S. N. (Eds.) (2004). *Iucn red list of threatened species: A global species assessment*. Gland, Switzerland: IUCN.

Barton, K. (2009). Mumin: Multi-model inference. R package version 1. 0. 0. *http://r-forge. r-project. org/projects/mumin/*.

Berenbaum, M. (2008). Insect conservation and the entomological society of america. *American Entomologist* 54, 117.

Brodie, J. F. (2009). Is research effort allocated efficiently for conservation? Felidae as a global case study. *Biodiversity and Conservation* 18, 2927.

Burnham, K. P., Anderson, D. R. (2004). Multimodel inference: Understanding aic and bic in model selection. *Sociological methods & research* 33, 261.

Burnham, K. P., Anderson, D. R., Huyvaert, K. P. (2011). Aic model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology* 65, 23.

Caro, T. (2007). Behavior and conservation: a bridge too far? *Trends in Ecology & Evolution* 22, 394.

Cassey, P., Blackburn, T. M. (2003). Publication rejection among ecologists. *Trends in Ecology & Evolution* 18, 375.

Chan, A. A. H. (2012). Anthropomorphism as a conservation tool. *Biodiversity and Conservation* 21, 1889.

Chetana, H. C., Ganesh, T. (2007). Survey and activity patterns of nocturnal mammals in a fragmented dry deciduous forest of Karnataka. *Zoo's Print Journal* 22, 2644.

Clark, J. A., May, R. M. (2002). Taxonomic bias in conservation research. *Science* 297, 191.

Clavero, M., García-Berthou, E. (2005). Invasive species are a leading cause of animal extinctions. *Trends in Ecology & Evolution* 20, 110.

Clucas, B., McHugh, K., Caro, T. (2008). Flagship species on covers of US conservation and nature magazines. *Biodiversity and Conservation* 17, 1517.

Colléony, A., Clayton, S., Couvet, D., Jalme, M. S., Prévot, A. (2017). Human preferences for species conservation: Animal charisma trumps endangered status. *Biological conservation* 206, 263.

Conenna, I., Rocha, R., Russo, D., Cabeza, M. (2017). Insular bats and research effort: a review of global patterns and priorities. *Mammal Review* 47, 169. doi:10.1111/mam.12090.

Cooke, S. J., Birnie-Gauvin, K., Lennox, R. J., Taylor, J. J., Rytwinski, T., Rummer, J. L., Franklin, C. E., Bennett, J. R., Haddaway, N. R. (2017). How experimental biology and ecology can support evidence-based decision-making in conservation: Avoiding pitfalls and enabling application. *Conservation Physiology* 5.

Correia, R. A., Jaric, I., Jepson, P., Malhado, A. C. M., Alves, J. A., Ladle, R. J. (2018). Nomenclature instability in species culturomic assessments: Why synonyms matter. *Ecological indicators* 90, 74.

Correia, R. A., Jepson, P., Malhado, A. C. M., Ladle, R. J. (2017). Internet scientific name frequency as an indicator of cultural salience of biodiversity. *Ecological indicators* 78, 549.

Correia, R. A., Malhado, A. C. M., Lins, L., Gamarra, N. C., Bonfim, W. A. G., Valencia-Aguilar, A., Bragagnolo, C., Jepson, P., Ladle, R. J. (2016). The scientific value of amazonian protected areas. *Biodiversity and Conservation* 25, 1503.

Courchamp, F., Jaric, I., Albert, C., Meinard, Y., Ripple, W. J., Chapron, G. (2018). The paradoxical extinction of the most charismatic animals. *Plos Biology* 16, e2003997.

De Lima, R. F., Bird, J. P., & Barlow, J. (2011). Research effort allocation and the conservation of restricted-range island bird species. *Biological Conservation* 144, 627. doi:10.1016/j.biocon.2010.10.021.

De Silva, P. U., Vance, C. K. (2017). Measuring the impact of scientific research. In *Scientific scholarly communication: The changing landscape*: 101-115. De Silva, P. U., Vance, C. K. (Eds.). Cham, Switzerland: Springer.

Donaldson, M. R., Burnett, N. J., Braun, D. C., Suski, C. D., Hinch, S. G., Cooke, S. J., Kerr, J. T. (2016). Taxonomic bias and international biodiversity conservation research. *FACETS* 1, 105.

Dreyfus, A. (1995). Biological knowledge as a prerequisite for the development of values and attitudes. *Journal of Biological Education* 29, 215.

Engemann, K., Enquist, B. J., Sandel, B., Boyle, B., Jørgensen, P. M., Morueta-Holme, N., Peet, R. K., Violle, C., Svenning, J. (2015). Limited sampling hampers “big data” estimation of species richness in a tropical biodiversity hotspot. *Ecology and Evolution* 5, 807.

Fazey, I., Fischer, J., Lindenmayer, J. D. B. (2005a). What do conservation biologists publish? *Biological Conservation* 124, 63.

Fazey, I. *et al*. (2005b). Who does all the research in conservation biology? *Biodiversity & Conservation* 14, 917.

Fisher, R., Radford, B. T., Knowlton, N., Brainard, R. E., Michaelis, F. B., Caley, M. J. (2011). Global mismatch between research effort and conservation needs of tropical coral reefs. *Conservation Letters* 4, 64.

Fleming, P. A., Bateman, P. W. (2016). The good, the bad, and the ugly: Which australian terrestrial mammal species attract most research? *Mammal Review* 46, 241.

Frynta, D., Šimková, O., Lišková, S., Landová, E. (2013). Mammalian collection on noah's ark: The effects of beauty, brain and body size. *PLoS One* 8, e63110.

Haddaway, N. R., Bayliss, H. R. (2015). Shades of grey: Two forms of grey literature important for reviews in conservation. *Biological Conservation* 191, 827.

Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., Ladle, R. J. (2015). Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics* 46, 523.

Ibáñez-Álamo, J. D., Rubio, E., Bitrus Zira, K. (2017). The degree of urbanization of a species affects how intensively it is studied: A global perspective. *Frontiers in Ecology and Evolution* 5, 41.

Isaac, N. J., Turvey, S. T., Collen, B., Waterman, C., Baillie, J. E. (2007). Mammals on the edge: Conservation priorities based on threat and phylogeny. *PloS one* 2, e296.

Jarić, I., Courchamp, F., Gessner, J., Roberts, D. L. (2016). Data mining in conservation research using Latin and vernacular species names. *PeerJ* 4, e2202.

Jepson, P., Canney, S. (2003). Values-led conservation. *Global Ecology and Biogeography* 12, 271.

Jepson, P., Barua, M. (2015). A theory of flagship species action. *Conservation and Society* 13, 95.

Jetz, W., Thomas, G. H., Joy, J. B., Redding, D. W., Hartmann, K., Mooers, A. O. (2014). Global distribution and conservation of evolutionary distinctness in birds. *Current Biology* 24, 919.

Koricheva, J. (2003). Non-significant results in ecology: A burden or a blessing in disguise? *Oikos* 102, 397.

Küper, W., Sommer, J. H., Lovett, J. C., Barthlott, W. (2006). Deficiency in african plant distribution data–missing pieces of the puzzle. *Botanical Journal of the Linnean Society* 150, 355.

Ladle, R., Jepson, P., Malhado, A. C. M., Jennings, S., Barua, M. (2011). The causes and biogeographical significance of species' rediscovery. *Frontiers of Biogeography* 3, 104.

Ladle, R. J., Jepson, P. (2008). Toward a biocultural theory of avoided extinction. *Conservation Letters* 1, 111.

Ladle, R. J., Correia, R. A., Do, Y., Joo, G., Malhado, A. C. M., Proulx, R., Roberge, J., Jepson, P. (2016). Conservation culturomics. *Frontiers in Ecology and the Environment* 14, 269.

Ladle, R. J., Jepson, P., Correia, R. A., Malhado, A. C. M. (2019). A culturomics approach to quantifying the salience of species on the global internet. *People & Nature* 00, 1.

Lampa, S., Mihoub, J-B., Gruber, B., Klenke, R., Henle, K. (2015). Non-invasive genetic mark-recapture as a means to study population sizes and marking behaviour of the elusive Eurasian Otter (*Lutra lutra*). *PloS one* 10, e0125684.

Lessa, T., dos Santos, J. W., Correia, R. A., Ladle, R. J., Malhado, A. C. M. (2019). Known unknowns: Filling the gaps in scientific knowledge production in the Caatinga. *PloS one* 14, e0219359.

Lišková, S., Frynta, D. (2013). What determines bird beauty in human eyes? *Anthrozoös* 26, 27.

Lorimer, J. (2007). Nonhuman charisma. *Environment and Planning D: Society and Space* 25, 911.

Marques, A., Pereira, H. M., Krug, C., Leadley, P. W., Visconti, P., Januchowski-Hartley, S. R., Krug, R. M., Alkemade, R., Bellard, C., Cheung, W. W. L., Christensen, V., Cooper, H. D., Hirsch, T., Hoft, R., van Kolck, J., Newbold, T., Noonan-Mooney, K., Regan, E. C., Rondinini, C., Sumaila, U. R., Teh, L. S. L., Walpole, M. (2014) A framework to identify enabling and urgent actions for the 2020 Aichi Targets. *Basic and Applied Ecology* 15, 633.

McKenzie, A. J. , Robertson, P. A. (2015). Which species are we researching and why? A case study of the ecology of british breeding birds. *PloS one* 10, e0131004.

Meyer, C., Kreft, H., Guralnick, R., Guralnick, R., Jetz, W. (2015) Global priorities for an effective information basis of biodiversity distributions. *Nature Communications* 6, 8221.

Miralles, A., Raymond, M., Lecointre, G. (2019). Empathy and compassion toward other species decrease with evolutionary divergence time. *Scientific Reports* 9, 1-8.

Murray, H. J., Green, E. J., Williams, D. R., Burfield, I. J., Brooke, M. de L. (2015). Is research effort associated with the conservation status of european bird species? *Endangered Species Research* 27, 193.

Pawar, S. (2003). Taxonomic chauvinism and the methodologically challenged. *Bioscience* 53, 861.

R Core Team (2013). *R: A language and environment for statistical computing*.

Reddy, S., Dávalos, L. M. (2003). Geographical sampling bias and its implications for conservation priorities in africa. *Journal of Biogeography* 30, 1719.

Remsen, D. (2016). The use and limits of scientific names in biological informatics. *ZooKeys* 550, 207.

Rodrigues, A. S., Pilgrim, J. D., Lamoreux, J. F., Hoffmann, M., Brooks, T. M. (2006). The value of the IUCN red list for conservation. *Trends in Ecology & Evolution* 21, 71.

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution* 1, 103.

Schipper, J., Chanson, J. S., Chiozza, F., *et al*. (2008). The status of the world's land and marine mammals: Diversity, threat, and knowledge. *Science* 322, 225.

Schulman, L., Toivonen, T. , Ruokolainen, K. (2007). Analysing botanical collecting effort in amazonia and correcting for it in species range estimation. *Journal of Biogeography* 34, 1388.

Strier, K. B. , Mendes, S. L. (2009). Long-term field studies of south american primates. In *South american primates*: 139-155: Springer.

Stropp, J., Ladle, R. J., Malhado, A. C. M., Hortal, J., Gaffuri, J., Temperley, W. H., Olav Skøien, J., Mayaux, P. (2016) Mapping ignorance: 300 years of collecting flowering plants in Africa. *Global Ecology and Biogeography* 25, 1085.

Sutherland, W. J., Pullin, A. S., Dolman, P. M., Knight, T. M. (2004). The need for evidence-based conservation. *Trends in Ecology & Evolution* 19, 305.

Tensen, L. (2018). Biases in wildlife and conservation research, using felids and canids as a case study. *Global Ecology and Conservation* 15, e00423.

Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R., Legendre, F. (2017). Taxonomic bias in biodiversity data and societal preferences. *Scientific Reports* 7, 9132.

Vine, S. J., Crowther, M. S., Lapidge, S. J., Dickman, C. R., Mooney, N., Piggott, M. P., English, A. W. (2009). Comparison of methods to detect rare and cryptic species: a case study using the red fox (*Vulpes vulpes*). *Wildlife Research* 36, 436.

Wilson, K. A., McBride, M., Bode, M., Possingham, H. P. (2006). Prioritizing global conservation efforts. *Nature* 440, 337.

Yarwood, M. R., Weston, M. A., Symonds, M. R. E. (2019). Biological determinants of research effort on Australian birds: a comparative analysis. *Emu-Austral Ornithology* 119, 38.

Zeileis, A., Kleiber, C. , Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software* 27, 1.

Zuur, A. F., Ieno, E. N. , Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution* 1, 3.

Table and Figure legends

Table 1. Explanatory variables used to explain the number of scientific publications on mammals. The table also provide a brief justification of why they were included and the source where they were collected.

Table 2. Results of the Hurdle models relating conservation-themed scientific production to proxy variables representing conservation need and researchability.

Figure 1. Relative volume of conservation-themed published documents for the 10 most studied terrestrial species of mammals.

Figure 2. Coefficient estimates (± 95% confidence intervals) showing the magnitude and direction of effects of different variables on conservation published documents for the Hurdle model analysis. Coefficients are shown for the a) Zero-hurdle model component and the b) Count model component. Blue and red symbols represent positive and negative effects, respectively. Black symbols represent no effect. For full description of predictors, see SI.

Table 1.

Factor	*Variable*	*Source*	*Level*	*Main Assumption*
Conservation Need	*Threat status*	IUCN Red List[1]	Threatened - No threatened	Researchers respond to conservation need by working on threatened species.
	Introduced species	IUCN Red List[1] GISD[3]	Introduced - No introduced	Researchers respond to conservation need by working on species which are a conservation threat.
	Evolutionary distinctiveness	EDGE of Existence[4]	-	Researchers work on more evolutionarily distinct species because they are more important for conserving evolutionary history.
Researchability	*Range size (km^2; log10 + 1)*	IUCN Red List[1]	-	Species with broad geographic ranges are more accessible to a greater number of researchers.
	Years since described	IUCN Red List[1]	-	Species discoveries earlier are less likely to present a broad baseline on which to base additional studies.
	Mean body mass (g; log10 + 1)	Elton traits[5] PanTHERIA[6] EoL[7] Primate Info Net[8] Animal DiversityWeb[9] Mammal Species of the World[10]	-	Larger species are, *ceteris paribus*, easier to observe and collect data on than smaller species.
	Scientific capacity (% global environmental science publications contributed by countries in species' range)	Scimago[11]	-	Countries with higher scientific capacity are likely to have more conservation scientists and expend a greater research effort per (native or introduced) species.
	Nocturnality	Elton traits[5] EoL[7]	Nocturnal - No nocturnal	Nocturnal species are generally more difficult to observe and study than diurnal species.

Data Souces: 1. www.iucnredlist.org/; 2. www.webofknowledge.com; 3. www.iucngisd.org/gisd; 4. www.edgeofexistence.org; 5. http://www.esapubs.org/archive/ecol/E095/178; 6. http://esapubs.org/archive/ecol/e090/184/; 7. www.eol.org; 8. www.pin.primate.wisc.edu; 9. www.animaldiversity.org; 10. https://www.departments.bucknell.edu/biology/resources/msw3; 11. www.scimagojr.com/countryrank.php.

Table 2.

Factor	Proxy Variables	**Zero-hurdle Model** Rel. Importance	**Zero-hurdle Model** Nos. Models	**Count Model** Rel. Importance	**Count Model** Nos. Models
Researchability	Range size	1	5	1	5
Researchasbility	Scientific Capacity	1	5	1	5
Conservation Need	Introduced Species	-	-	1	5
Conservation Need	Threat Status	0.86	4	1	5
Conservation Need	Evol. Distinctiveness	0.74	3	1	5
Researchability	Nocturnality	0.24	2	1	5
Researchability	Years since Described	1	5	1	5
Researchability	Body Mass	1	5	1	5

Figure 1

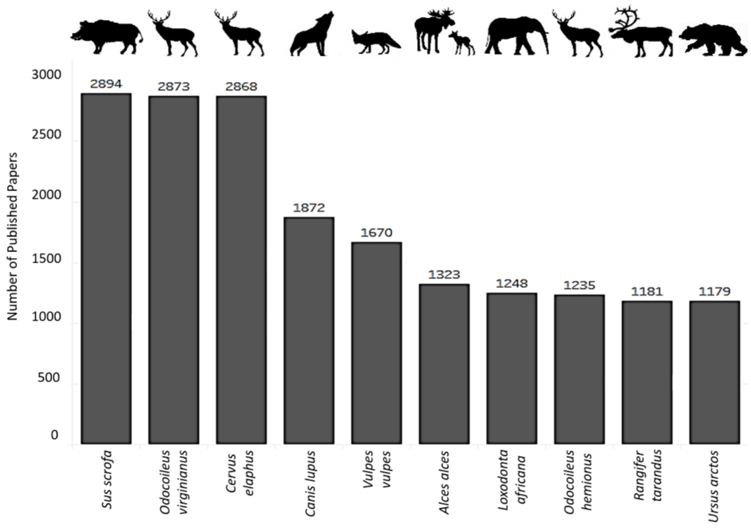
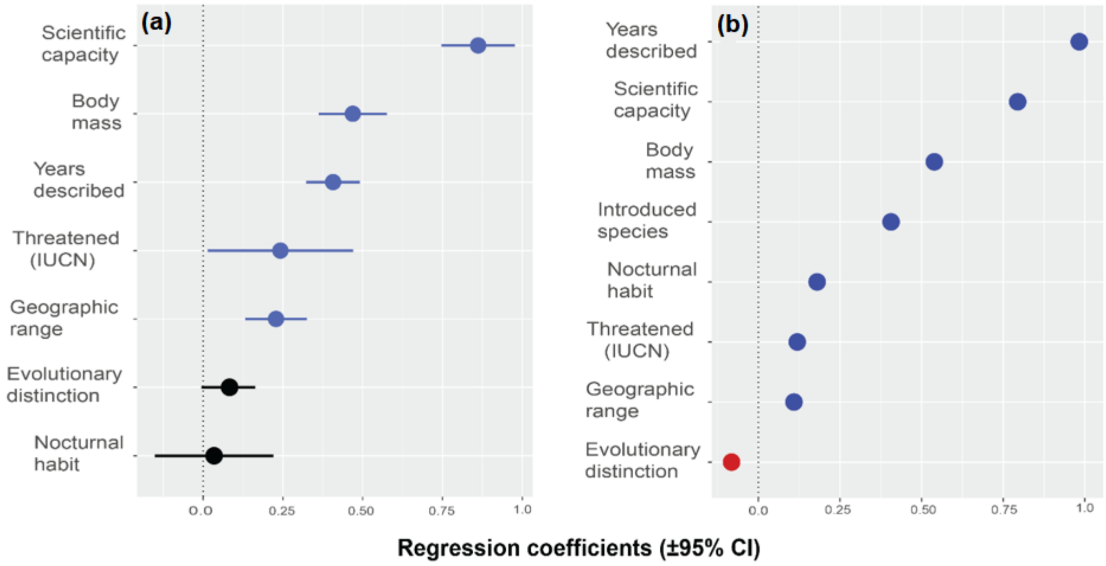

Number of Published Papers
3000
2500
2000
1500
1000
500
0
2894
2873
2868
1872
1670
1323
1248
1235
1181
1179
Sus scrofa
Odocoileus virginianus
Cervus elaphus
Canis lupus
Vulpes vulpes
Alces alces
Loxodonta africana
Odocoileus hemionus
Rangifer tarandus
Ursus arctos

Figure 2


(a)
Scientific capacity
Body mass
Years described
Threatened (IUCN)
Geographic range
Evolutionary distinction
Nocturnal habit
0.0
0.25
0.50
0.75
1.0
(b)
Years described
Scientific capacity
Body mass
Introduced species
Nocturnal habit
Threatened (IUCN)
Geographic range
Evolutionary distinction
0.0
0.25
0.50
0.75
1.0
Regression coefficients (±95% CI)

Supplementary material

Figure S1. Relative value of conservation published documents for the 26 orders of mammals. The colour standards per bar represent the relative values of published documents for the distinct levels of threat. In the legend, "NO_THR" represents the non-threatened species, while the "THR" represents the threatened species.

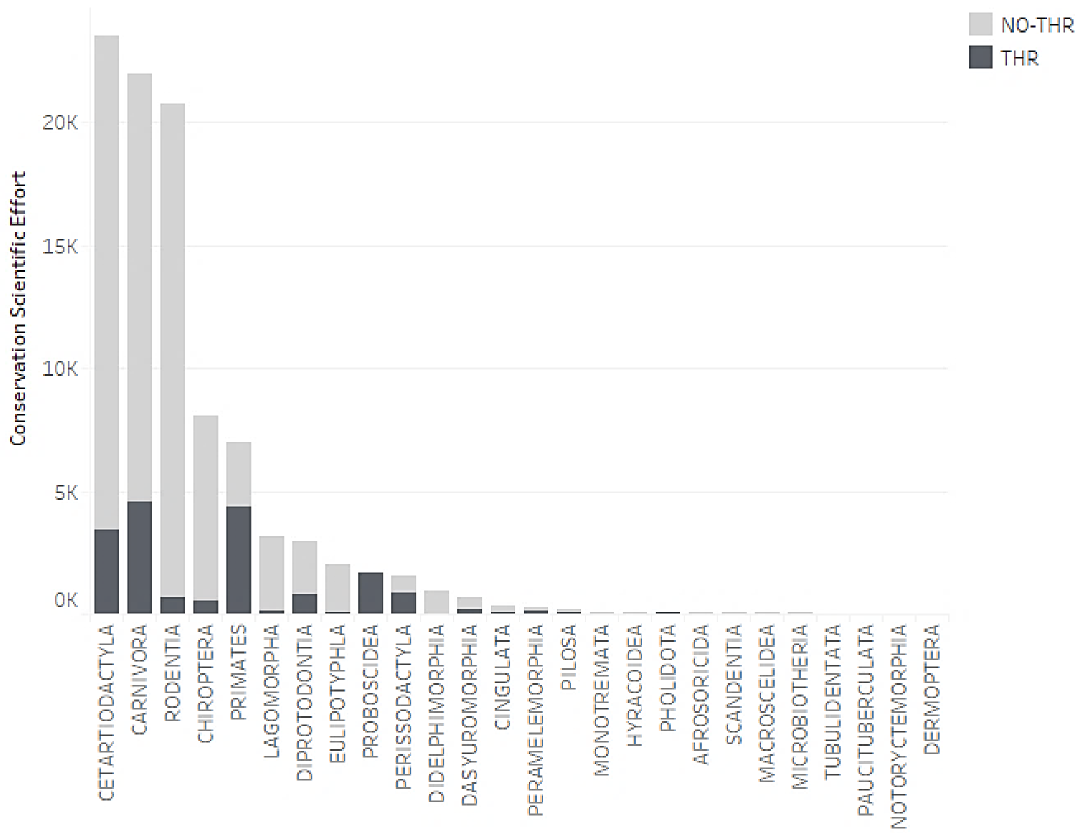


Table S1. Set of best models used in Hurdle Zero Model analysis. The 0-1 values in the columns of the variables represent the absence and presence of the variables in the zero and count models, respectively.

Model rank	**Variables in *count* model**	**Variables in *hurdle* model**	**AICc**	**ΔAICc**	**wi**
1	Body mass + Scientific capacity + Years describing + E.D. score + Geographic range + Nocturnal + Introduced species + Threatened	Body mass + Scientific capacity + Years describing + E.D. score + Geographic range + Threatened	106209.5	0.00	0.44
2	Body mass + Scientific capacity + Years describing + E.D. score + Geographic range + Nocturnal + Introduced species + Threatened	Body mass + Scientific capacity + Years describing + Geographic range + Threatened	106211.3	1.74	0.18
3	Body mass + Scientific capacity + Years describing + E.D. score + Geographic range + Nocturnal + Introduced species + Threatened	Body mass + Scientific capacity + Years describing + E.D. score + Geographic range + Nocturnal + Threatened	106211.4	1.91	0.17
4	Body mass + Scientific capacity + Years describing + E.D. score + Geographic range + Nocturnal + Introduced species + Threatened	Body mass + Scientific capacity + Years describing + E.D. score + Geographic range	106211.8	2.31	0.14
5	Body mass + Scientific capacity + Years describing + E.D. score + Geographic range + Nocturnal + Introduced species + Threatened	Body mass + Scientific capacity + Years describing + Geographic range + Nocturnal + Threatened	106213.1	3.56	0.07