Universidade Federal do Rio de Janeiro

University of Regina

Práticas em Ciência Aberta

Bruno Soares, Ph.D.
Guilherme Maricato, Ph.D.
Bruno Umbelino, M.Sc.

Replicabilidade não é a norma

RESEARCH ARTICLE SUMMARY

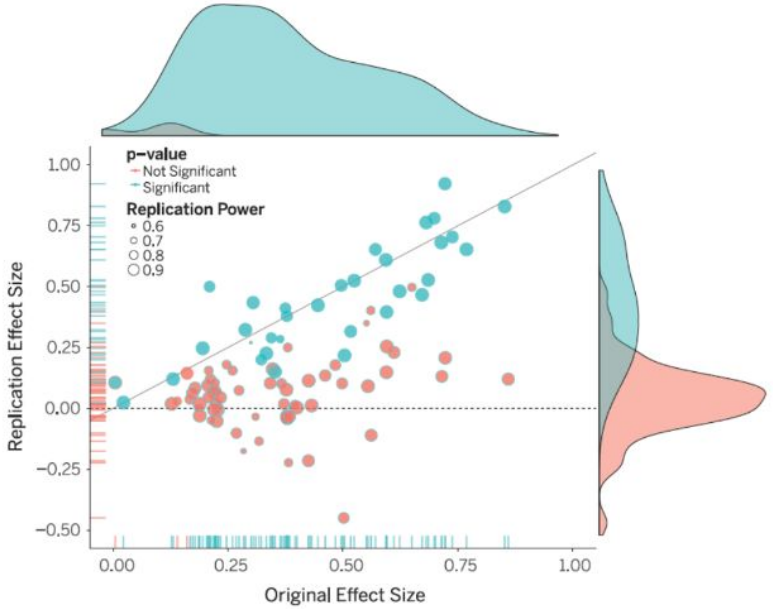
PSYCHOLOGY

Estimating the reproducibility of psychological science (2015)

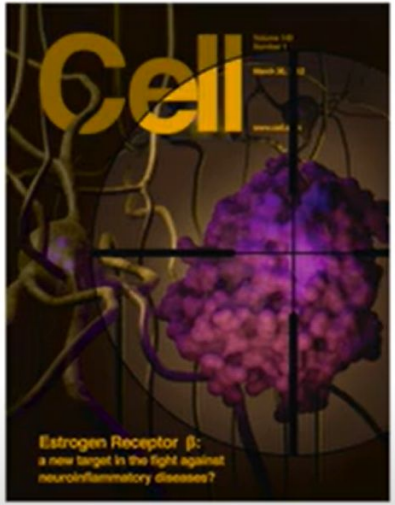
Open Science Collaboration*

100 estudos publicados em 3 revistas de psicologia

- Apenas 36% dos estudos foram capazes de ser reproduzidos



Original study effect size versus replication effect size (correlation coefficients). Diagonal line represents replication effect size equal to original effect size. Dotted line represents replication effect size of 0. Points below the dotted line were effects in the opposite direction of the original. Density plots are separated by significant (blue) and nonsignificant (red) effects.



News

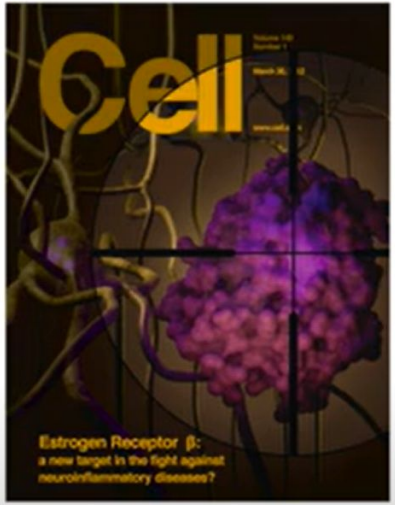
Most laboratory cancer studies cannot be replicated, study shows

BMJ 2012 ; 344 doi: https://doi.org/10.1136/bmj.e2555 (Published 04 April 2012)
Cite this as: *BMJ* 2012;344:e2555

Nigel Hawkes

Author affiliations

Most laboratory studies of cancer are wrong, says a former head of global cancer research at the biotechnology company Amgen. Of 53 "landmark" publications in the literature that Glenn Begley and colleagues attempted to double check, only six could be successfully replicated.



11% (6 de 53 estudos)

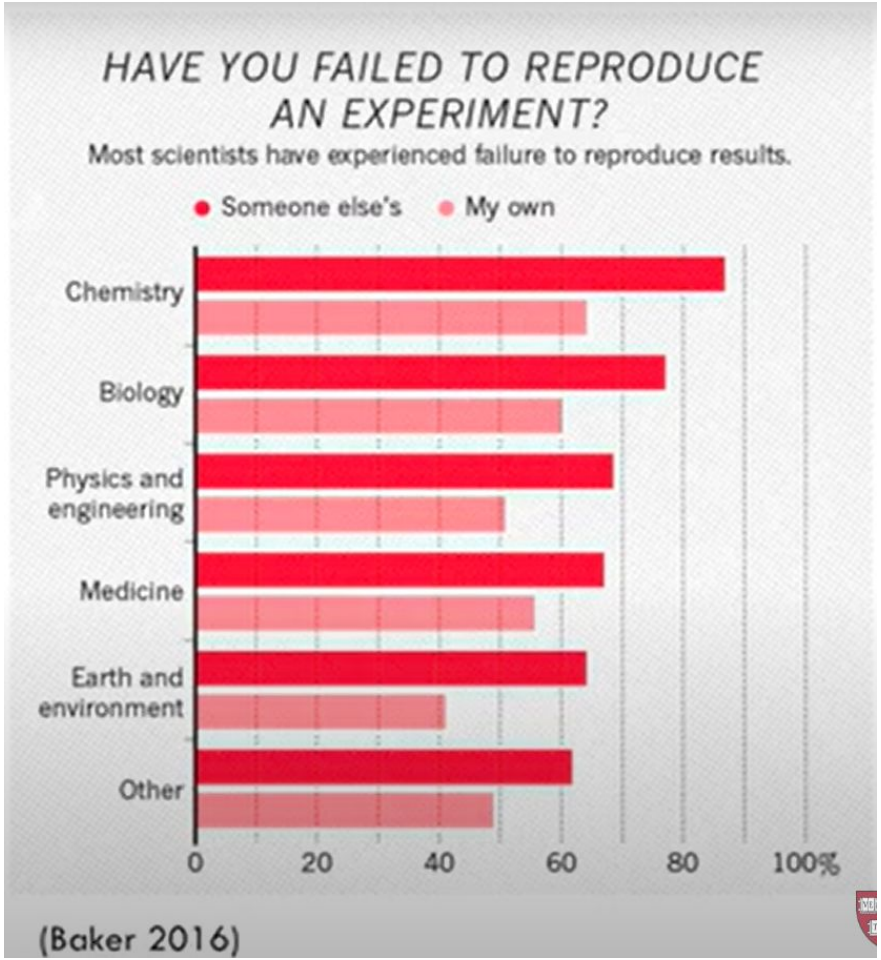
News

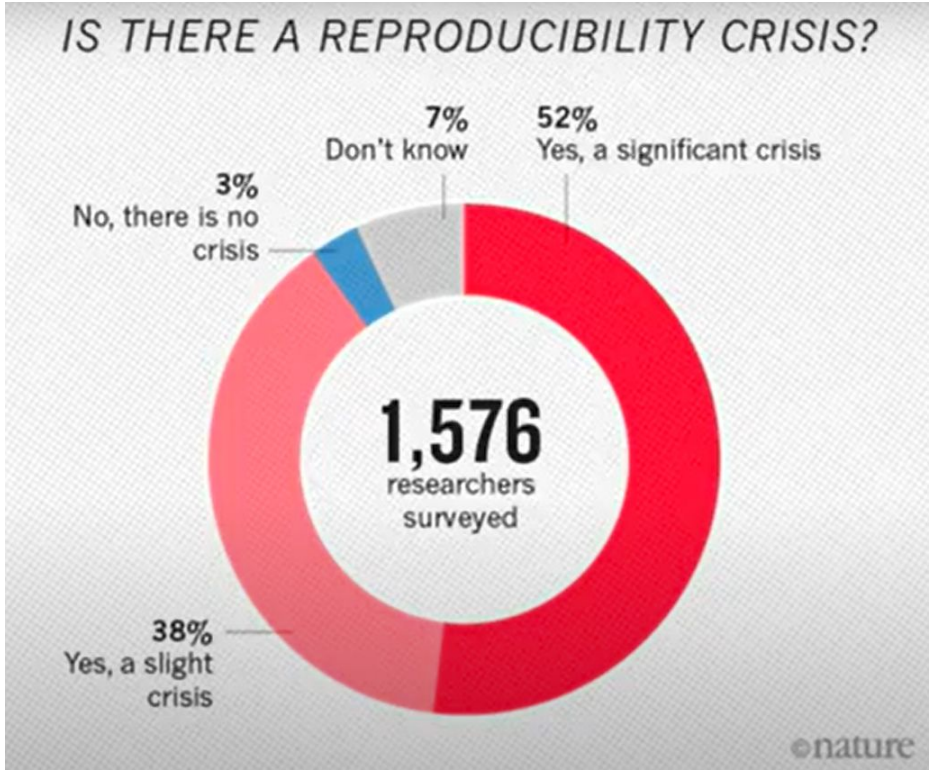
Most laboratory cancer studies cannot be replicated, study shows

BMJ 2012 ; 344 doi: https://doi.org/10.1136/bmj.e2555 (Published 04 April 2012)
Cite this as: *BMJ* 2012;344:e2555

Nigel Hawkes
Author affiliations

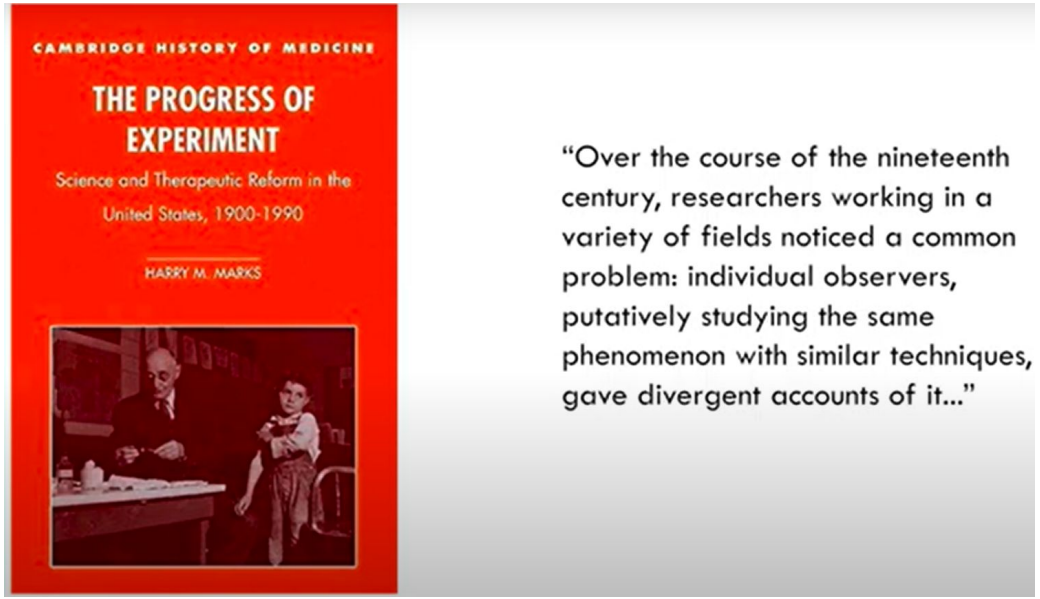
Most laboratory studies of cancer are wrong, says a former head of global cancer research at the biotechnology company Amgen. Of 53 "landmark" publications in the literature that Glenn Begley and colleagues attempted to double check, only six could be successfully replicated.

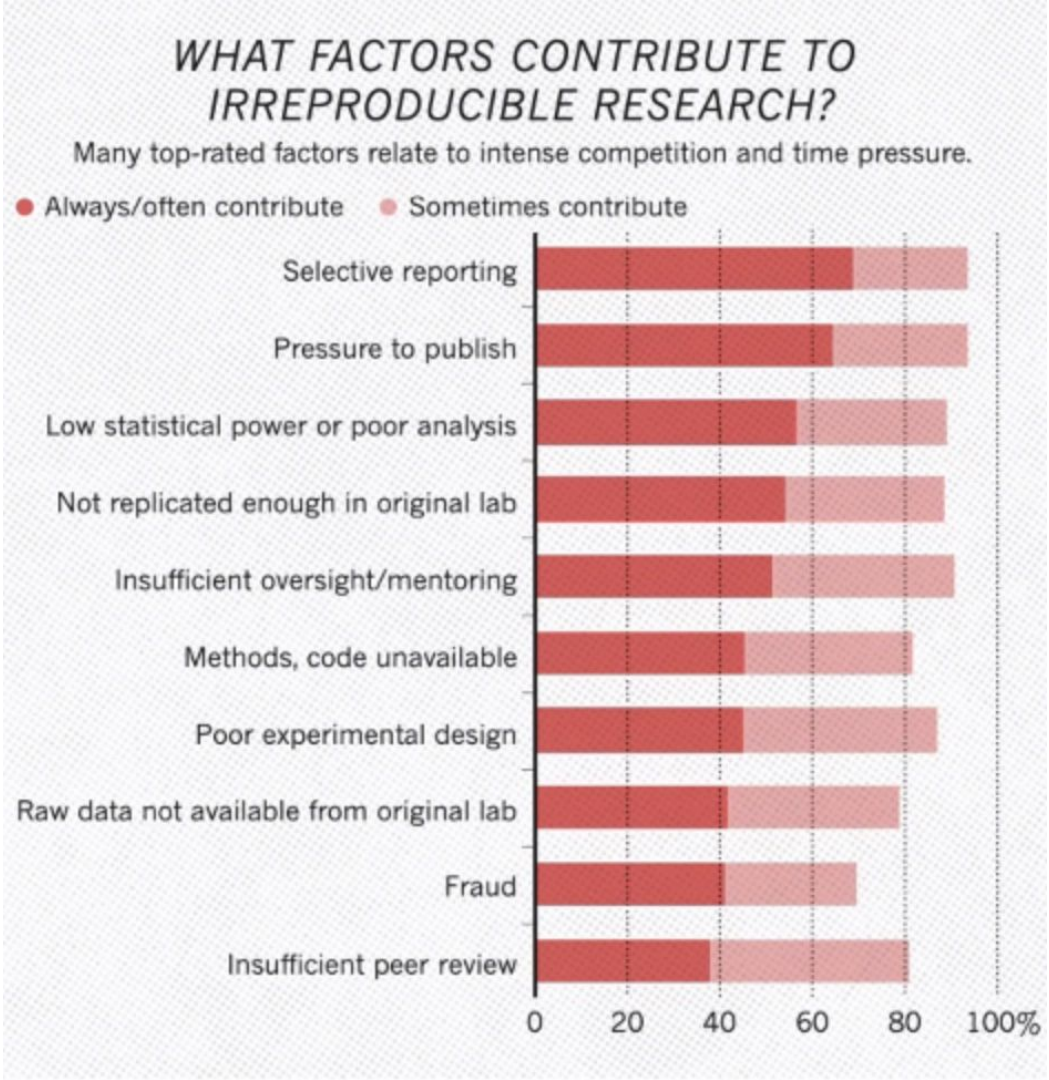

HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?
Most scientists have experienced failure to reproduce results.
Someone else's
My own
Chemistry
Biology
Physics and engineering
Medicine
Earth and environment
Other
0
20
40
60
80
100%
(Baker 2016)


IS THERE A REPRODUCIBILITY CRISIS?
7%
Don't know
52%
Yes, a significant crisis
3%
No, there is no crisis
1,576
researchers surveyed
38%
Yes, a slight crisis
©nature

Como a falta de reprodutibilidade e replicabilidade é algo que todos sabem e tão prevalente?

Como a falta de reprodutibilidade e replicabilidade é algo que todos sabem e tão prevalente?




WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?
Many top-rated factors relate to intense competition and time pressure.
Always/often contribute
Sometimes contribute
Selective reporting
Pressure to publish
Low statistical power or poor analysis
Not replicated enough in original lab
Insufficient oversight/mentoring
Methods, code unavailable
Poor experimental design
Raw data not available from original lab
Fraud
Insufficient peer review
0
20
40
60
80
100%


Literatura tendenciosa
Ecossistema editorial perverso
Problema na gaveta de arquivos
Falha no controle do viés na literatura
Vieses cognitivos
HARKing
Vieses cognitivos
Design de estudo ruim
Baixo poder estatístico
P-hacking
Flexibilidade analítica
Análises inadequadas
Materiais e métodos de documentação incompletos e/ou inacessíveis
Dados inacessíveis e/ou inutilizáveis
Publicar relatório
Pesquise e descubra
Escrever
ideia
Interpretar descobertas
Estudo de projeto
Analisar dados
Adquirir Materiais
Armazenar dados
Coletar dados

BIOLOGICAL REVIEWS
Cambridge Philosophical Society
Biol. Rev. (2017), **92**, pp. 1941–1968. 1941
doi: 10.1111/brv.12315

Detecting and avoiding likely false-positive findings – a practical guide

Wolfgang Forstmeier[1]*, Eric-Jan Wagenmakers[2] and Timothy H. Parker[3]

[1]*Department of Behavioural Ecology and Evolutionary Genetics, Max Planck Institute for Ornithology, 82319 Seewiesen, Germany*
[2]*Department of Psychology, University of Amsterdam, PO Box 15906, 1001 NK Amsterdam, The Netherlands*
[3]*Department of Biology, Whitman College, Walla Walla, WA 99362, U.S.A.*

Table 2. A collection of cognitive biases that may hinder objectivity of researchers. Names and explanations were adopted from Wikipedia (www.wikipedia.org) and inspired by a compilation of 175 cognitive biases by Buster Benson (https://betterhumans.coach.me/cognitive-bias-cheat-sheet-55a472476b18)

Bias	Explanation
Confirmation bias	The tendency to search for, interpret, favour, and recall information in a way that confirms one's pre-existing beliefs or hypotheses, while giving disproportionately less consideration to alternative possibilities
Selective perception	The tendency not to notice and more quickly forget stimuli that cause emotional discomfort and contradict our prior beliefs
Bias blind spot	The cognitive bias of recognizing the impact of biases on the judgement of others, while failing to see the impact of biases on one's own judgment
Confabulation	The production of fabricated, distorted or misinterpreted memories about oneself or the world, without the conscious intention to deceive. This may help us in making sense of what we see
Clustering illusion	The tendency to erroneously consider the inevitable 'streaks' or 'clusters' arising in small samples from random distributions to be non-random
Illusion of validity	A cognitive bias in which a person overestimates his or her ability to interpret and predict accurately the outcome when analysing a set of data, in particular when the data analysed show a very consistent pattern – that is, when the data 'tell' a coherent story
Belief bias	The tendency to judge the strength of arguments based on the plausibility of their conclusion rather than how strongly they support that conclusion. This is an error in reasoning, such as accepting an invalid argument because it supports a conclusion that is plausible
Hindsight bias	The inclination, after an event has occurred, to see the event as having been predictable, despite there having been little or no objective basis for predicting it
Overconfidence effect	A bias in which a person's subjective confidence in his or her judgments is reliably greater than the objective accuracy of those judgments
Appeal to novelty	A fallacy in which one prematurely claims that an idea or proposal is correct or superior, exclusively because it is new and modern

Play

Confirmation bias

nature ecology & evolution

Perspective

https://doi.org/10.1038/s41559-023-01986-1

Positive citation bias and overinterpreted results lead to misinformation on common mycorrhizal networks in forests

Received: 18 August 2022

Justine Karst[1], Melanie D. Jones[2] & Jason D. Hoeksema[3]

VIEWPOINTS

REPUBLISH

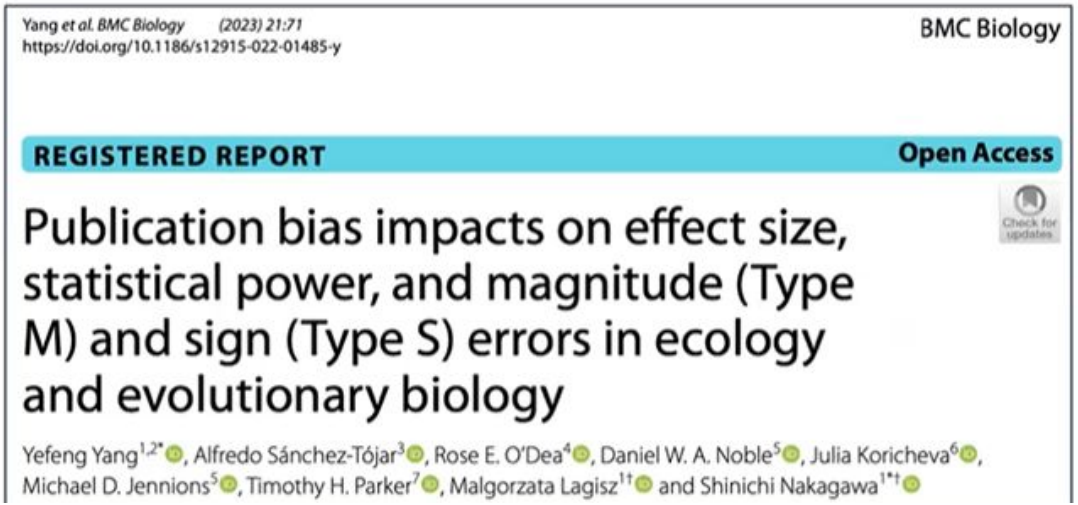
Opinion: Where the 'Wood-Wide Web' Narrative Went Wrong

A compelling story about how forest fungal networks communicate has garnered much public interest. Is any of it true?

Visual: DigitalVision Vectors via Getty Images

BY MELANIE JONES, JASON HOEKSEMA, & JUSTINE KARST 05.25.2023 6 COMMENTS

Low statistical power



Yang et al. BMC Biology (2023) 21:71
https://doi.org/10.1186/s12915-022-01485-y

BMC Biology

REGISTERED REPORT Open Access

Publication bias impacts on effect size, statistical power, and magnitude (Type M) and sign (Type S) errors in ecology and evolutionary biology

Yefeng Yang[1,2*], Alfredo Sánchez-Tójar[3], Rose E. O'Dea[4], Daniel W. A. Noble[5], Julia Koricheva[6], Michael D. Jennions[5], Timothy H. Parker[7], Malgorzata Lagisz[1†] and Shinichi Nakagawa[1*†]

"The statistical power of ecological and evolutionary studies and experiments was found to be consistently low at 15%.

Low statistical power

- Estudos ecológicos em geral possuem baixo poder estatístico
- Tamanhos de efeito são geralmente baixos
- Estudos incluindo interações entre fatores possuem poder ainda menor

Flexibilidade analítica

- Muitas formas diferentes de analisar os mesmos dados para responder uma mesma pergunta
- Mais proeminente em desenhos amostrais complexos
- Idealmente, todos os métodos convergiriam para as mesmas conclusões

aps | ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Journal indexing and metrics

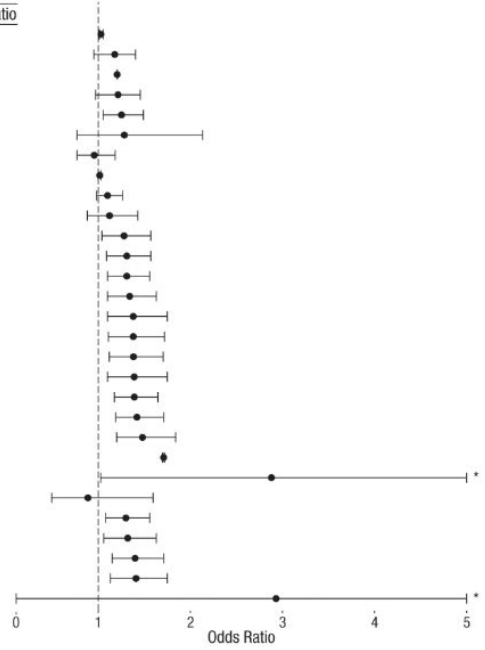
Open access | Research article | First published online August 23, 2018

Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Result

R. Silberzahn, E. L. Uhlmann, [...], and B. A. Nosek (+62) View all authors and affiliations

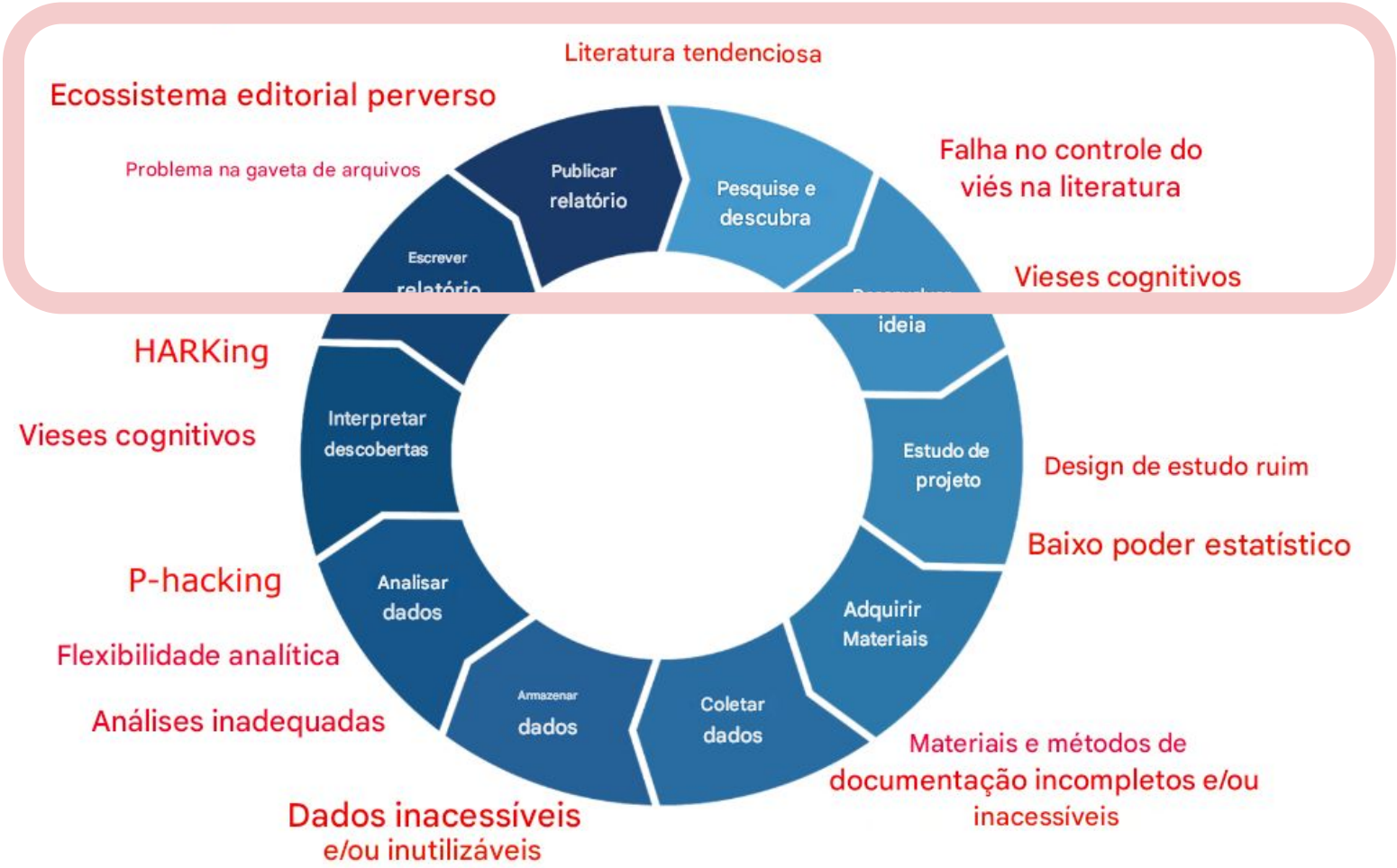
All Articles | https://doi.org/10.1177/2515245917747646 | View correction

Team	Analytic Approach	Distribution	Odds Ratio
10	Multilevel Regression and Logistic Regression	Linear	1.03
1	OLS Regression With Robust Standard Errors, Logistic Regression	Linear	1.18
4	Spearman Correlation	Linear	1.21
14	WLS Regression With Clustered Standard Errors	Linear	1.21
11	Multiple Linear Regression	Linear	1.25
6	Linear Probability Model	Linear	1.28
17	Bayesian Logistic Regression	Logistic	0.96
15	Hierarchical Log-Linear Modeling	Logistic	1.02
18	Hierarchical Bayes Model	Logistic	1.10
31	Logistic Regression	Logistic	1.12
30	Clustered Robust Binomial Logistic Regression	Logistic	1.28
3	Multilevel Logistic Regression Using Bayesian Inference	Logistic	1.31
23	Mixed-Model Logistic Regression	Logistic	1.31
2	Linear Probability Model, Logistic Regression	Logistic	1.34
5	Generalized Linear Mixed Models	Logistic	1.38
24	Multilevel Logistic Regression	Logistic	1.38
28	Mixed-Effects Logistic Regression	Logistic	1.38
32	Generalized Linear Models for Binary Data	Logistic	1.39
8	Negative Binomial Regression With a Log Link	Logistic	1.39
25	Multilevel Logistic Binomial Regression	Logistic	1.42
9	Generalized Linear Mixed-Effects Models With a Logit Link	Logistic	1.48
7	Dirichlet-Process Bayesian Clustering	Misc	1.71
21	Tobit Regression	Misc	2.88
12	Zero-Inflated Poisson Regression	Poisson	0.89
26	Hierarchical Generalized Linear Modeling With Poisson Sampling	Poisson	1.30
16	Hierarchical Poisson Regression	Poisson	1.32
20	Cross-Classified Multilevel Negative Binomial Model	Poisson	1.40
13	Poisson Multilevel Modeling	Poisson	1.41
27	Poisson Regression	Poisson	2.93

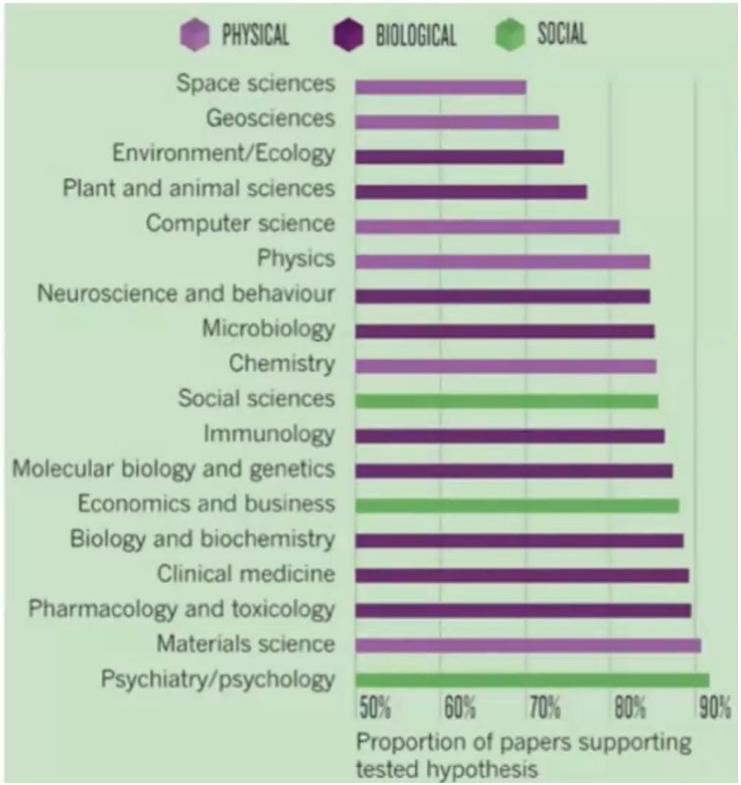


Harking

- Hipotetizar depois que os resultados são conhecidos;
- Encontrar um padrão em um estudo exploratório como se fosse uma pesquisa confirmatória de uma hipótese;
- Taxa de falsos positivos é maior em estudos assim;
- Em geral não se toma nota de todas as decisões feitas;


Literatura tendenciosa
Ecossistema editorial perverso
Problema na gaveta de arquivos
Publicar relatório
Pesquise e descubra
Falha no controle do viés na literatura
Escrever relatório
Vieses cognitivos
ideia
HARKing
Interpretar descobertas
Vieses cognitivos
Estudo de projeto
Design de estudo ruim
Baixo poder estatístico
P-hacking
Analisar dados
Adquirir Materiais
Flexibilidade analítica
Análises inadequadas
Armazenar dados
Coletar dados
Materiais e métodos de documentação incompletos e/ou inacessíveis
Dados inacessíveis e/ou inutilizáveis

Literatura favorece alguns artigos



Práticas questionáveis

PLOS One

Publish About Browse

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Questionable research practices in ecology and evolution

Hannah Fraser, Tim Parker, Shinichi Nakagawa, Ashley Barnett, Fiona Fidler

Published: July 16, 2018 • https://doi.org/10.1371/journal.pone.0200303

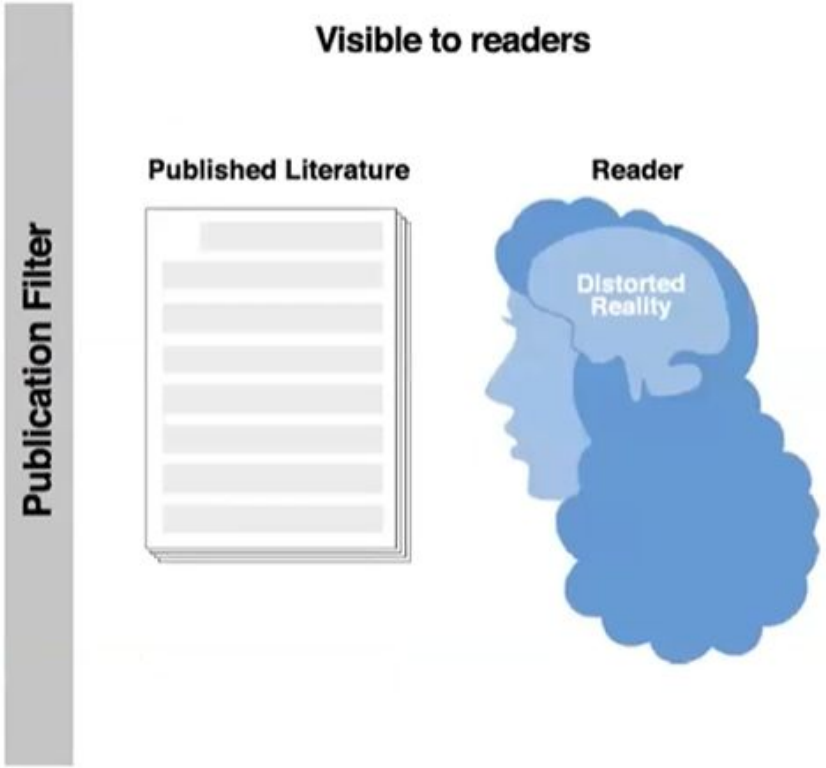
Article Authors Metrics Comments Media Coverage

Práticas questionáveis

Table 2. Percentage (with 95% CIs) of researchers in psychology, ecology and evolution who reported having used each Questionable Research Practice at least once. n = 555–626.

Questionable Research Practice	Psychology Italy Agnoli et al. [16]	Psychology USA John et al. [17]	Ecology	Evolution
Not reporting response (outcome) variables that failed to reach statistical significance#	47.9 *(41.3–54.6)*	63.4 *(59.1–67.7)*	64.1 *(59.1–68.9)*	63.7 *(57.2–69.7)*
Collecting more data after inspecting whether the results are statistically significant#	53.2 *(46.6–59.7)*	55.9 *(51.5–60.3)*	36.9 *(32.4–42.0)*	50.7 *(43.9–57.6)*
Rounding-off a p value or other quantity to meet a pre-specified threshold#	22.2 *(16.7–27.7)*	22.0 *(18.4–25.7)*	27.3 *(23.1–32.0)*	17.5 *(13.1–23.0)*
Deciding to exclude data points after first checking the impact on statistical significance	39.7 *(33.3–46.2)*	38.2 *(33.9–42.6)*	24.0 *(19.9–28.6)*	23.9 *(18.5–30.2)*
Reporting an unexpected finding as having been predicted from the start#	37.4 *(31.0–43.9)*	27.0 *(23.1–30.9)*	48.5 *(43.6–53.6)*	54.2 *(47.7–60.6)*
Filling in missing data points without identifying those data as simulated*	2.3 *(0.3–4.2)*	0.6 *(0.0–1.3)*	4.5 *(2.8–7.1)*	2.0 *(0.8–5.1)*

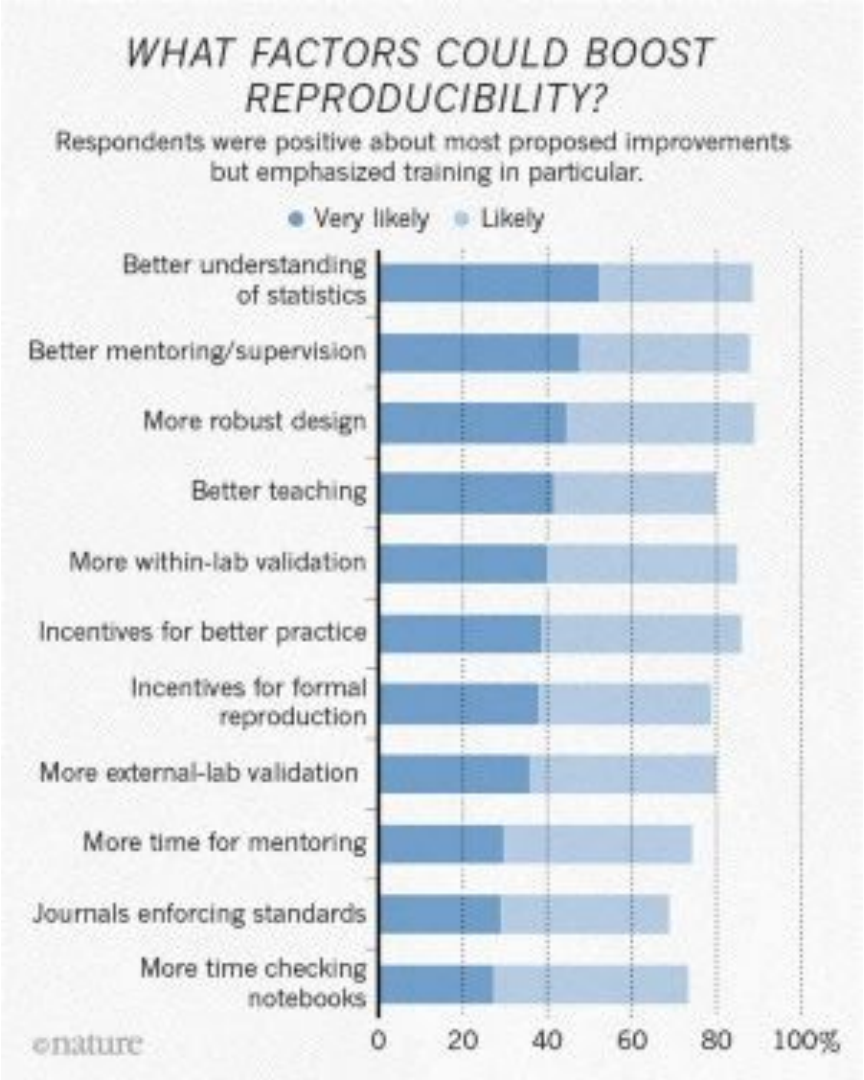
- Bias towards positive results
- File drawer problem
- Power failures yielding high false positive rates and exaggerated effect sizes
- Analytical flexibility
- P-hacking
- HARKing



Ciência Aberta como solução

A pesquisa científica deve ser conduzida e comunicada de forma honesta, acessível e transparente*, de modo que, no mínimo, um estudo possa ser **reproduzido**, mas, idealmente, **replicado**.

A **replicação** fortalece as evidências.


WHAT FACTORS COULD BOOST REPRODUCIBILITY?
Respondents were positive about most proposed improvements but emphasized training in particular.
Very likely
Likely
Better understanding of statistics
Better mentoring/supervision
More robust design
Better teaching
More within-lab validation
Incentives for better practice
Incentives for formal reproduction
More external-lab validation
More time for mentoring
Journals enforcing standards
More time checking notebooks
0
20
40
60
80
100%
©nature

Data archiving

Data archiving

PLOS Biology

Browse Publish About

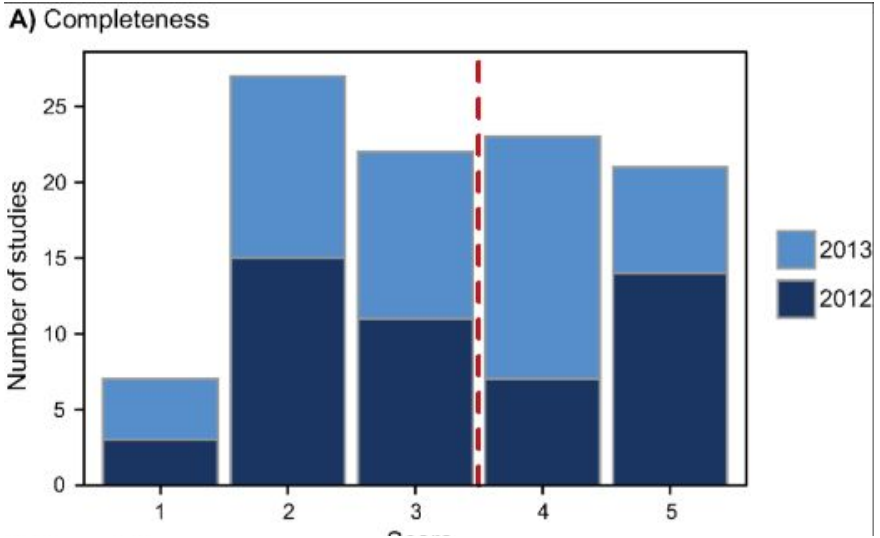
OPEN ACCESS

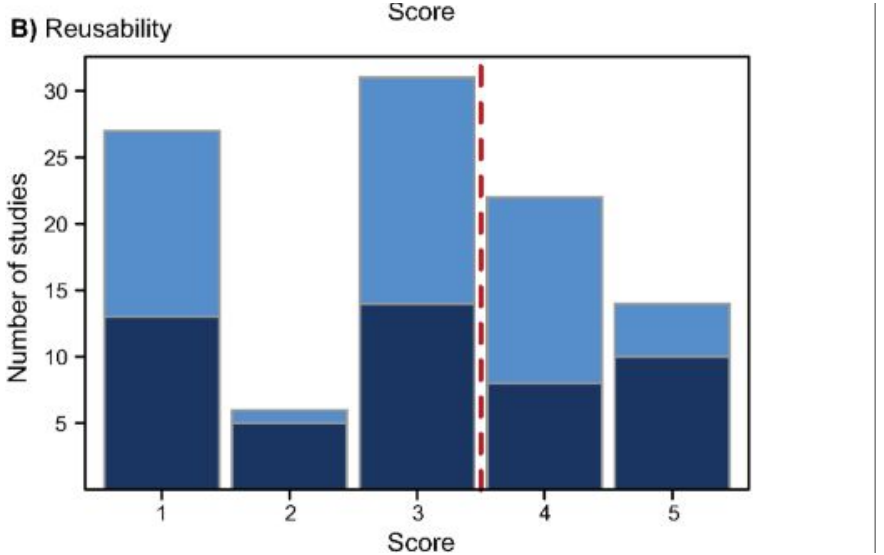
PERSPECTIVE

Public Data Archiving in Ecology and Evolution: How Well Are We Doing?

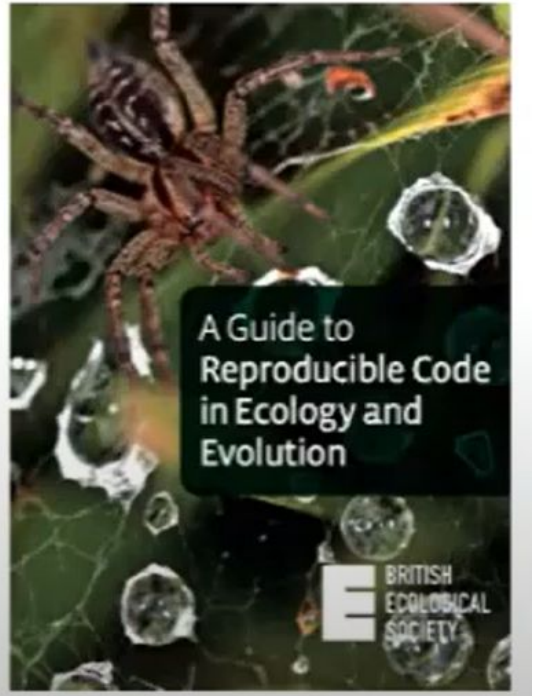
Dominique G. Roche, Loeske E. B. Kruuk, Robert Lanfear, Sandra A. Binning

Published: November 10, 2015 • https://doi.org/10.1371/journal.pbio.1002295





Code archiving



Code archiving

PROCEEDINGS
OF THE ROYAL SOCIETY B

BIOLOGICAL SCIENCES

You have access

Check for updates

View PDF

Tools Share

Cite this article

Biological science practices

The promise of community-driven preprints in ecology and evolution

Daniel W. A. Noble, Zoe A. Xirocostas, Nicholas C. Wu, April Robin Martinig, Rafaela A. Almeida, Kevin R. Bairos-Novak, Heikel Balti, Michael G. Bertram, Louis Bliard, Jack A. Brand, Ilha Byrne, Ying-Chi Chan, Dena Jane Clink, ... See all authors

Published: 29 January 2025 https://doi.org/10.1098/rspb.2024.1487

greater than in many papers published in research journals [32,33]. Despite this, we found that 54.4% ($n = 398$) of articles relying on data (i.e. classified as 'research articles') on *EcoEvoRxiv* did not share data, and 58.1% ($n = 425$) did not share code.

What can you do?

Pre-Project checklist

- create GitHub account and ensure Git installed on local computer
- create ORCID account
- create OSF account
- create Zotero account, install Zotero desktop app and Zotero connector in browser
- install "R"
- install "RStudio"

PROJECT SETUP

RStudio/ GitHub

- create GitHub repository for the project (can optionally be private for now)
- create new RStudio project, enabling version control by connecting to GitHub repo
  - enable 'renv' option to facilitate package management
- set up appropriate directory structure (can automate this with R script)
- initialize a _README.md file for project root directory (can be fleshed out later)

OSF

- create new OSF project (or component on existing project) for managing this project, materials and for serving as a central 'hub' for all project components (e.g. github, google drive etc...)
- bookmark the project in your browser
- include all collaborators as contributors, order names appropriately, tailor their permissions
- provide useful keywords/tags to enhance discoverability
- edit the 'metadata' entries appropriately
- include an informative project description on home Wiki page
- use another Wiki page to provide estimated project timeline
- this OSF project can be linked to pre-registration when it is submitted (see below)
- [OPTIONAL] create component for any administrative materials (restricted permissions)
- [OPTIONAL] connect GitHub repo to OSF project

Zotero

- create a Zotero library (or shared group library if collaborating)
  - store attachments (PDFs) on LOCAL or OneDrive (not on Zotero cloud)

- Initiate a Data Management Plan (see https://assistant.portagenetwork.ca)

PREREGISTRATION

Part I: Preparation

- lots of reading to formulate rough ideas / questions / hypotheses
- devise preliminary study / experimental design (see next step)

Part II: Analysis to inform pre-registration [OPTIONAL]

- create annotated R Markdown script that walks through the process of generating simulated data for the purpose of working out the study design and analysis plan
- can use data from previous research to inform data generation
- prospective power analysis to inform study design
- Finalized analysis plan, with example code for simulated data, can be linked in pre-registration (either through GitHub, or via HTML or PDF file)

Part III: Submission

- complete & submit OSF pre-registration & optionally link to OSF project
  - optionally include knitted document showing power analysis etc...
- the pre-registration can be kept private for a time period, and this is recommended if you wish to retain anonymity during review process; you can create an anonymized link to the pre-registration for use during review process
- bookmark the registration in your browser (it will issue a DOI)

- Finalize data management plan and store in OSF project

ANALYSIS / WRITING

RStudio/ GitHub

- regular versioning (push/pull) of coding work to GitHub
- ensure that only non-sensitive data and those with acceptable license get stored on GitHub
- any sensitive data only store locally, backed up on private OSF component
- beware of storage limitations on GitHub & OSF
- keep the 'renv' lock file up to date for package management

[OPTIONAL]

- R Markdown + "truckdown" package for collaborative writing via google docs
  - using citation tools within RStudio (e.g. Zotero within Visual editing mode)
  - versioning via push/pull to GitHub

OSF

- regular backup / versioning of binary files (e.g. Word docs)
- use Wiki pages to provide progress updates (referring to timeline Wiki)
- use other Wiki pages for any other relevant details
- use as location for backup storage of data
- [OPTIONAL] use 'check-out'/'check-in' feature for collaborating on documents
- [OPTIONAL] keep all admin materials up to date

Zotero

- keep Zotero library up to date and clean regularly

- regularly check project reproducibility (can ask labmate to help)

* Follow best practices for sensitive data

PROJECT DISSEMINATION

RStudio/ GitHub

- knit the manuscript to PDF (journal may have direct markdown submission option)
- ensure it rendered correctly, including symbols, figures, formatting, citations etc.
- create _DATA-DICTIONARY.md file for both rawdata and outputs directories and ensure _README.md files in root and rawdata directories correct

OSF

- create new OSF private project that will eventually serve as the project archive
- add all the relevant information as described in "Project Setup"
  - metadata, wiki pages, etc... (but don't put info in wiki pages that could be used to identify authors)
- copy all directories and files (excluding any sensitive data) from local RStudio project folder (repo) into a SIGNALE folder on the main OSF project storage (this makes it easier for someone to download materials)
- in settings, create a "view only" link, and select option to anonymize
  - this link is what you'll provide with manuscript submission to journal and in preprint
- use different browser to double-check anonymity of OSF project

- double-check project reproducibility from OSF materials (can ask labmate to help)

Submission

- be sure to state that the study was pre-registered, and provide anonymized link
- explain any deviations from pre-registered plans
- [OPTIONAL] submit PREPRINT to a preprint server like BioRxiv/Arxiv or OSF Preprints
  NOTE: this will preclude anonymity during review process
- submit manuscript preferably to OPEN ACCESS journal (but if not, simply ensure the journal you submit to allows posting of a non-copy edited version)

Revisions

- complete revisions using versioning as usual
- update all project materials accordingly

PAPER ACCEPTED!

- include the DOI for the OSF archive (next step) in the final submission!

PROJECT ARCHIVING

OSF

- make the OSF archive project public
- note it will create a permanent DOI for this archive
- choose appropriate copyright license
- ensure contributor (author) order on the archive is appropriate for the work done on the archived materials, e.g. scripts / data (not the main manuscript itself)
- include a clear statement illustrating how the archive should be cited
  NOTE: this archive should have its own unique citation (separate from main paper)
- ensure all the metadata are complete
- explain any sensitive data issues
- include link to published paper
- include link to GitHub repo
- ensure all components are view/read only
- maximize discoverability with useful tags
- can use home Wiki page as the landing page, and govern the content (e.g. link to video abstract, posters, etc...)
- can use OSF analytics to track site visits and downloads
- can "link" this archive to the original OSF project
- optionally link the Zotero library to the archive
- provide a link to the public Zotero library associated with paper
- bookmark this archive in your browser

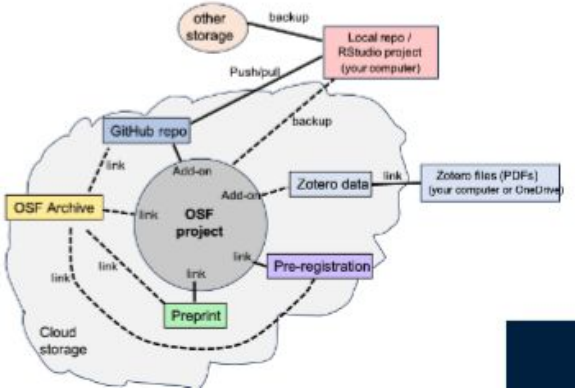
Borealis Dataverse [OPTIONAL]

- additionally create a dataverse archive through your institution
- alternatively, rather than archiving the project materials, simply enter the metadata and then link to the OSF archive

CLEANUP

OSF

- remove any unnecessary files/components from your OSF projects
- make sure all URL links are working
- in the home Wiki page of the hub project, put a clear statement of project status near the top (e.g. indicating the project was wrapped up, with date)
- sign out of OSF and double check that only those files/components that should be publicly visible are.



What you can do:

Update CV to showcase your efforts around Open Science

Become more proficient at identifying QRPs in the literature

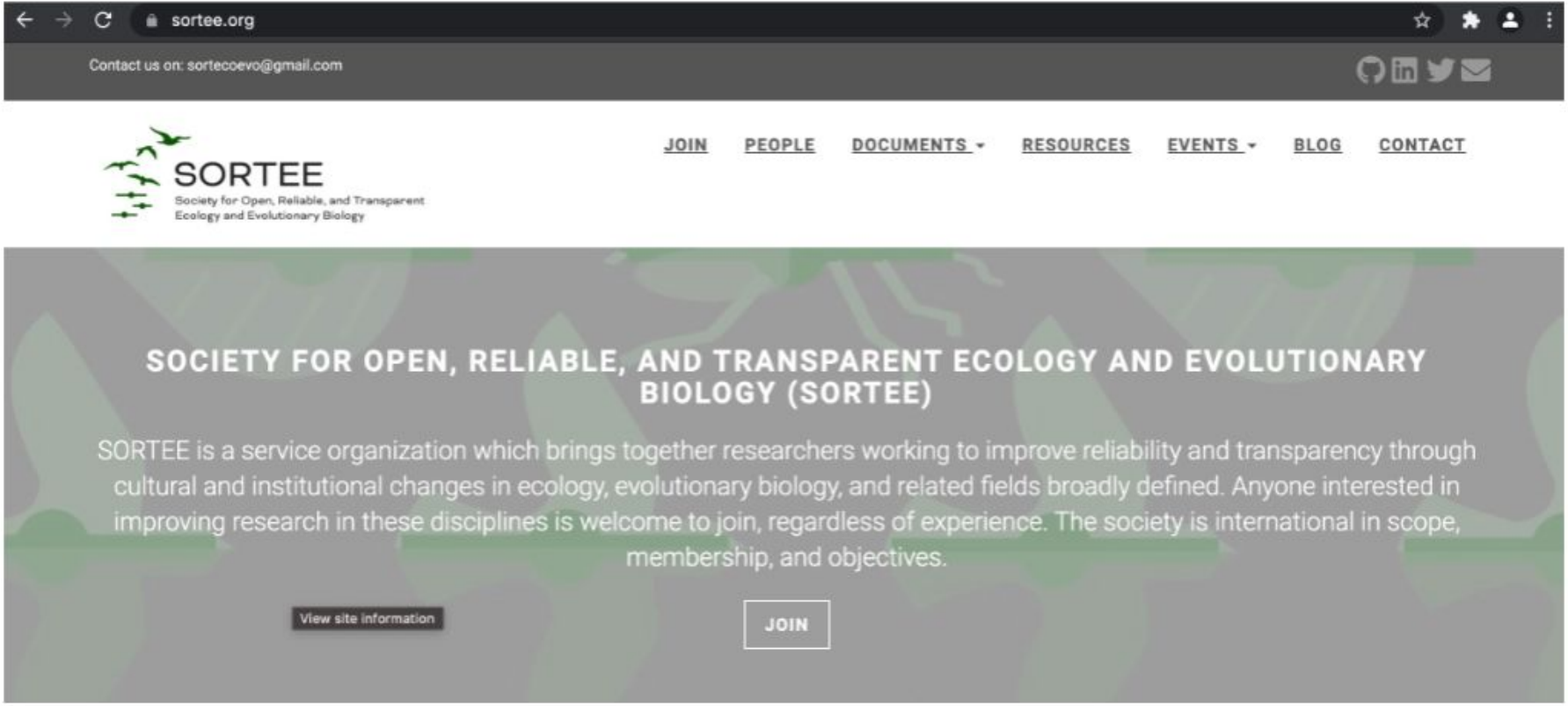
- Maximize accessibility to relevant audiences
- Maximize discoverability
- Submit a pre-print

- Be aware of cognitive biases
- Think of all possible interpretations

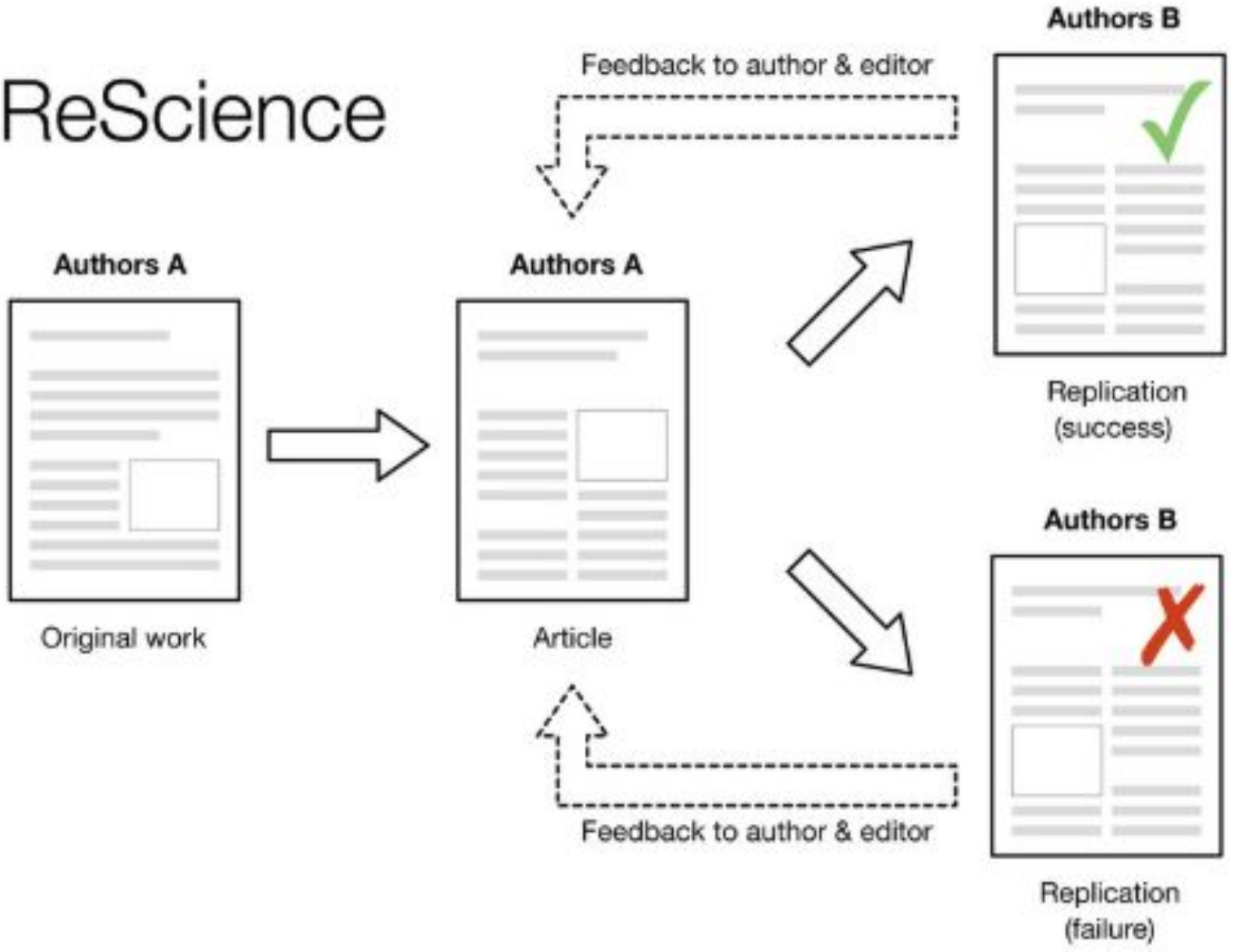
- Implement a reproducible workflow
- Take advantage of available tools
- Use version control

- Build expertise in stats and study design
- Play with fake / existing data
- Think carefully about who might be interested in your research, and who it might impact; modify design, data/access/communication plans accordingly
- Use helpful checklists
- Data Management Plan
- Pre-register your study

Join societies that help promote open science



A. ReScience


Authors B
Feedback to author & editor
Authors A
Authors A
Replication
(success)
Original work
Article
Authors B
Feedback to author & editor
Replication
(failure)

Coffee break

Reprodutibilidade em Ecologia

Reprodutibilidade em Ecologia

Ao utilizar Modelagem de nicho climático:

nature ecology & evolution

PERSPECTIVE

https://doi.org/10.1038/s41559-019-0972-5

OPEN

A checklist for maximizing reproducibility of ecological niche models

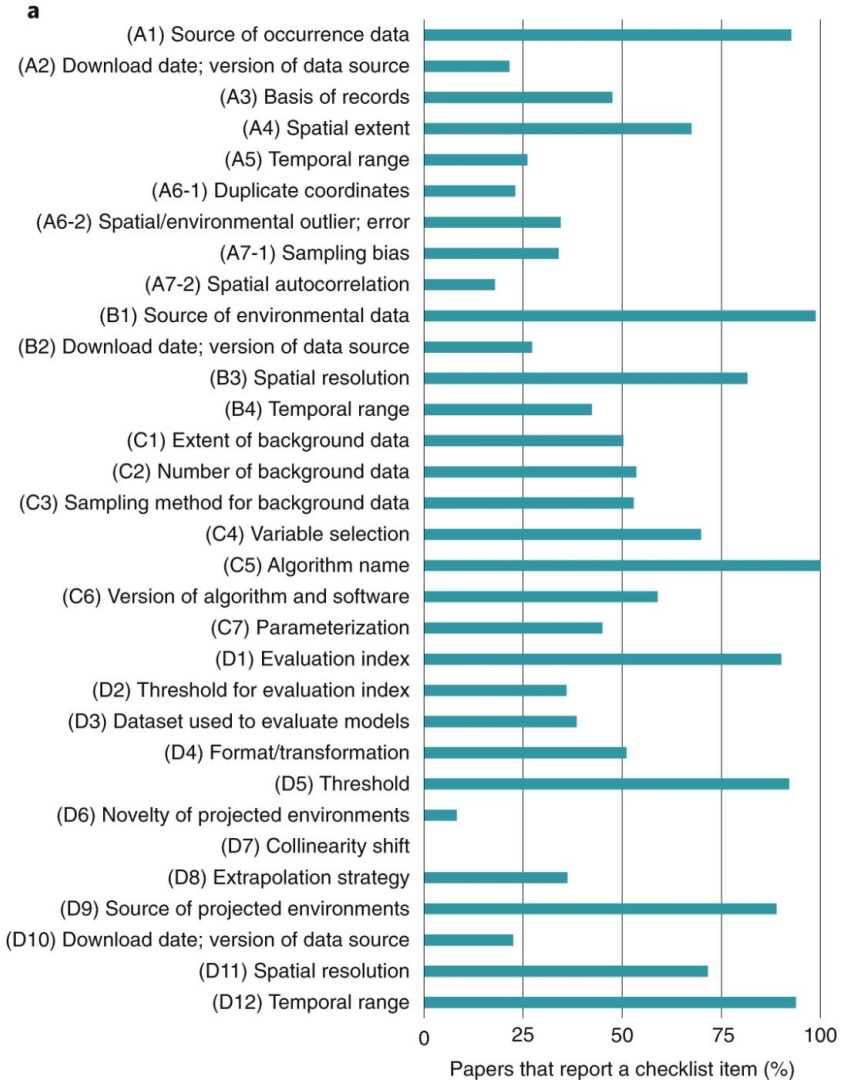
Xiao Feng[1,2,8*], Daniel S. Park[3,8], Cassondra Walker[4], A. Townsend Peterson[5], Cory Merow[6] and Monica Papeș[7]

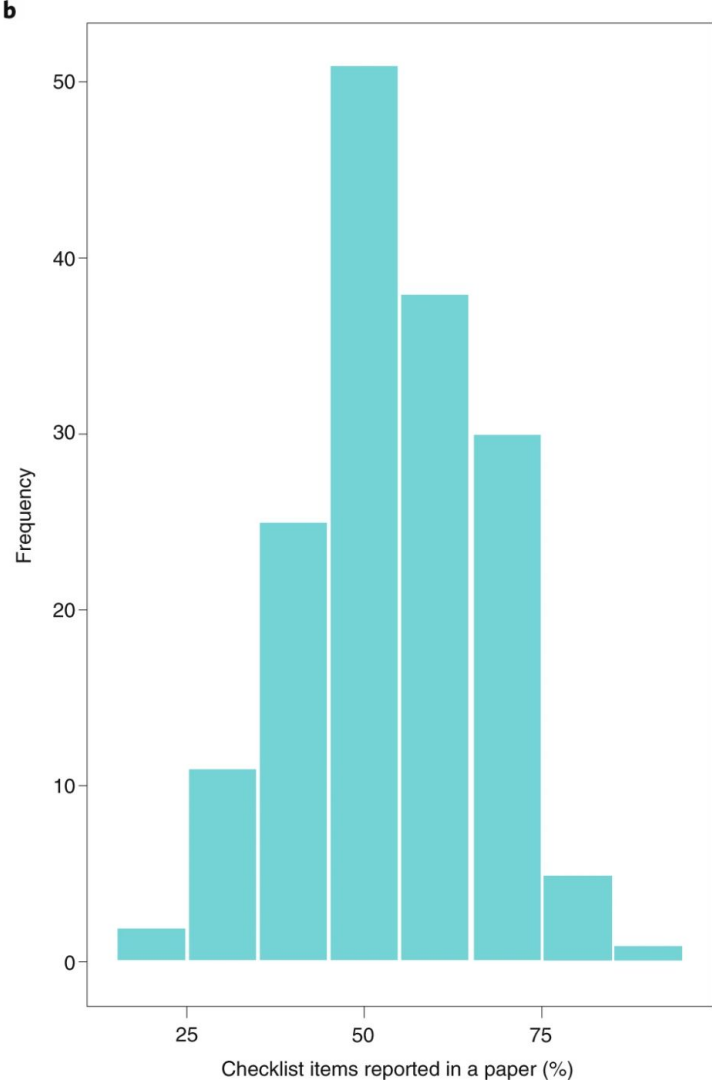
(A) Dados de ocorrência

(B) Dados ambientais

(C) Calibração do modelo

(D) Transferência e avaliação do modelo





Reprodutibilidade em Ecologia

Ao utilizar Modelagem de nicho climático:

1) Obtenção dos dados de ocorrência

(A1) Source of occurrence data	Reporting occurrence data sources allows one to assess data quality and trace/correct any possible issues that may be detected.	"Species distribution records were collected from the Ocean Biogeographic Information System (OBIS; http://iobis.org, accessed February 2016), from the Global Biodiversity Information Facility (GBIF; http://gbif.org, accessed January 2016), the Reef Life Survey (RLS; http://reeflifesurvey.com, accessed February 2016) and for a few species via personal communications."[109]	NA	93
(A2) Download date; version of data source	Databases and datasets change over time.	"Occurrences were downloaded from GBIF.org on 28 January 2016 (https://doi.org/10.15468/dl.iou7qq)."[110]	NA	22

Reprodutibilidade em Ecologia

Ao utilizar Modelagem de nicho climático:

1) Obtenção dos dados de ocorrência

(A3) Basis of records	Biodiversity databases comprise many different types of data, each with specific uses and caveats. Relevant distinctions include whether data are collected opportunistically, as part of structured surveys, as part of repeated surveys, as part of comprehensive checklists of co-occurring species, by scientists, by citizen scientists and so on.	"Before fieldwork, we obtained locality information from *C. canescens* herbarium specimens and online biodiversity databases such as the Southwest Environmental Information Network and the Rocky Mountain Herbarium (University of Wyoming). In addition, the Rocky Mountain Herbarium and the Colorado State University Herbarium were visited to examine potentially misidentified specimens from outlying portions of the species' distribution."[111]	112–114	48

Reprodutibilidade em Ecologia

Ao utilizar Modelagem de nicho climático:

1) Obtenção dos dados de ocorrência

(A4) Spatial extent	Spatial extent of occurrences is crucial for interpretation of model predictions, including whether potential sink populations are included, whether sampling is biased or whether records outside the native range are used.	"We integrated missing countries by obtaining occurrences from the literature — that is, for France, Italy and Switzerland. To increase the accuracy of the analysis, we excluded the following records: (1) localities for which we were not able to obtain precise coordinates; [...] (4) record of *M. bourneti* in the Canary Islands, due to taxonomical issues currently unresolved (C. Ribera, personal communication, 2016)."[115]	116,117	67
(A5) Temporal range	Environments can change over time, thus the timestamp of occurrence records is crucial for linking them to the relevant environmental conditions experienced by the species, and hence correctly describing the niche.	"Although the sightings dataset extended over 257 years, 79% of sightings occurred between 2000 and 2015. Therefore only this subset of 5,419 sightings was retained for further analysis. These sightings were divided into each quarter of the year (Jan–Mar, Apr–May, Jun–Aug and Sep–Dec) and matched with recent climate data available through online data sharing platforms."[118]	70	26

Reprodutibilidade em Ecologia

Ao utilizar Modelagem de nicho climático:

2) Processamento de dados de ocorrência

(A6-1) Duplicate coordinates	Duplicated coordinates can potentially bias model training. Also, different modelling algorithms may have different default options for handling duplicated coordinates, either at point level or at pixel level.	"We constructed potential distributions for each species in the program Maxent 3.3.3k (Phillips et al., 2006) using the default settings, including removing duplicate species records from the same grid square."[119]	120	23
(A6-2) Spatial and environmental outliers; error	Outliers or errors may lead to model errors. Also, the model prediction may be sensitive to outliers or errors.	"Finally, we plotted all the points on maps and excluded any point falling far outside the proven distribution described in Krapovickas et al. (2007)."[121]	116	35

Reprodutibilidade em Ecologia

Ao utilizar Modelagem de nicho climático:

2) Processamento de dados de ocorrência

(A6-3) Spatial and coordinate uncertainty	The coordinates of a record may not represent the exact location of collection. Coordinates are often recorded or processed to different degrees of specificity (for example, two decimal points versus four). Further, coordinates are often georeferenced from locality descriptions to, for example, centroids of political boundaries. The mismatch between the coordinate uncertainty and spatial resolution	"For this study, precise locality coordinates for *P. solenopsis* were not available, so the district-level occurrence data published by Nagrare et al. (2009) were used ($n = 42$ records). The centroid method may be acceptable if the target scale of prediction is global but may not be appropriate at national, state or finer scales; districts are not homogeneous, and some of them can be quite large. We calculated district-level averages of climatic variables in ArcMap (version 9.3, ESRI, Redlands, CA, USA) and used those as predictors. This is a relatively unconventional use of ENM/SDM, and the results may be useful for	33,40–43	NA

Reprodutibilidade em Ecologia

Ao utilizar Modelagem de nicho climático:

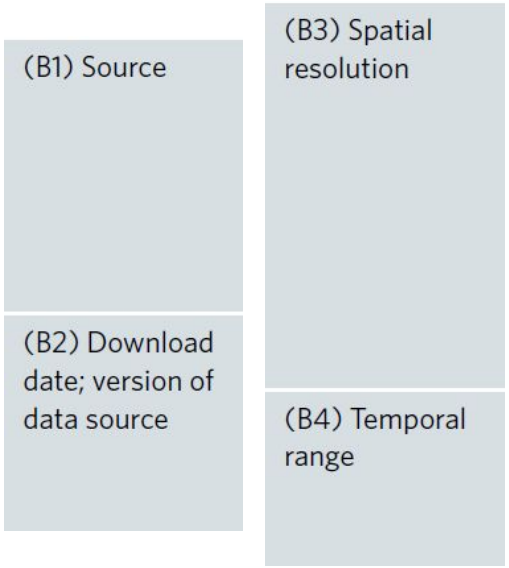
2) Processamento de dados de ocorrência

(A7-1) Sampling bias	Biased sampling, unequal sampling of a species' distribution, may cause the model to overfit environmental conditions associated with such samples. Also, different algorithms may have different default methods for handling spatially biased occurrences.	"To reduce the effects of sampling bias, we spatially filtered the occurrence dataset to ensure that no two localities were within 10 km of one another."[124]	50–52, 120, 125–129	34
(A7-2) Spatial autocorrelation	Spatial autocorrelation, here referring to the non-independent spatial distribution of occurrences, could violate the modelling assumption of independent and identical residuals, thus could bias estimations of model parameters.	"In order to account for autocorrelation in the observations, models were also fitted in which contagion (see below: spatial interpolators) was included as an autocovariate term in the initial variable set (AGLM). These models are termed autologistic (Smith, 1994; Augustin et al., 1996; Araújo & Williams, 2000). Measures of aggregation for point	53,127, 131–133	18

Reprodutibilidade em Ecologia

Ao utilizar Modelagem de nicho climático:

3) Obtenção dos dados ambientais



Reprodutibilidade em Ecologia

Ao utilizar Modelagem de nicho climático:

4) Calibração do modelo

(C1) Modelling domain	The geographic domain of a model has to be specified because it is associated with the underlying assumptions of the relationship between species' distribution and the environments, as well as background selection for some ENM algorithms.	"In the second approach, locality data were overlaid on terrain base maps in ArcGIS 10.2 (Environmental Systems Research Institute, 2011) together with a world ecoregions layer (World Wildlife Fund, 2011). These were used to identify breaks in habitat and ecological regions in topographically homogeneous areas. [...] Restricting calibration areas to regions bounded by significant abiotic barriers (for example, large rivers, mountain ranges) and known or hypothesized dispersal distances yielded more accurate models and reduced these errors (Barve et al., 2011; Owens et al., 2013; Royle, Chandler, Yackulic, & Nichols, 2012; Saupe et al., 2012). Thus, in our study, Ms were constrained by deep valleys (for example, the Maranon Valley), the crests of mountains (for example, the Andes) and ~other distinct features likely to act as barriers to species distributions (for example, the llanos of northern South America)."[137]	52,71, 72,74, 138–140	50

Reprodutibilidade em Ecologia

Ao utilizar Modelagem de nicho climático:

4) Calibração do modelo

(C2) Number of background data	Background data are assumed to represent the environmental composition of species' accessible area, thus the optimal number of background data may depend on the extent of the study area and resolution of environmental data, as well as computation capacity.	"For each geographical background we selected 10,000 random cells that did not hold a species presence record (or all available cells if fewer than 10,000 were available)."[141]	73,142	54

Reprodutibilidade em Ecologia

Ao utilizar Modelagem de nicho climático:

4) Calibração do modelo

(C3) Sampling method for background data	Random selection of background data has been used as the default strategy in some algorithms, but new methods have been developed for different purposes.	"We used Maxent with default settings, except that we applied a targeted background sampling to reduce the influence of sample selection bias (Phillips et al., 2009) by using 666 vertebrate fossil site localities (excluding moa bones [Order Dinornithiformes] and swamp sites) throughout New Zealand as background points."[143]	72,73, 75,76, 142,144, 145–147	53
(C4) Variable selection	Selection of variables is biologically and/or statistically relevant, thus criteria and justification are needed.	"Four 'bioclimatic' layers were used to calibrate models: mean temperature of the warmest quarter, mean temperature of the coldest quarter, precipitation of the wettest quarter, and precipitation of the driest quarter. These four layers were chosen because they represent the climatic extremes that often constrain species distributions and because most	77,87,148	70

Reprodutibilidade em Ecologia

Ao utilizar Modelagem de nicho climático:

5) Algoritmo

(C5) Name	Reporting the name of modelling algorithm is the basis of reproducibility.	"ENM was performed using the maximum entropy approach as implemented in MAXENT 3.3.3k (Phillips, Anderson, & Schapire, 2006)."[149]	NA	100
(C6) Version of algorithm and software	Modelling algorithm, default settings, and dependent libraries can change over time, so providing the version will enhance the reproducibility of a study.	"For BRTs, different combinations of learning rates (0.005, 0.01, 0.05) and tree complexity (1, 2, 3) were tested. Folds were set at random and other parameters were left as default in the gbm R package (version 2.1.1). Runs on R version 3.3.2."[150]	17,151,152	59
(C7) Parameterization	Parameter or modelling settings can influence the resulting model, and default settings may not be appropriate for a study. Thus specific settings should be reported, including default ones.	"Selecting the best settings for the regularization multipliers and number of feature classes, which determine the model complexity, requires quantitative evaluation (Merow et al., 2013). The optimal model parameters were tuned using the function ENMevaluate in the package 'ENMeval' (Muscarella et al., 2014) for R. Within ENMevaluate, we jackknifed	72,81,83, 154–157	45

Reprodutibilidade em Ecologia

Ao utilizar Modelagem de nicho climático:

6) Transferência de modelo e avaliação

(D1) Evaluation index	Proper understanding of model performance requires the use of model evaluation indices. Also, different evaluation indices may be informative of different aspects of a model.	"We evaluated the performance of the models by three different methods using an independent dataset of occurrences for model evaluation: (a) an omission error test [...] (b) the binomial cumulative probability [...] (c) the partial receiver operating characteristic [...]"[158]	89,90, 159–165	90
(D2) Threshold for evaluation index	Calculation of some evaluation indices requires a threshold. The threshold will vary by study because there is no single, default method for choosing a threshold.	"A threshold to convert continuous predicted probabilities into a binomial output was estimated for each model run, using the threshold value that maximized specificity (true negative rate) and sensitivity (true positive rate) over the evaluation dataset predictions (Liu, Newell, & White, 2016).	92,93,166	36

Reprodutibilidade em Ecologia

Ao utilizar Modelagem de nicho climático:

6) Transferência de modelo e avaliação

(D3) Dataset used to evaluate models	Evaluation of a model is usually based on another independent dataset, or part of the dataset not used in model training. The choice of dataset can influence the evaluation results and the subsequent interpretations.	"All these methods used observed presences as input with a 70% random sample for model development and the remaining 30% sample for model evaluation."[167]	81,91,168	39

Reprodutibilidade em Ecologia

Ao utilizar Modelagem de nicho climático:

7) Output

(D4) Format/ transformation	The raw model predictions are sometimes transformed (for example, logistic transformation) via different methods under different assumptions.	"[...] we used the logistic output format [...]"[124]	80,169,170	51
(D5) Threshold	Often, the model predictions are in continuous format, which is subsequently transformed into a binary prediction under a particular threshold. Researchers have proposed different ways of thresholding for different purposes and under varied assumptions.	"We repeated this procedure 20 times for each algorithm and used the Lowest Present Threshold values (Pearson et al., 2007) to transform each map in binary."[171]	92,93,172	92

Reprodutibilidade em Ecologia

Ao utilizar Modelagem de nicho climático:

8) Extrapolação

(D6) Novelty of projected environments relative to training environments	Transferring a model across space and/or time may lead to extrapolation if the projected environments are novel compared with training environments. Quantification of novel environments could help understand the uncertainties associated with model predictions.	"To assess the effect of model extrapolation on values of predictor variables lying outside the training range, that is, projecting models on non-analogous climates (cf. Nogues-Bravo, 2009), we conducted a multivariate environmental similarity surfaces (MESS) analysis, following Elith et al. (2011)."[173]	48,95,97, 174,175	8
(D7) Collinearity shift between training and projected environments	Transferring a model outside training data may be affected by differences in collinearity structure between training and projection environments, which can lead to degraded prediction performance. Therefore, quantification of collinearity shift or any steps towards correcting for it should be specified.	"We compared the correlation matrix of the 6 variables in the training region to the average of the correlation matrices of present and future climate layers in the projected area (Tables S2 & S3 in Supplement 2). The highest absolute change of *r* was 0.3 for bio4 and bio17, and *r* increased above the 0.7 threshold for 2 pairs of variables (−0.78 for bio3 and bio4; 0.71 for bio16 and bio17; Supplement 2)."[176]	77,96,177	0

Reprodutibilidade em Ecologia

Ao utilizar Modelagem de nicho climático:

8) Extrapolação

(D8) Extrapolation strategy	Model extrapolation is statistically challenging. Different extrapolation strategies can lead to very different model predictions, therefore the choice of extrapolation, even the default setting of an algorithm, should be provided.	"Five replicates of each model were conducted with no clamping or extrapolation and with all the default 'features' used."[137]	94,97, 98,178	36

Reprodutibilidade no seu trabalho

- Fazer subgrupos de 3 ou 4 pessoas
- Selecionar o trabalho de uma das pessoas
- Listar todas as informações que devem ser listadas para tornar o trabalho reprodutível e os motivos