

1 **Cost-effective assessment of extinction risk with limited**  
2 **information**

3 **Standard Paper**

4 Lucie M. Bland<sup>1,2,3\*</sup>, C. David L. Orme<sup>2</sup>, Jon Bielby<sup>3</sup>, Ben Collen<sup>4</sup>, Emily Nicholson<sup>1,5</sup>, Michael A.  
5 McCarthy<sup>1</sup>

6 <sup>1</sup> School of BioSciences, The University of Melbourne, Parkville, Victoria 3010, Australia

7 <sup>2</sup> Division of Biology, Imperial College London, Silwood Park, Ascot, SL5 7PY, UK

8 Institute of Zoology, Zoological Society of London, Regent's Park, London,  
9 NW1 4RY, UK

10 <sup>3</sup> Institute of Zoology, Zoological Society of London, Regent's Park, London,  
11 NW1 4RY, UK

12 <sup>4</sup> Centre for Biodiversity and Environment Research, University College London, Gower Street,  
13 London, WC1 E6BT, UK

14 <sup>5</sup> School of Life and Environmental Sciences, Deakin University, Burwood, Victoria 3125, Australia

15

16 \* **Correspondence author. E-mail:** [l.bland@unimelb.edu.au](mailto:l.bland@unimelb.edu.au)

17 **Running title:** Cost-effective assessment of extinction risk

18

## 19 Summary

- 20 1. Cost-effective reduction of uncertainty in global biodiversity indicators is a central goal of  
21 conservation. Comprising a sixth of the 74,000+ species currently on the IUCN Red List, Data  
22 Deficient species contribute to considerable uncertainty in estimates of extinction risk.  
23 Estimating levels of risk in Data Deficient species will require large resources given the costs  
24 of surveys and Red List assessments. Predicting extinction risk from species traits and  
25 geographical information could provide a cheaper approach for determining the proportion  
26 of Data Deficient species at risk of extinction.
- 27 2. We use double sampling theory to compare the cost-effectiveness of predictive models and  
28 IUCN Red List assessments for estimating risk levels in Data Deficient terrestrial mammals,  
29 amphibians, reptiles and crayfish. For each group, we calibrate Machine Learning models of  
30 extinction risk on species of known conservation status, and assess their cost and reliability  
31 relative to field surveys followed by Red List assessments.
- 32 3. We show that regardless of model type used or species group examined, it is always more  
33 cost-effective to determine the conservation status of all species with models and assess a  
34 small proportion of species with IUCN criteria (double sampling), rather than spend the  
35 same resources on field surveys and Red List assessments alone (single sampling).
- 36 4. We estimate that surveying and re-assessing all Data Deficient species currently listed on the  
37 IUCN Red List (12,206 species) with IUCN criteria would cost a minimum of US \$323 million.  
38 Double sampling reduces the cost of determining the proportion of Data Deficient species at  
39 risk of extinction by up to 68%, because less than 6% of Data Deficient species would need to  
40 be surveyed and assessed with IUCN criteria.
- 41 5. **Synthesis and applications.** Double sampling with models cost-effectively estimates  
42 extinction risk levels in poorly-known species, and can be used to reduce the impact of  
43 uncertainty in the Red List and Red List Index. We provide recommendations for uptake by  
44 managers and a sampling planner spreadsheet. Double sampling could be applied more  
45 widely in ecology and conservation to formally compare the cost-effectiveness of sampling  
46 methods differing in cost and reliability.

47 **Key-words:** Aichi biodiversity targets, amphibians, biodiversity indicators, cost-effectiveness,  
48 crayfish, IUCN Red List, mammals, reptiles.

## 49 **Introduction**

50 Global indicators of biodiversity status are central to monitoring progress towards the 2020 Aichi  
51 targets (Convention on Biological Diversity 2010), and assessing the outcomes of conservation  
52 actions globally. Resources for conservation are orders of magnitude below what is needed to  
53 reverse declines in biodiversity (McCarthy *et al.* 2012), so biodiversity monitoring needs to inform  
54 conservation decisions cost-effectively (McDonald-Madden *et al.* 2010). Representativeness and  
55 reliability have been identified as desirable properties of successful indicators (Dobson 2005; Jones  
56 *et al.* 2011), but the costs of achieving these are not well understood. Developing reliable  
57 biodiversity indicators with limited funds is therefore a pressing challenge for conservation science.

58 The taxonomic coverage of the IUCN Red List has improved in recent years (Collen & Bailie 2010;  
59 Böhm *et al.* 2013), with more than 74,000 species assessed as of 2014 (IUCN 2014). However, one in  
60 six species on the Red List are too poorly known to assign to a category of extinction risk, and are  
61 listed as Data Deficient (DD). This gap in knowledge contributes to considerable uncertainty in global  
62 patterns of extinction risk (Bland *et al.* 2012) and conservation prioritization (Trindade-Filho *et al.*  
63 2012). Re-assessment of the 12,206 species currently listed as DD to data-sufficient categories will  
64 require substantial financial resources, given the costs of biodiversity surveys (Gardner *et al.* 2008)  
65 and Red List assessments (Stuart *et al.* 2010). As a consequence, cost-effectively estimating risk  
66 levels in DD species is crucial to reducing uncertainty in the IUCN Red List and Red List Index (Bailie  
67 *et al.* 2008).

68 Comparative studies of extinction risk based on species trait data have yielded insight into the  
69 determinants of risk among groups (Purvis *et al.* 2000; Cardillo & Meijaard 2012), and could  
70 underpin a preliminary re-assessment of DD species (Davidson *et al.* 2009; Jones & Safi 2011). Good  
71 coverage of species' trait data is available for a large number of DD species and includes life-history,  
72 ecological and phylogenetic information. The geographic distribution of many DD species is known,  
73 allowing inference of species' geographical range size, environmental niche and exposure to

74 anthropogenic threats. These data alone are insufficient for making a decision on formal Red List  
75 status, but could be used to inform global estimates of risk. Recently developed Machine Learning  
76 models of extinction risk based on species trait data have shown excellent predictive performance,  
77 and have been used to predict the likely status of DD mammals (Bland *et al.* 2014).

78 Models may be cheaper to apply than collecting field-based data to update Red List assessments of  
79 Data Deficient species, but model predictions may be inaccurate and bias estimates of extinction  
80 risk. Given the importance of reducing uncertainty in global biodiversity indicators, and the trade-off  
81 between the cost of a monitoring method and its reliability (McDonald-Madden *et al.* 2010), how  
82 can we cost-effectively estimate extinction risk levels in DD species? We use sampling theory to  
83 answer this question. Specifically, we compare the variance in the estimated proportion of DD  
84 species threatened with extinction with two methods:

85 i) *Single sampling.* The proportion of DD species at risk of extinction is inferred from surveying  
86 and updating Red List assessments for a random subset of DD species.

87 ii) *Double sampling.* The same financial resources are shared between developing predictive  
88 models of extinction risk based on species data for all species, and updating Red List  
89 assessments for a smaller set of species. Given the relative costs of these two procedures and  
90 expected accuracy of model classifications, double sampling theory (Tenenbein 1970)  
91 identifies both the optimal allocation of funds to each process and the resulting variance in  
92 the estimated proportion of threatened species. If model development is sufficiently cheap  
93 and accurate, double sampling can give more precise estimates of risk prevalence than single  
94 sampling. Double sampling theory is frequently used in medical diagnostics (Zhou, McClish &  
95 Obuchowski 2002) and quality control (Poduri 2005), but few ecological applications exist  
96 (Harper *et al.* 2004; Rayner, Ellis & Taylor 2011).

97 In the context of this paper, species can be assessed as threatened or non-threatened as defined  
98 within the Red List categories (threatened: Vulnerable, Endangered and Critically Endangered. Non-  
99 threatened: Near Threatened and Least Concern; IUCN 2001), with Red List assessments or with  
100 predictive models of extinction risk. We use four taxonomic groups with varying levels of data  
101 deficiency as case studies: terrestrial mammals (n=4,997; 22.1% DD), amphibians (n=4,449; 41.7%  
102 DD), reptiles (n=1,500; 20.1% DD) and crayfish (n=586; 31.3% DD). For each group, we calibrate  
103 Machine Learning models of extinction risk on species of known conservation status (data-sufficient  
104 species), and assess their reliability compared to Red List assessments. We compute the costs of field  
105 data collection and updating Red List assessments, and compare them with the costs of model  
106 development. We then devise the most cost-effective strategy for determining the proportion of DD  
107 species threatened with extinction in each group.

## 108 **Materials and methods**

### 109 **Double sampling**

110 We estimate the proportion of threatened species ( $p$ ) and its associated variance with double  
111 sampling theory. In practice an investigator may wish to minimize the variance in the estimation of  $p$   
112 for a given budget; alternatively, she may wish to obtain a given variance in the estimation of  $p$  for a  
113 minimum budget. Tenenbein (1970) derives identical solutions for these two problems; we outline  
114 his main results and refer the reader to Tenenbein (1970) for complete statistical derivations. We  
115 compare two estimates of variance of  $p$ :

$$116 \quad V_s = \frac{pq}{n_s} \quad \text{Equation 1}$$

$$117 \quad V_d = \frac{pq}{n_d}(1 - K) + \frac{pq}{N}K \quad \text{Equation 2}$$

118 First, the variance under single sampling ( $V_s$ ) is the binomial variance: we conduct a small set of  
119 expensive assessments of size  $n_s$  and find the proportion of threatened ( $p$ ) and non-threatened

120 species ( $q=1 - p$ ). Second, for the variance under double sampling ( $V_d$ ), we share the cost between  
 121 cheap modelling for a larger set of species ( $N$ ) and assessments for a small subset of modelled  
 122 species ( $n_d$ ) and again find the proportions of threatened and non-threatened species. Note that  
 123  $n_d < n_s < N$ : by modelling some species, we cannot afford to assess as many species.

124 The comparison of these two variances hinges on the coefficient of reliability of the model ( $K$ ), which  
 125 lies in the range  $[0,1]$ . If the model is perfect ( $K=1$ ), then  $V_d=(pq)/N$ , and since  $n_s < N$ , we gain a more  
 126 precise estimate of  $p$  than from single sampling. If the model is useless ( $K=0$ ), we only have  
 127  $V_d=(pq)/n_d$ , and since  $n_d < n_s$ , we have a less precise estimate of  $p$ . For intermediate values of  $K$ ,  $V_d$  is  
 128 weighted average of these two extremes.

129 To use this approach in practice, we derive:

130 i) The coefficient of reliability of the model ( $K$ ). Below, we estimated  $K$  from Machine Learning  
 131 predictions of the conservation status of data-sufficient species based on species data. In the  
 132 absence of a data set, a preliminary survey would be required to estimate  $K$ . The size of this  
 133 survey can be estimated separately (see Discussion and Tenenbein (1971)). The calculation of  
 134  $K$  uses key values calculated from a confusion matrix: the assessed proportions of threatened  
 135 ( $p$ ) and non-threatened species ( $q$ ), the model misclassification probabilities for threatened ( $\phi$ )  
 136 and non-threatened ( $\theta$ ) species, and the modelled proportion of threatened species ( $\pi$ ). From  
 137 these values, Tenenbein (1970) derives:

138 
$$K = \frac{pq(1-\theta-\phi)^2}{\pi(1-\pi)}$$
 Equation 3

139 The example below shows the cell probabilities, and a confusion matrix for the classification of  
 140 109 crayfish species in the validation set by assessments (rows) and the best model (columns):

	thr.	n.thr.		thr.	n.thr.		
thr.	$p(1-\phi)$	$p\phi$	$p$	thr.	28	6	34
n.thr.	$q\theta$	$q(1-\theta)$	$q$	n.thr.	6	69	75

141 With thr.: threatened; n.thr.: non-threatened species. From this we calculate:

142  $p=34/109=0.312$ ,  $q=75/109=0.688$ ,  $\phi=6/34=0.176$ ,  $\theta=6/75=0.08$ ,  $\pi = 34/109=0.312$ , hence

143  $K=0.555$ .

144 ii) The costs of risk assessments ( $c_1$ ) and modelling ( $c_2$ ) per species, and their cost ratio ( $R=c_1/c_2$ ).

145 Below, we estimated these values from the cost of previous assessments and the combined  
 146 costs of collating data bases and modelling.

147 iii) The sampling ratio ( $f_0$ ), giving an optimal division of costs between modelling and assessment  
 148 ( $n_d=N*f_0$ ) that minimizes the variance  $V_d$ . This is derived by Tenenbein (1970) as:

149 
$$f_0 = \min \left[ \sqrt{\frac{1-K}{KR}}, 1 \right] \quad \text{Equation 4}$$

150 If  $f_0$  is close to 1, it is unlikely that double sampling will be cost effective since nearly all  
 151 modelled species must also be assessed, but if  $f_0 < 1$  then double sampling may generate more  
 152 precise estimates for the same cost. A crucial metric is the proportional reduction in cost (or  
 153 variance) achieved by double sampling ( $\lambda$ ):

154 
$$\lambda = 1 - \frac{\left(R + \frac{1}{f_0}\right)(1-K-Kf_0)}{R} \quad \text{Equation 5}$$

155 The threshold  $\lambda > 0$  (Fig. 1a) gives the region in which double sampling is a cost effective alternative  
 156 to single sampling.

157 **Estimating the coefficient of reliability  $K$**

158 We developed predictive models of extinction risk for four taxonomic groups: terrestrial mammals  
 159 (hereafter, mammals), amphibians, reptiles and crayfish (Table 1). We defined data-sufficient species  
 160 as threatened (Critically Endangered, Endangered or Vulnerable) or non-threatened (Near  
 161 Threatened or Least Concern). For each group, we predicted the conservation status of data-

162 sufficient species with life-history, ecological, environmental and threat exposure information. We  
163 collected new data sets for reptiles and crayfish. For reptiles, we collected the following life-history  
164 and ecological traits: maximum snout-vent length, reproductive mode, trophic level, habitat type,  
165 and number of IUCN-listed habitats (Böhm *et al.* 2013). For crayfish, we collected: maximum  
166 carapace length, habitat type, and number of IUCN-listed habitats (IUCN 2010) (Appendix S1 in  
167 Supporting Information). Using mean values from within species' geographic ranges, we compiled  
168 species' spatial data with ArcGIS 9.2 as follows:

- 169 i) Niche. For both reptiles (Böhm *et al.* 2013) and crayfish (IUCN 2010), we extracted:  
170 temperature, temperature seasonality, precipitation, precipitation seasonality, minimum  
171 elevation, and elevation range (Hijmans *et al.* 2005). We also extracted the latitude of the  
172 range centroid and extent of occurrence.
- 173 ii) Threat exposure. For reptiles, we extracted: Human Footprint (CIESIN 2005a), mean and  
174 minimum human population density for the year 2000 (CIESIN 2005b). For crayfish, we  
175 extracted: water consumption, wetland disconnectivity, river fragmentation, mercury  
176 deposition, pesticide loading and sediment loading (Vorosmarty *et al.* 2010).

177 We used an existing mammal data set (Bland *et al.* 2014), and collated an amphibian data set  
178 (Appendix S1) based on Bielby *et al.* (2008) and Cooper *et al.* (2008). Biological traits were  
179 phylogenetically imputed for 32–58% of mammal species (Appendix S1); all other data for all groups  
180 were 100% complete. Species data varied among groups, due to differences in variable  
181 measurement, variable availability, and variable relevance to risk prediction. Data sets remain  
182 comparable in the sense that they use the best macroecological data available to date to predict  
183 extinction risk in poorly known species.

184 Machine Learning (ML) tools are increasingly used in ecology for statistical pattern recognition  
185 (Cutler *et al.* 2007; Olden, Lawler & Poff 2008). For mammals and reptiles, we trained classification  
186 trees, boosted trees, random forest, k-nearest neighbours, support vector machines and neural



187 networks (Bland *et al.* 2014). For amphibians and crayfish, we only trained classification trees,  
188 random forests and boosted trees, as necessary data pre-processing for other ML tools increased  
189 model misclassifications (Appendix S1). For all groups, we trained decision stumps based on  
190 geographical range size alone (IUCN criterion B) to assess its predictive power. Range boundaries  
191 may be more uncertain for DD species than data-sufficient species. To assess the influence of  
192 uncertainty in range size on model predictions, we coarsened species range sizes by rounding log-  
193 transformed range sizes to the nearest higher integer (e.g. 1 = 0 to 1 km<sup>2</sup>, 8 = 10,000,000 to  
194 100,000,000 km<sup>2</sup>). We then recalibrated all models of extinction risk.

195 We partitioned data-sufficient species into a training set comprising 75% of species and a validation  
196 set comprising 25% of species. For each ML tool and data set in turn, we optimized tuning  
197 parameters using ten-fold cross-validation on the training set. For each combination of tuning  
198 parameters, we measured area under the receiver operating characteristic curve (AUC) in the cross-  
199 validation test folds. We selected AUC as measure of model performance, as it is insensitive to class  
200 imbalance and does not require the specification of misclassification costs (Fawcett 2006). ML tools  
201 were compared independently on the validation sets previously set aside. As predictions of risk were  
202 probabilistic, predicting the risk category of a species required a threshold of predicted risk above  
203 which a species should be classified as threatened. For each trained model we calculated the  
204 reliability coefficient  $K$  among all realizable thresholds and selected the threshold maximizing  $K$ .

### 205 **Estimating the cost ratio $R$**

206 For each taxon we calculated the cost of risk assessments ( $c_1$ ) and the cost of predictive models ( $c_2$ ),  
207 expressed in US dollars (\$) per species.

- 208 i) *Assessment costs ( $c_1$ )*. The cost of a risk assessment includes the cost of collecting  
209 information to a level suitable for the application of IUCN Red List criteria, and re-  
210 assessment by the IUCN. Collecting sufficient data for poorly known species to estimate

211 population size or conduct quantitative analyses will be difficult, considering the short  
212 timeframe relevant to most global conservation targets (e.g. Aichi targets; Convention  
213 on Biological Diversity 2010). In addition, most species in poorly known groups are  
214 assessed under IUCN criterion B (restricted range size) (e.g. 80% of threatened crayfish).  
215 We therefore focused on criterion B, which can be predominantly investigated with  
216 presence/absence surveys. We estimated survey costs through consultation with  
217 experts from the IUCN/SCC Specialist Groups and a range of funding bodies for  
218 threatened species research (Appendix S1). We computed three survey costs for  
219 mammals based on geographical range size (Appendix S1). We computed one survey  
220 cost for amphibians, reptiles and crayfish as range size is less variable among species  
221 (Appendix S1). We derived IUCN Red List assessments costs from published sources  
222 (Stuart *et al.* 2010) and consultation with IUCN assessors (mammals: B. Collen,  
223 amphibians: A. Angulo, reptiles: M. Böhm, crayfish: N. Richman).

224 ii) *Predictive model costs ( $c_2$ )*. Predictive model building involves the following stages:  
225 collection of species trait data, GIS extractions of species range maps, data cleaning, and  
226 ML model calibration. We computed the project and staff costs of collecting species data  
227 from data base compilers for mammals (Jones *et al.* 2009; Bland *et al.* 2014), amphibians  
228 (Bielby *et al.* 2008), reptiles (M. Böhm, *pers. comm.*), and crayfish (this study). We  
229 computed three costs of mammal trait data as costs for the panTHERIA data base were  
230 uncertain (Jones *et al.* 2009). We computed the cost of data cleaning and ML model  
231 calibration based on the recorded task time and staff costs of a postdoctoral researcher.  
232 Details of costs for both risk assessments and predictive models are available in  
233 Appendix S1.

## 234 **Results**

### 235 **Estimating the coefficient of reliability $K$**

236 Machine Learning tools achieved high classification performance in all groups as measured by AUC  
237 (Table 2). Values of the coefficient of reliability  $K$  ranged between 0 and 0.7 among models and taxa,  
238 where 1 indicates perfect congruence between predictive models and IUCN assessments. The best  
239 models were random forests in mammals ( $K=0.7$ ) and crayfish ( $K=0.555$ ), boosted trees in  
240 amphibians ( $K=0.629$ ), and neural networks in reptiles ( $K=0.485$ ). Models calibrated on a coarse  
241 measure of range size achieved lower maximum  $K$  values than models calibrated on raw range size  
242 (Table 2). Decision stumps based on geographical range size alone achieved lowest  $K$  values in all  
243 taxa (Table 2).

#### 244 **Estimating the cost ratio $R$**

245 We calculated cost ratios  $R$  ( $c_1/c_2$ ) of 233, 1,877 and 2,489 for mammals, contingent on the three  
246 cost estimates of trait data. We computed cost ratios of 836 for amphibians, 1,375 in reptiles, and  
247 1,401 for crayfish. We present results for a medium cost ratio of  $R=1,500$  among all groups, as the  
248 choice of cost ratio did not qualitatively affect results (see Appendix S1 for alternative cost ratios).  
249 Models based on geographical range size alone achieved very low costs relative to risk assessments  
250 (mammals:  $R=2,409,673$ ; amphibians:  $R=235,902$ ; reptiles:  $R=481,397$ ; crayfish:  $R=272,131$ ). We  
251 present results for a medium cost ratio of  $R=250,000$  among all groups, as the choice of cost ratio  
252 did not qualitatively affect results (Appendix S1).

#### 253 **Double sampling**

254 It was always more cost-effective to determine the status of all DD species with predictive models  
255 and assess a small sample of species with risk assessments (double sampling), rather than spend the  
256 same resources on risk assessments alone (single sampling) (Fig. 1). If all DD species were modelled  
257 with the best models calibrated on raw range size, assessments by the IUCN were required for a  
258 random selection of 11 mammals, 25 amphibians, 8 reptiles, and 3 crayfish. The number of risk  
259 assessments increased when models were calibrated on a coarse measure of range size, requiring

260 the random selection of 43 mammals, 68 amphibians, 25 reptiles, and 9 crayfish. For the best model  
261 calibrated on precise data, reduction in cost achieved by double sampling was 68% in mammals, 60%  
262 in amphibians, 46% in reptiles, and 53% in crayfish (Fig. 1). Reduction in cost decreased when  
263 coarsening range size: the best models achieved reductions in cost of 36–59% among groups.  
264 Reductions in cost achieved by models calibrated on range size alone were low: 32% in mammals,  
265 47% in amphibians, 25% in reptiles, and 16% in crayfish (Fig. 1).

## 266 **Discussion**

267 We find that it is always more cost-effective to model the risk status of all DD species in a group and  
268 update IUCN Red List assessments for a small number of DD species (double sampling), compared to  
269 allocating all financial resources to updating IUCN Red List assessments (single sampling). Double  
270 sampling reduces the cost of determining the proportion of DD species at risk of extinction by up to  
271 68%, as pre-existing biological data are used to minimize the number of field surveys to perform.

272 Assuming DD species not included in this study can be surveyed and assessed for similar costs as  
273 amphibians, reptiles and crayfish (US \$25,400 per species), we estimate the total cost of surveying  
274 and risk assessments for all 12,206 DD species on the Red List (IUCN 2013) to US \$323 million. Our  
275 figure does not reflect efficiencies in surveying multiple species simultaneously (Gardner *et al.* 2008),  
276 or the costs of assessing species under criteria other than B, which may provide more complete  
277 information on risk status. The cost of increasing the number of species on the Red List to 160,000  
278 has been estimated at US \$60 million (Barometer of Life: Stuart *et al.* 2010 also see Collen & Bailie  
279 2010). Many invertebrate, plant and fungi species to be included in the Barometer of Life are not  
280 well studied and may be assessed as DD, so the initiative is likely to require considerable additional  
281 investment in field surveys. Under current funding of the Red List, more than 90% of the Barometer  
282 of Life assessments will become outdated in the next decade (Rondinini *et al.* 2014). Limited  
283 resources for tracking the status of biodiversity create a trade-off between expanding the taxonomic  
284 coverage of biodiversity assessments (Collen *et al.* 2009), keeping assessments up-to-date (Rondinini

285 *et al.* 2014), and ensuring reliability of risk assessments (Bland *et al.* 2012). Improving our  
286 understanding of the costs and trade-offs involved in creating biodiversity indicators is therefore a  
287 key topic for further research. These are the costs merely for understanding extinction risks; the cost  
288 of reducing the extinction risk of all globally threatened species was estimated at US \$3.41 to \$4.76  
289 billion, of which only 12% is currently funded (McCarthy *et al.* 2012).

290 Data Deficient species contribute to considerable uncertainty in conservation prioritization (Bland *et*  
291 *al.* 2012; Trindade-Filho *et al.* 2012) and may jeopardize the measurement of progress towards Aichi  
292 targets (Tittensor *et al.* 2014), particularly towards improving the conservation status of threatened  
293 species (Target 12: Convention on Biological Diversity 2010). Yet, they receive very little  
294 conservation investment: for example 3% of the awards from the Mohamed bin Zayed Species  
295 Conservation Fund are directed toward DD species (MBZSCF 2013). We show that using existing  
296 biological data and conducting risk assessments for poorly known species could enable the cost-  
297 effective monitoring of progress towards international biodiversity targets. Extinction risk models  
298 are not only a cheaper option than risk assessments for monitoring broad-scale changes in risk; they  
299 are also more likely to be developed within time scales relevant to biodiversity targets. Group  
300 assessments require extensive workshops, administration and training and typically take several  
301 years to complete (Rondinini *et al.* 2014). On the other hand, models require collection of data from  
302 species descriptions and other natural history resources, which can be carried out rapidly by non-  
303 experts. Whilst calibration of ML tools requires statistical expertise, accessibility could be improved  
304 by developing user-friendly platforms.

### 305 **Comparison of models and taxonomic groups**

306 The utility of risk models for conservation depends on their reliability and cost relative to risk  
307 assessments. In the focal groups considered in this study, models achieved very high AUC in  
308 validation sets, indicating excellent discrimination between threatened and non-threatened species  
309 (Table 2). Predictive performance varied among groups (Table 2) likely due to complex interactions

310 among taxon size, risk prevalence and species data availability. Most models performed better on  
311 mammals and amphibians than crayfish or reptiles, likely due to the high number of mammals  
312 modelled and the high prevalence of risk in amphibians. Random Forests, Boosted Trees, Neural  
313 Networks and Support Vector Machine achieved high coefficients of reliability. We recommend  
314 testing multiple ML tools for predicting risk in new species groups.

315 The biggest savings were achieved by improving model performance, whilst savings were less  
316 sensitive to the estimate of risk assessment and model costs. Approximate cost ratios may therefore  
317 be sufficiently informative when designing double sampling schemes (Appendix S1). Double  
318 sampling remains cost-effective under poor data quality: models calibrated on a coarse measure of  
319 range size still achieved 34–56% reduction in cost among groups. Reductions in cost achieved by  
320 models calibrated on range size alone were smaller (15–47%), indicating that collecting biological  
321 data is necessary to achieve the highest cost savings.

322 Double sampling may not be cost-effective under certain conditions. With levels of congruence  
323 between predictive models and IUCN Red List assessments of  $0.4 < K < 0.7$ , double sampling is not cost-  
324 effective when the costs of modelling and updating Red List assessments are about equal ( $R < 1.5$ ),  
325 which is unlikely to occur. With poor models ( $K < 0.1$ ), reductions in cost are small ( $< 10\%$ ) so  
326 managers may decide the overhead costs of calibrating models are not worthwhile. If risk  
327 assessments are at least 250 times more expensive than predictive models, and models relatively  
328 reliable ( $K > 0.4$ ), double sampling reduces cost by 40% or more – a good rule of thumb for managers  
329 wishing to use predictive models. To facilitate exploration of possible savings, we provide an Excel  
330 planning model (Appendix S2).

### 331 **Limitations and prospects**

332 We modelled binomial threat status (threatened vs. non-threatened) rather than Red List categories,  
333 due to difficulties in modelling highly imbalanced response categories with the available data

334 (Hastie, Tibshirani & Friedman 2009). A multinomial double sampling scheme (Tenenbein 1972)  
335 could investigate the cost-effectiveness of estimating the prevalence of individual Red List  
336 categories. We assume that the relationship between predictor variables and extinction risk is  
337 similar in data-sufficient and DD species. Accurate predictions require the range of predictor values  
338 exhibited by DD species to be represented by modelled data-sufficient species (Appendix S1; DD  
339 data not available for amphibians). Modelled species are also assumed to be representative of the  
340 wider data-sufficient species pool (Table 2). It should be noted that the utility of models is  
341 contingent on the quality of IUCN Red List assessments. The Red List status of some species may  
342 change due to genuine improvements and deteriorations in conservation status, as well as previous  
343 misclassifications (non-genuine reasons; Butchart *et al.* 2004).

344 Estimation of predictor variables may be less accurate in DD species, which could affect model  
345 performance. We used the best available data and investigated the role of uncertainty in  
346 geographical range size, and show that data uncertainty can be readily incorporated into a double  
347 sampling scheme. We also assume that geographical range maps are available for all species to  
348 assess in a sample, which may not be the case for all DD species (although only 3 DD crayfish species  
349 could not be mapped; B. Collen *pers. comm.*) or for species not assessed by the IUCN (e.g. species  
350 not selected in the Sampled Red List assessment of their taxonomic group; Bailie *et al.* 2008). For  
351 such species, the cost of constructing a range map from occurrence records and atlases would need  
352 to be incorporated in the costs of predictive models.

353 The estimation of  $K$  may depend on the number of species used to calibrate models of extinction  
354 risk. To deal with this problem, a three-stage sampling scheme (Tenenbein 1971) can determine the  
355 size of a pilot study to estimate  $K$ . Rarefaction analyses indicate that fewer than 1,000 mammals  
356 could have been assessed to obtain  $K > 0.5$  (unpublished data). Understanding the relationship  
357 between  $K$  and sample size is therefore a useful avenue for further research, especially for species

358 assessed with the Sampled Red List approach. Similarly, future studies could incorporate dynamic  
359 updating of  $K$  as species of unknown conservation status are assessed by the IUCN.

360 The double sampling scheme relies on binomial sampling from an infinite population (Tenenbein  
361 1970). In reality, DD species represent populations of finite size, which are more adequately  
362 modelled by a hypergeometric distribution. As the single sample size approaches the total  
363 population size, the variance in the estimated proportion of DD species at risk decreases faster for a  
364 hypergeometric distribution than for a binomial distribution, eventually reaching zero when all  
365 species have been assessed. Double sampling theory has not been extended to the hypergeometric  
366 distribution, but we have included single sampling under a hypergeometric model in our planning  
367 model (Appendix S2). Under realistic conditions ( $K=0.4$  and  $R=1,500$ ), double sampling with a  
368 binomial distribution performs better than single sampling with a hypergeometric distribution when  
369 funds for red listing are small (e.g. fewer than 188 out of 500 DD species can be assessed). Double  
370 sampling as implemented in this study will therefore yield adequate results under limited budgets,  
371 which are commonplace in conservation biology (McCarthy *et al.* 2012).

372 Finally, our study addresses only one objective of the IUCN Red List, the quantification of global  
373 patterns and trends in extinction risk globally (IUCN 2014). The IUCN Red List also aims to pinpoint  
374 individual species at high risk of extinction (IUCN 2014). Models of extinction risk could address this  
375 objective by identifying high-risk DD species for preferential re-assessment to data-sufficient  
376 categories. Observed or predicted species extinction risk is only part of the information required for  
377 efficient resource allocation (Possingham *et al.* 2002), hence survey costs and likelihood of survey  
378 success should be taken into account during prioritization (Joseph *et al.* 2009).

## 379 **Conclusions**

380 To measure progress towards international targets and halt the current loss biodiversity, reliable  
381 indicators of biodiversity status are needed. We show that double sampling with predictive models



382 cost-effectively estimates the proportion of IUCN DD species at risk of extinction, and reduces  
383 assessments costs by up to 68%. Double sampling remains cost-effective under poor data quality  
384 and availability, demonstrating the method's capacity to cheaply determine extinction risk levels in  
385 poorly known groups of plants and invertebrates. We conclude that double sampling could reduce  
386 the impact of uncertainty in the Red List and Red List Index, and cost-effectively monitor progress  
387 towards Target 12 of the Aichi Biodiversity Targets. The technique could also be applied to local and  
388 national risk assessment programmes, and to species excluded from Sampled Red List assessments.  
389 Double sampling schemes are available for multinomial data (Tenenbein 1972), continuous data  
390 (Gilbert 1987), and for designing pilot studies in multiple stages (Tenenbein 1971). Double sampling  
391 could be applied more widely in ecology and conservation to formally compare the cost-  
392 effectiveness of sampling methods differing in cost and reliability. Given the urgency of the  
393 biodiversity crisis and the limited availability of conservation funds and biological data, designing  
394 efficient monitoring schemes is imperative.

## 395 **Acknowledgements**

396 This research was supported by a Travel Fellowship to L.B. from the Australian Research Council  
397 (ARC) Centre of Excellence for Environmental Decisions, an ARC Future Fellowship to M.A.M. and a  
398 Centenary Research Fellowship to E.N. We thank the collaborative effort among the National  
399 Autonomous University of Mexico, Stony Brook University, Nature Serve and the Institute of Zoology  
400 for the collection of reptile data, particularly Andres Garcia, Monika Böhm and Ana Davidson. We  
401 thank the curation staff of the Natural History Museum London, Musee d'Histoire Naturelle Paris,  
402 Museum Victoria and the Australian Museum for assistance in collecting crayfish data.

403

## 404 **Data accessibility**

## 405 **Supporting Information**

406 Additional Supporting Information may be found in the online version of this article:  
407 **Appendix S1.** Supplementary methods and results.  
408 **Appendix S2.** Sampling planner for double sampling with a binomial distribution, single sampling  
409 with a binomial distribution and single sampling with a hypergeometric distribution.

## 410 References

Baillie, J. E. M., Collen, B., Amin, R., Akcakaya, H. R., Butchart, S. H. M., Brummitt, N., Meagher, T. R., Ram, M., Hilton-Taylor, C., Mace, G. M. (2008). Toward monitoring global biodiversity. *Conservation Letters*, **1**, 18–26.

Bielby, J., Cooper, N., Cunningham, A.A., Garner, T.W.J. & Purvis, A. (2008) Predicting susceptibility to future declines in the world's frogs. *Conservation Letters*, **1**, 82–90.

Bland, L.M., Collen, B., Orme, C.D.L. & Bielby, J. (2012) Data uncertainty and the selectivity of extinction risk in freshwater invertebrates. *Diversity and Distributions*, **18**, 1211–1220.

Bland, L.M., Collen, B., Orme, C.D.L. & Bielby, J. (2014) Predicting the conservation status of Data Deficient species. *Conservation Biology*.

Böhm, M., Collen, B., Baillie, J.E.M., Bowles, P., Chanson, J., Cox, N., Hammerson, G., Hoffmann, M., Livingstone, S.R., Ram, M., Rhodin, A.G.J., Stuart, S.N., van Dijk, P.P., Young, B.E., Afuang, L.E., Aghasyan, A., García, A., Aguilar, C., Ajtic, R., Akarsu, F., Alencar, L.R.V., Allison, A., Ananjeva, N., Anderson, S., Andrén, C., Ariano-Sánchez, D., Arredondo, J.C., Auliya, M., Austin, C.C., Avci, A., Baker, P.J., Barreto-Lima, A.F., Barrio-Amorós, C.L., Basu, D., Bates, M.F., Batistella, A., Bauer, A., Bennett, D., Böhme, W., Broadley, D., Brown, R., Burgess, J., Captain, A., Carreira, S., Castañeda, M.D.R., Castro, F., Catenazzi, A., Cedeño-Vázquez, J.R., Chapple, D.G., Cheylan, M., Cisneros-Heredia, D.F., Cogalniceanu, D., Cogger, H., Corti, C., Costa, G.C., Couper, P.J., Courtney, T., Crnobrnja-Isailovic, J., Crochet, P.-A., Crother, B., Cruz, F., Daltry, J.C., Daniels, R.J.R., Das, I., de Silva, A., Diesmos, A.C., Dirksen, L., Doan, T.M., Dodd, C.K., Doody, J.S., Dorcas, M.E., Duarte de Barros Filho, J., Egan, V.T., El Mouden, E.H., Embert, D., Espinoza, R.E., Fallabrino, A., Feng, X., Feng, Z.-J., Fitzgerald, L., Flores-Villela, O., França, F.G.R., Frost, D., Gadsden, H., Gamble, T., Ganesh, S.R., Garcia, M. a., García-Pérez, J.E., Gatus, J., Gaulke, M., Geniez, P., Georges, A., Gerlach, J., Goldberg, S., Gonzalez, J.-C.T., Gower, D.J., Grant, T., Greenbaum, E., Grieco, C., Guo, P., Hamilton, A.M., Hare, K., Hedges, S.B., Heideman, N., Hilton-Taylor, C., Hitchmough, R., Hollingsworth, B., Hutchinson, M., Ineich, I., Iverson, J., Jaksic, F.M., Jenkins, R., Joger, U., Jose, R., Kaska, Y., Kaya, U., Keogh, J.S., Köhler, G., Kuchling, G., Kumlutaş, Y., Kwet, A., La Marca, E., Lamar, W., Lane, A., Lardner, B., Latta, C., Latta, G., Lau, M., Lavin, P., Lawson, D., LeBreton, M., Lehr, E., Limpus, D., Lipczynski, N., Lobo, A.S., López-Luna, M. a., Luiselli, L., Lukoschek, V., Lundberg, M., Lymberakis, P., Macey, R., Magnusson, W.E., Mahler, D.L., Malhotra, A., Mariaux, J., Maritz, B., Marques, O. a. V., Márquez, R., Martins, M., Masterson, G., Mateo, J. a., Mathew, R., Mathews, N., Mayer, G., McCranie, J.R., Measey, G.J., Mendoza-Quijano, F., Menegon, M., Métrailler, S., Milton, D. a., Montgomery, C., Morato, S. a. a., Mott, T., Muñoz-Alonso, A., Murphy, J., Nguyen, T.Q., Nilson, G., Nogueira, C., Núñez, H., Orlov, N., Ota, H., Ottenwalder, J., Papenfuss, T., Pasachnik, S., Passos, P., Pauwels, O.S.G., Pérez-Buitrago, N., Pérez-Mellado, V., Pianka, E.R., Pleguezuelos, J., Pollock, C., Ponce-Campos, P., Powell, R., Pupin, F., Quintero Díaz, G.E., Radder, R., Ramer, J., Rasmussen, A.R., Raxworthy, C., Reynolds, R., Richman, N., Rico, E.L., Riservato, E., Rivas, G., da Rocha, P.L.B., Rödel, M.-O., Rodríguez Schettino, L., Roosenburg, W.M., Ross, J.P., Sadek, R., Sanders, K., Santos-Barrera, G., Schleich, H.H., Schmidt, B.R., Schmitz, A., Sharifi, M., Shea, G., Shi, H.-T., Shine, R., Sindaco, R., Slimani, T., Somaweera, R., Spawls, S., Stafford, P., Stuebing, R., Sweet, S., Sy, E., Temple, H.J., Tognelli, M.F., Tolley, K., Tolson, P.J., Tuniyev, B., Tuniyev, S., Üzümlü, N., van Buurt, G., Van Sluys, M., Velasco, A., Vences, M., Veselý, M., Vinke, S., Vinke, T., Vogel, G., Vogrin, M., Vogt, R.C., Wearn, O.R., Werner, Y.L., Whiting, M.J., Wiewandt, T., Wilkinson, J., Wilson, B., Wren, S., Zamin, T., Zhou, K. & Zug, G. (2013) The conservation status of the world's reptiles. *Biological Conservation*, **157**, 372–385.

Butchart, S.H.M., Stattersfield, A.J., Bennun, L. a, Shutes, S.M., Akçakaya, H.R., Baillie, J.E.M., Stuart, S.N., Hilton-Taylor, C. & Mace, G.M. (2004) Measuring global trends in the status of biodiversity: red list indices for birds. *PLoS biology*, **2**, e383.

Cardillo, M. & Meijaard, E. (2012) Are comparative studies of extinction risk useful for conservation? *Trends in Ecology & Evolution*, **27**, 167–171.

CIESIN. (2005a) Last of the Wild Data Version 2 (LWP-2): Global Human Footprint dataset (HF), <http://sedac.ciesin.columbia.edu/data/collection/wildareas-v2> Accessed on the 10<sup>th</sup> of October 2011.

CIESIN. (2005b) Gridded Population of the World (2000), Version 3 (GPWv3), <http://www.ciesin.columbia.edu/datasets/downscaled/> Accessed on the 10<sup>th</sup> of October 2011.

Collen, B. & Bailie, J.M. (2010) The barometer of life: sampling. *Science*, **329**, 140.

Collen, B., Ram, M., Dewhurst, N., Clausnitzer, V., Kalkman, V.J., Cumberlidge, N. & Bailie, J.E.M. (2009) Broadening the coverage of biodiversity assessments. *Wildlife in a changing world. An analysis of the 2008 IUCN Red List of Threatened Species* pp. 67–75. IUCN, Gland, Switzerland.

Convention on Biological Diversity. (2010) TARGET 12 - Technical Rationale. *COP10 Decisions Tenth meeting of the Conference of the Parties to the Convention on Biological Diversity* CBD, Nagoya, Japan.

Cooper, N., Bielby, J., Thomas, G.H. & Purvis, A. (2008) Macroecology and extinction risk correlates of frogs. *Global Ecology and Biogeography*, **17**, 211–221.

Cutler, R.D., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J. & Lawler, J.J. (2007) Random forests for classification in ecology. *Ecology*, **88**, 2783–92.

Davidson, A.D., Hamilton, M.J., Boyer, A.G., Brown, J.H. & Ceballos, G. (2009) Multiple ecological pathways to extinction in mammals. *Proceedings of the National Academy of Sciences*, **106**, 10702–10705.

Dobson, A. (2005) Monitoring global rates of biodiversity change: challenges that arise in meeting the Convention on Biological Diversity (CBD) 2010 goals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**, 229–41.

Elith, J. & Graham, C.H. (2009) Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography*, **32**, 66–77.

Fawcett, T. (2006) An introduction to ROC analysis. *Pattern Recognition Letters*, **27**, 861–874.

Gardner, T.A., Barlow, J., Araujo, I.S., Avila-Pires, T.C., Bonaldo, A.B., Costa, J.E., Esposito, M.C., Ferreira, L. V., Hawes, J., Hernandez, M.I.M., Hoogmoed, M.S., Leite, R.N., Lo-Man-Hung, N.F., Malcolm, J.R., Martins, M.B., Mestre, L. a M., Miranda-Santos, R., Overal, W.L., Parry, L., Peters, S.L., Ribeiro-Junior, M.A., da Silva, M.N.F., da Silva Motta, C. & Peres, C. a. (2008) The cost-effectiveness of biodiversity surveys in tropical forests. *Ecology Letters*, **11**, 139–50.

Gilbert, R.O. (1987) *Statistical Methods for Environmental Pollution Monitoring*. Van Nostrand Reinhold, NY, USA.

Harper, M.J., McCarthy, M. a, van der Ree, R. & Fox, J.C. (2004) Overcoming bias in ground-based surveys of hollow-bearing trees using double-sampling. *Forest Ecology and Management*, **190**, 291–300.

Hastie, T., Tibshirani, R. & Friedman, J. (2009) *The Elements of Statistical Learning*. Springer, NY, USA.

Hijmans, S.E., Cameron, J.L., Parra, P.G., Jones, A. & Jarvis, R.J. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

IUCN. (2001) IUCN Red List Categories and Criteria: version 3.1, 30 pp., Gland, Switzerland.

IUCN. (2010) 2010 IUCN Red List of threatened species. Version 2010.3, [www.iucnredlist.org](http://www.iucnredlist.org) Accessed on the 10<sup>th</sup> of October 2010.

IUCN. (2014) Red List Overview, <http://www.iucnredlist.org/about/red-list-overview>, Accessed on the 10<sup>th</sup> of

October 2014.

Jones, K.E., Bielby, J., Cardillo, M., Fritz, S.A., O'Dell, J., Orme, C.D.L., Safi, K., Sechrest, W., Boakes, E.H., Carbone, C., Connolly, C., Cutts, M., Foster, J.K., Grenyer, R., Habib, M. & Plaster, ChChristop, and W.K.M. (2009) PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology*, **90**, 2648–2648.

Jones, J.P.G., Collen, B., Atkinson, G., Baxter, P.W.J., Bubb, P., Illian, J.B., Katzner, T.E., Keane, A., Loh, J., McDonald-Madden, E.V.E., Nicholson, E., Pereira, H.M., Possingham, H.P., Pullin, A.S., Rodrigues, A.S.L., Ruiz-Gutierrez, V., Sommerville, M. & Milner-Gulland, E.J. (2011) The Why, What, and How of Global Biodiversity Indicators Beyond the 2010 Target. *Conservation Biology*, **25**, 450–457.

Jones, K.E. & Safi, K. (2011) Ecology and evolution of mammalian biodiversity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **366**, 2451–2461.

Joseph, L. N., Maloney, R. F., Possingham, H. P. (2009). Optimal allocation of resources among threatened species: a project prioritization protocol. *Conservation Biology* **23**, 328-338.

McCarthy, D.P., Donald, P.F., Scharlemann, J.P.W., Buchanan, G.M., Balmford, A., Green, J.M.H., Bennun, L.A., Burgess, N.D., Fishpool, L.D.C., Garnett, S.T., Leonard, D.L., Maloney, R.F., Morling, P., Schaefer, H.M., Symes, A., Wiedenfeld, D.A. & Butchart, S.H.M. (2012) Financial costs of meeting global biodiversity conservation targets: current spending and unmet needs. *Science*, **338**, 946–9.

McDonald-Madden, E., Baxter, P.W.J., Fuller, R.A., Martin, T.G., Game, E.T., Montambault, J. & Possingham, H.P. (2010) Monitoring does not always count. *Trends in Ecology & Evolution*, **25**, 547–50.

Mohamed bin Zayed Species Conservation Fund. (2013) The Mohamed bin Zayed Species Conservation Fund, <http://www.speciesconservation.org/> Accessed on the 10<sup>th</sup> of April 2013.

Olden, J.D., Lawler, J.J. & Poff, N.L. (2008) Machine learning methods without tears: a primer for ecologists. *The Quarterly review of biology*, **83**, 171–93.

Poduri S. R. S. Rao. (2005) Double Sampling. *Encyclopedia of Biostatistics* (eds P. Armitage & T. Colton), p. 718 John Wiley & Sons, Ltd, Chichester, UK.

Possingham, H. P., Andelman, S.J., Burgman, M. A., Medellin, R. A., Master, L. L, Keith, D. A. (2002). Limits to the use of threatened species. *Trends in Ecology & Evolution* **17**, 503-507.

Purvis, A., Gittleman, J.L., Cowlshaw, G. & Mace, G.M. (2000) Predicting extinction risk in declining species. *Proceedings of the Royal Society B: Biological Sciences*, **267**, 1947–1952.

Rayner, L., Ellis, M. & Taylor, J.E. (2011) Double sampling to assess the accuracy of ground-based surveys of tree hollows in eucalypt woodlands. *Austral Ecology*, **36**, 252–260.

Rondinini, C., Di Marco, M., Visconti, P., Butchart, S.H.M. & Boitani, L. (2014) Update or outdate: long-term viability of the IUCN Red List. *Conservation Letters*, **7**, 126-130.

Stuart, S.N., Wilson, E.O., McNeely, J.A., Mittermeier, R.A. & Rodríguez, J.P. (2010) The barometer of life. *Science*, **328**, 177.

Tenenbein, A. (1970) A double sampling scheme for estimating from binomial data with misclassifications. *Journal of the American Statistical Association*, **65**, 1350–1361.

Tenenbein, A. (1971) A Double Sampling Scheme for Estimating from Binomial Data with Misclassifications: Sample size Determination. *International Biometric Society*, **27**, 935–944.

Tenenbein, A. (1972) A Double Sampling Scheme for Estimating from Misclassified Multinomial Data with Applications to Sampling Inspection. *Technometrics*, **14**, 187–202.

Tittensor, D.P., Walpole, M., Hill, S.L.L., Boyce, D.G., Britten, G.L., Burgess, N.D., Butchart, S.H.M., Leadley, P.W., Regan, E.C., Alkemade, R., Baumung, R., Bellard, C., Bouwman, L., Bowles-Newark, N.J., Chenery, A.M., Cheung, W.W.L., Christensen, V., Cooper, H.D., Crowther, A.R., Dixon, M.J.R., Galli, A., Gaveau, V., Gregory, R.D., Gutierrez, N.L., Hirsch, T.L., Höft, R., Januchowski-Hartley, S.R., Karmann, M., Krug, C.B., Leverington, F.J., Loh, J., Lojenga, R.K., Malsch, K., Marques, A., Morgan, D.H.W., Mumby, P.J., Newbold, T., Noonan-Mooney, K., Pagad, S.N., Parks, B.C., Pereira, H.M., Robertson, T., Rondinini, C., Santini, L., Scharlemann, J.P.W., Schindler, S., Sumaila, U.R., Teh, L.S.L., van Kolck, J., Visconti, P. & Ye, Y. (2014) A mid-term analysis of progress toward international biodiversity targets. *Science*, **346**, 241-244.

Trindade-Filho, J., Carvalho, R.A., Brito, D. & Loyola, R.D. (2012) How does the inclusion of Data Deficient species change conservation priorities for amphibians in the Atlantic Forest? *Biodiversity and Conservation*, **21**, 2709–2718.

Vorosmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Liermann, C.R. & Davies, P.M. (2010) Global threats to human water security and river biodiversity. *Nature*, **467**, 555–561.

Zhou, X.H., McClish, D.K. & Obuchowski, N.A. (2002) *Statistical Methods in Diagnostic Accuracy*. Wiley, NY, USA.

412 **Tables**

413 **Table 1** Description of IUCN Red List assessments and predictive models of extinction risk for  
 414 terrestrial mammals, amphibians, reptiles and crayfish

	Number of data-sufficient species*	Number of Data Deficient species	Percentage of threatened data-sufficient species	Number of data-sufficient species in models	Number of predictors of extinction risk	Number of models of extinction risk
<b>Mammals</b>	4,300	677	22.1	3,967	35	7
<b>Amphibians</b>	4,449	1,294	42	478	15	4
<b>Reptiles†</b>	1,199	301	20.1	982	29	7
<b>Crayfish</b>	467	125	31.3	440	24	4

415  
 416  
 417  
 418  
 419  
 420  
 421

\* Data-sufficient species are listed as Least Concern, Near Threatened, Vulnerable, Endangered or Critically Endangered on the Red List.

† Sampled Red List of 1,500 randomly selected reptiles.

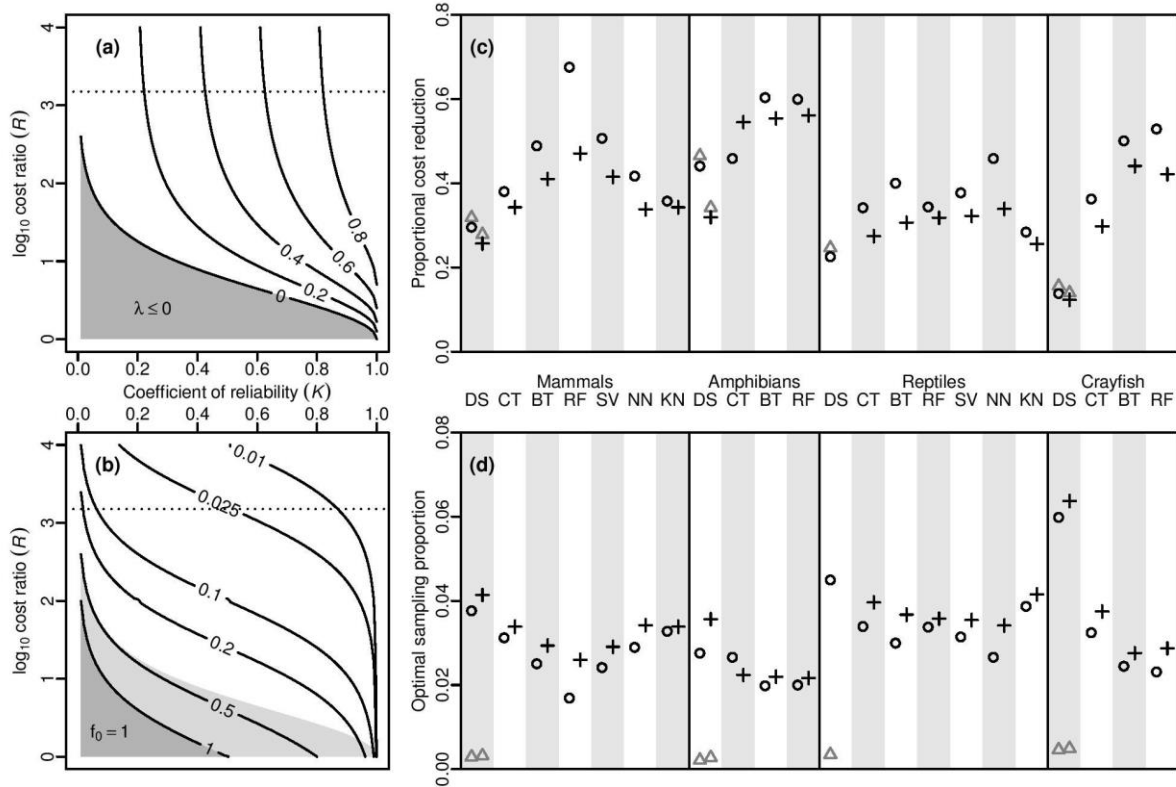
422 **Table 2** Model performances among predictive models and taxonomic groups, for (a) models  
423 calibrated on fine geographical range size, and (b) models calibrated on coarsened geographical  
424 range size



	AUC*	Cutoff†	θ‡	φ¶	π§	κ**	
							425
<b>a) Fine geographical range size</b>							
<b>Mammals</b>							426
Decision Stump	0.75	0.731	0.05	0.447	0.161	0.32	
Classification Tree	0.895	0.3	0.102	0.233	0.249	0.406	427
Boosted Trees	0.935	0.317	0.069	0.201	0.231	0.515	
Random Forests	0.971	0.604	0.014	0.196	0.189	0.7	428
Support Vector Machines	0.932	0.385	0.059	0.21	0.221	0.533	
Neural Networks	0.922	0.448	0.082	0.242	0.231	0.443	
K-Nearest Neighbours	0.906	0.345	0.069	0.333	0.201	0.383	429
<b>Amphibians</b>							
Decision Stump	0.842	0.731	0.136	0.18	0.569	0.467	430
Classification Tree	0.898	0.846	0.1	0.196	0.5	0.485	
Boosted Trees	0.949	0.269	0.03	0.2	0.638	0.629	431
Random Forests	0.953	0.428	0.045	0.18	0.621	0.625	
<b>Reptiles</b>							432
Decision Stump	0.726	0.731	0.059	0.488	0.135	0.248	
Classification Tree	0.895	0.192	0.196	0.049	0.322	0.367	433
Boosted Trees	0.928	0.164	0.147	0.073	0.277	0.426	
Random Forests	0.916	0.354	0.107	0.219	0.22	0.369	
Support Vector Machines	0.925	0.214	0.113	0.171	0.233	0.403	434
Neural Networks	0.943	0.283	0.108	0.097	0.24	0.485	
K-Nearest Neighbours	0.894	0.255	0.117	0.268	0.22	0.308	435
<b>Crayfish</b>							
Decision Stump	0.698	0.731	0.026	0.706	0.11	0.157	436
Classification Tree	0.874	0.828	0.053	0.382	0.229	0.388	
Boosted Trees	0.927	0.38	0.093	0.176	0.321	0.527	437
Random Forests	0.919	0.456	0.08	0.176	0.312	0.555	
<b>b) Coarse geographical range size</b>							438
<b>Mammals</b>							
Decision Stump	0.718	0.731	0.038	0.525	0.135	0.28	439
Classification Tree	0.875	0.75	0.045	0.411	0.165	0.368	
Boosted Trees	0.912	0.456	0.062	0.297	0.204	0.436	440
Random Forests	0.927	0.408	0.046	0.279	0.196	0.497	
Support Vector Machines	0.915	0.394	0.058	0.301	0.199	0.441	
Neural Networks	0.892	0.36	0.096	0.292	0.231	0.363	441
K-Nearest Neighbours	0.897	0.276	0.124	0.228	0.267	0.368	442
<b>Amphibians</b>							
Decision Stump	0.769	0.731	0.4	0.06	0.706	0.344	443
Classification Tree	0.9	0.286	0.12	0.12	0.551	0.571	
Boosted Trees	0.94	0.69	0.08	0.151	0.517	0.58	
Random Forests	0.946	0.666	0.06	0.167	0.5	0.587	
<b>Reptiles</b>							
Decision Stump	0.5	0	0	0	0	0	
Classification Tree	0.854	0.09	0.147	0.219	0.253	0.298	
Boosted Trees	0.901	0.162	0.152	0.171	0.265	0.331	
Random Forests	0.89	0.242	0.157	0.146	0.273	0.343	
Support Vector Machines	0.907	0.207	0.142	0.171	0.257	0.347	
Neural Networks	0.919	0.427	0.064	0.341	0.163	0.364	
K-Nearest Neighbours	0.88	0.246	0.122	0.293	0.22	0.279	
<b>Crayfish</b>							
Decision Stump	0.633	0.731	0.133	0.471	0.256	0.141	
Classification Tree	0.823	0.727	0.12	0.323	0.294	0.322	
Boosted Trees	0.868	0.432	0.107	0.206	0.321	0.467	
Random Forests	0.883	0.38	0.2	0.088	0.422	0.447	

- 444 \* AUC: area under receiver-operator characteristic curve,  
445 † Cutoff: predicted probability of risk above which a species is classified as threatened.  
446 ‡  $\theta$ : probability of misclassification for genuinely threatened species.  
447 ¶  $\phi$ : probability of misclassification for genuinely non-threatened species.  
448 §  $\pi$ : proportion of threatened species estimated by the model. The true proportion of threatened  
449 species in the sample ( $p$ ) for each group is: mammals = 0.221, amphibians = 0.568, reptiles = 0.169,  
450 crayfish = 0.312.  
451 \*\*  $K$ : coefficient of reliability of the model.

452 **Figures**



453

454

**Figure 1** Proportional reduction in cost and optimal sampling proportion for double sampling

455

assessments of extinction risk. (a) Proportional reduction in cost for double sampling assessments,

456

given model reliability (K) and cost ratio (R), showing λ < 0 (light grey). (b) Optimal sampling

457

proportion given model reliability (K) and cost ratio (R), showing f<sub>0</sub> = 1 (dark grey) and λ < 0 (light

458

grey). (a) and (b): horizontal dotted line indicates the estimated cost ratio (R = 1,500). (c) Proportional

459

reduction in cost and (d) optimal sampling proportion among predictive models and taxonomic

460

groups. (c) and (d); circles: models calibrated on fine geographical range size. Crosses: models

461

calibrated on coarsened geographical range size. Triangles: models calibrated on range size alone

462

(left: fine geographical range size; right: coarse geographical range size). CT: classification tree. RF:

463

random forests. BT: boosted trees. SV: support vector machines. NN: neural networks. KN: k-nearest

464

neighbours. DS: decision stumps.