



Beyond Boolean Search: Tools to identify comprehensive and inclusive scientific citations

Anna R. Willoughby, Kaylee Arnold, Donat Agosti, Jorrit Poelen,
Aja C. Sherman, Nancy Simmons, Nathan S Upham, Maxwell Farrell

Digital Libraries Forum
St. Louis, Missouri, USA
13 November 2023

Follow Along

Slides at DLF Website

Data sets and Resource document at:
<https://github.com/arw36/dlf-workshop-2023>



Workshop Objectives

01 Basics and Barriers of Systematic Searches

02 Case Studies Of Hybrid Approaches

Hello from the Natural Sciences

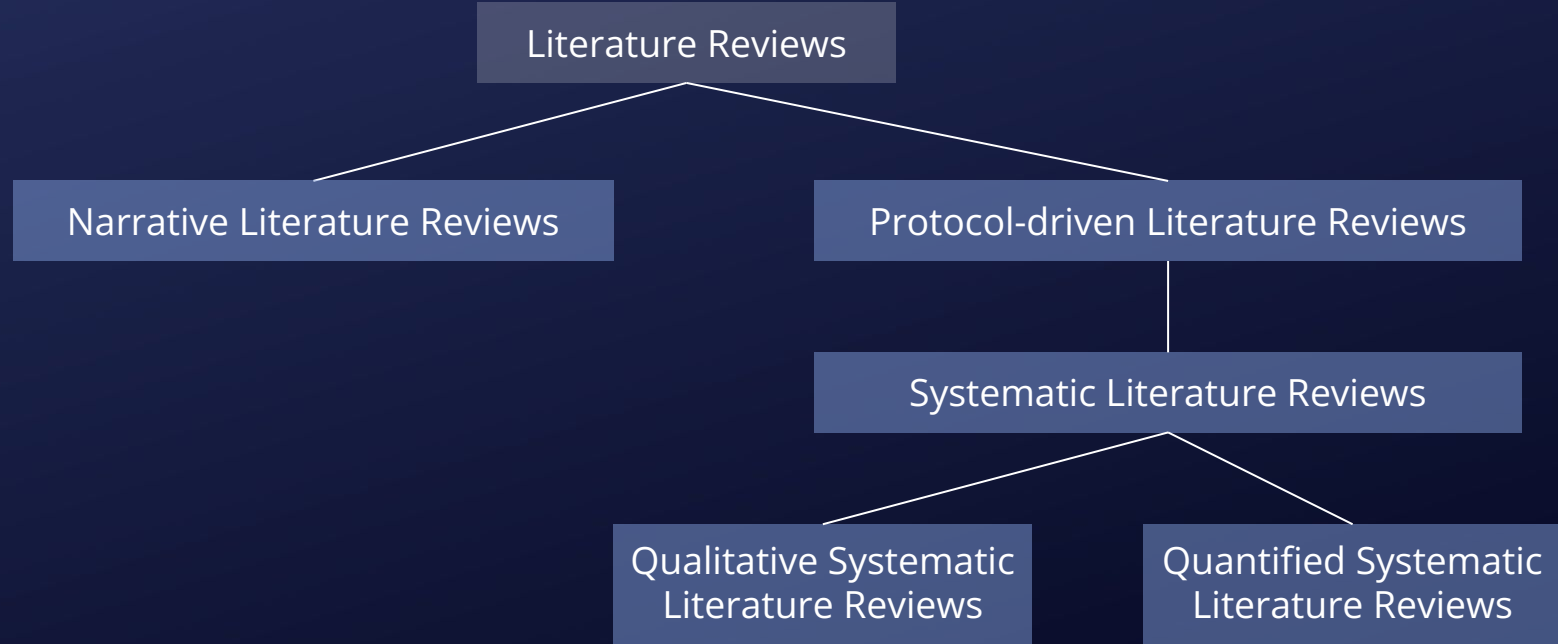


Anna Willoughby



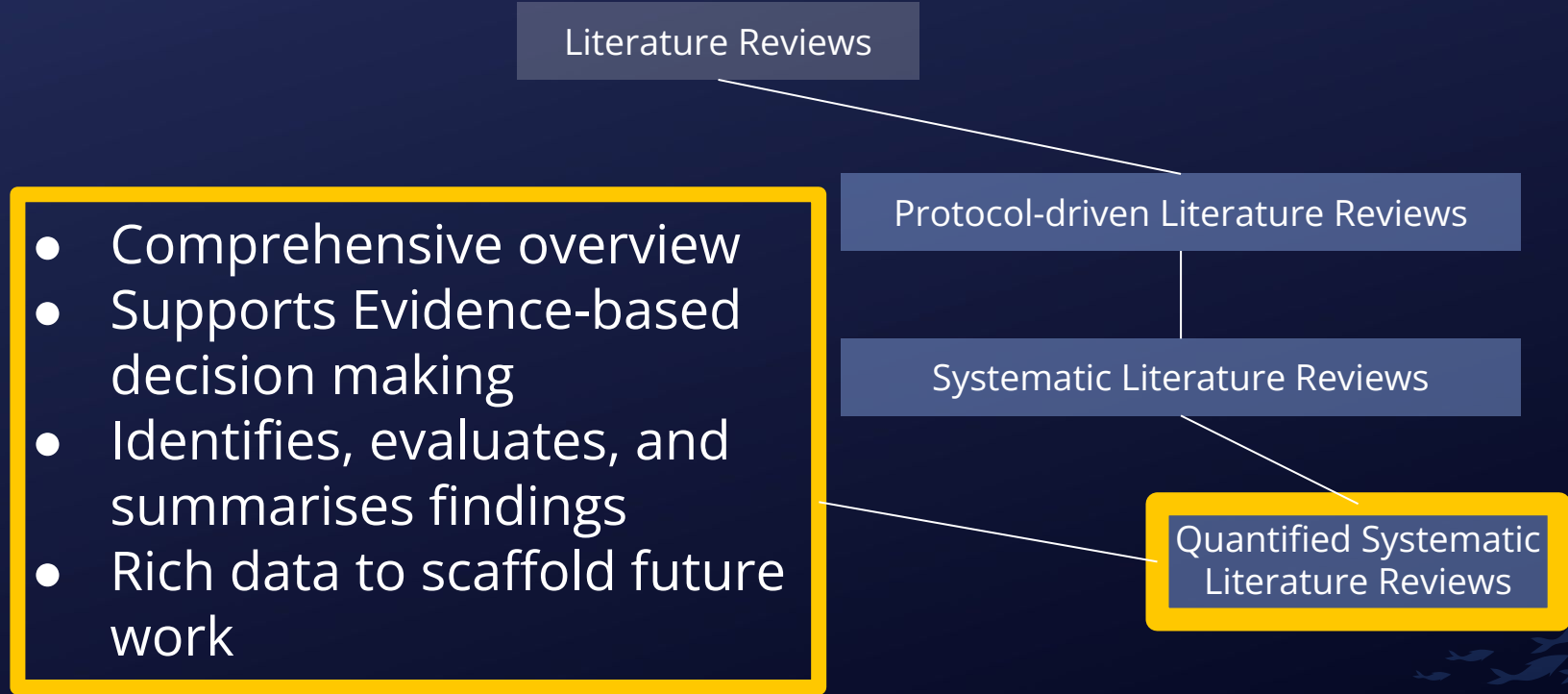
Dr. Kaylee Arnold

Systematic Reviews: a scholarly stepping stone



Adapted from Dekkers et al. 2022

Systematic Reviews: a scholarly stepping stone



Adapted from Dekkers et al. 2022

The PRISMA Flow Diagram

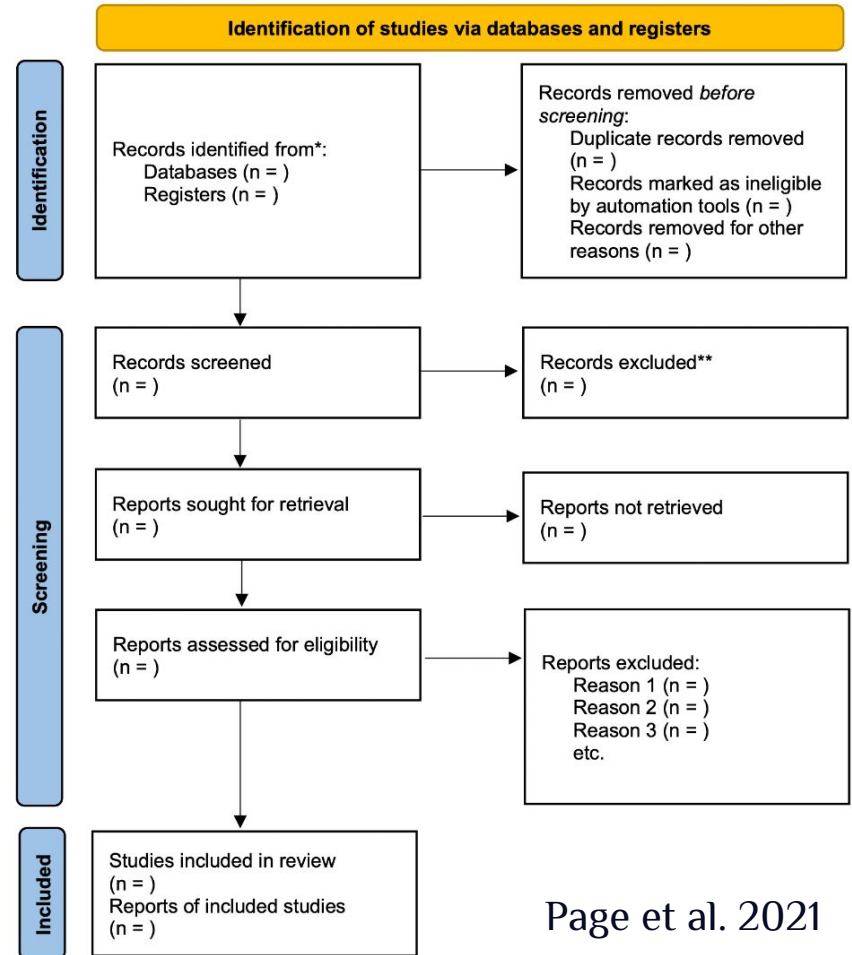
Identification

Search String + Databases

Screening

Access + Criteria

= List of Included Studies



PRISMA facilitates replicability, not rigor

Method Section	Weakness	Possible solutions
Identification	Comprehensiveness Unclear	Create metrics that validate search strategies (e.g., relevancy ratio)
		Test literature accumulation across databases and approaches
Inclusion	Bias: <i>Geographic, Language, Accessibility</i>	Implement hybrid approaches; Retain search strategy metadata with the document outputs so corpora can be assessed separately and compared

Case Study 1: What do ringtails eat?



Photo courtesy of Saguaro National Park

Bassariscus astutus

native to the US and Mexico

resistant to urbanization?

What data do we need?

Rodríguez-Estrella et al. 2000

Journal of Arid Environments (2000) **44**: 241–246

doi:10.1006/jare.1999.0579, available online at <http://www.idealibrary.com> on IDEAL®



Spring diet of the endemic ring-tailed cat (*Bassariscus astutus insulicola*) population on an island in the Gulf of California, Mexico

Ricardo Rodríguez-Estrella, Angel Rodríguez Moreno
& Karina Grajales Tam

Centro de Investigaciones Biológicas del Noroeste (CIBNOR), San Juan de la Costa, El Comitán, La Paz 23000 B.C.S., Mexico

(Received 24 December 1998, accepted 26 July 1999)

The spring diet of an endemic ring-tailed cat (*Bassariscus astutus insulicola*) population of Isla San José, Gulf of California, Mexico, is presented based on the analysis of 104 faeces. The ring-tailed cat is an opportunistic species, concentrating on the predation of mainly small prey, and has a richness of prey of 36 animal taxa and 12 plant species. The most important prey were found to be mainly Orthoptera, Tenebrionida, and Scorpionoidea, *Chaetodipus spinatus* and *Neotoma lenida*. *Chilomeniscus cinctus*, *Cnemidophorus*, and *Sceloporus* were

What data do we need?

Rodríguez-Estrella et al. 2000

Journal of Arid Environments (2000) **44**: 241–246

doi:10.1006/jare.1999.0579, available online at <http://www.idealibrary.com> on IDEAL®



Spring diet of the endemic ring-tailed cat (*Bassariscus astutus insulicola*) population on an island in the Gulf of California, Mexico

Ricardo Rodríguez-Estrella, Angel Rodríguez Moreno
& Karina Grajales Tam

Centro de Investigaciones Biológicas del Noroeste (CIBNOR), San Juan de la Costa, El Comitán, La Paz 23000 B.C.S., Mexico

(Received 24 December 1998, accepted 26 July 1999)

The spring diet of an endemic ring-tailed cat (*Bassariscus astutus insulicola*) population of Isla San José, Gulf of California, Mexico, is presented based on the analysis of 104 faeces. The ring-tailed cat is an opportunistic species, concentrating on the predation of mainly small prey, and has a richness of prey of 36 animal taxa and 12 plant species. The most important prey were found to be mainly Orthoptera, Tenebrionida, and Scorpioneidea, *Chaetodipus spinatus* and *Neotoma lepida*. *Chilomeniscus cinctus*, *Cnemidophorus*, and *Sceloporus* were

Diet Survey!

	Frequency	Occurrence
Mammalia		
<i>Chaetodipus spinatus</i>	33	17
<i>Dipodomys insularis</i>	4	4
<i>Peromyscus eremicus</i>	12	12
<i>Neotoma lepida</i>	17	17
<i>Odocoileus hemionus*</i>	1	1
<i>Caprinus capra*</i>	3	3
Total	70 (15.2)†	61 (58.7)
Aves		
<i>Campylorhynchus brunneicapillus</i>	1	1
Unidentified birds	4	4
Total	5 (1.1)	5 (4.8)
Reptilia		
<i>Phyllodactylus nocticolus</i>	2	2
<i>Dipsosaurus dorsalis</i>	5	5
<i>Callisaurus draconoides</i>	2	2
<i>Sauromalus ater</i>	3	3
<i>Sceloporus</i> spp.	9	9
<i>Urosaurus microscutatus</i>	2	2
<i>Uta stansburiana</i>	3	3
<i>Cnemidophorus</i> spp.	6	6
<i>Chylomeniscus cinctus</i>	12	12
<i>Hypsigena torquata</i>	5	5
Unidentified lizards	8	8
Total	57 (12.3)	52 (50.0)
Invertebrata		
Scorpioneidea	43	38
Solifugae	1	1
Unidentified Arachnida	6	6
Chilopoda	9	6
Scarabaeidae	3	1
Tenebrionidae	87	50
Cerambycidae	3	3



Hybrid Approach

1) Systematic Review

performed by *Willoughby et al. In Prep*

2) Snowballing and Citation Chaining

performed by *Willoughby et al. In Prep*



1) Systematic Review

A. Library Search (through University of Georgia Libraries in May 7, 2022)

Library databases: **BIOSIS, CAB Abstract , Zoological Record, Web Of Science,**
Wildlife & Ecology Studies

Search terms:

("ringtail" **OR** "ring-tailed" OR "ringtailed" **OR** "Bassariscus" OR "cacomistle" OR
"cacomixtle" **OR** "babisuri" OR "bassarisk") **AND**
("diet" **OR** "foraging" **OR** "alimentary" **OR** "dieta" **OR** "food" **OR** "forrajeo" or
"recurso" **OR** "alimento") **NOT** "lemur" **NOT** "possum"




2) Snowballing and Citation Chaining

B. Targeted Academic Sources

i. Citation Chaining from identified primary studies

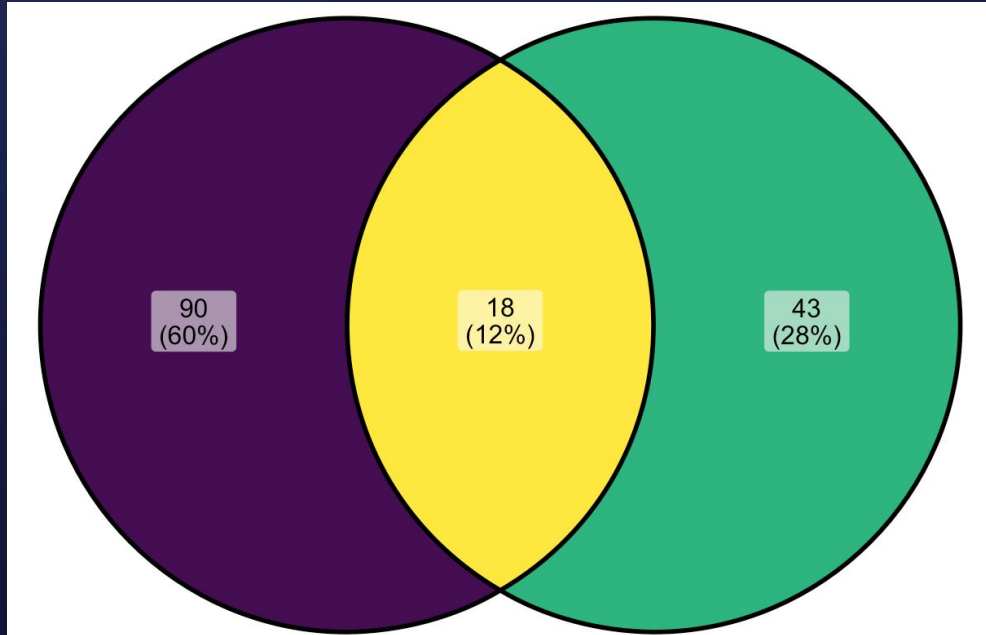
ii. Public University and College Library Queries

- Limited to States within ringtail range
 - (“ringtail” **OR** “ring-tailed” OR “ringtailed” **OR** “Bassariscus” OR “cacomistle” OR “cacomixtle” **OR** “babisuri” OR “bassarisk” OR “mammal” OR “mamifero”)
- 

What did literature did we find?

Systematic Search

Snowballing



151 **papers**

94 **journals**

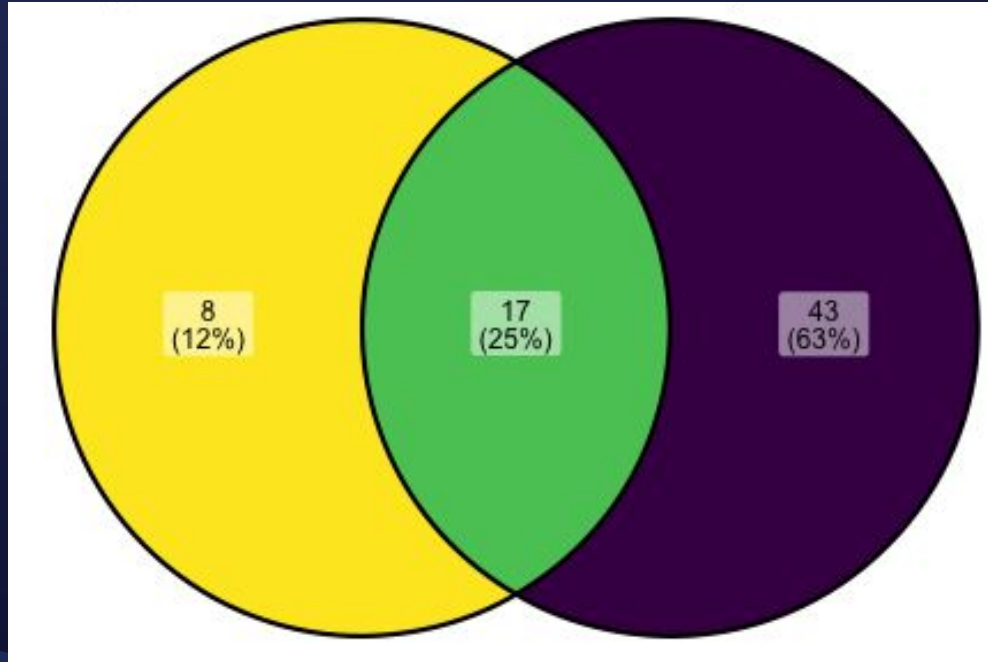
4 **languages**

40% **open access**

What did relevant literature did we find?

Systematic Search

Snowballing



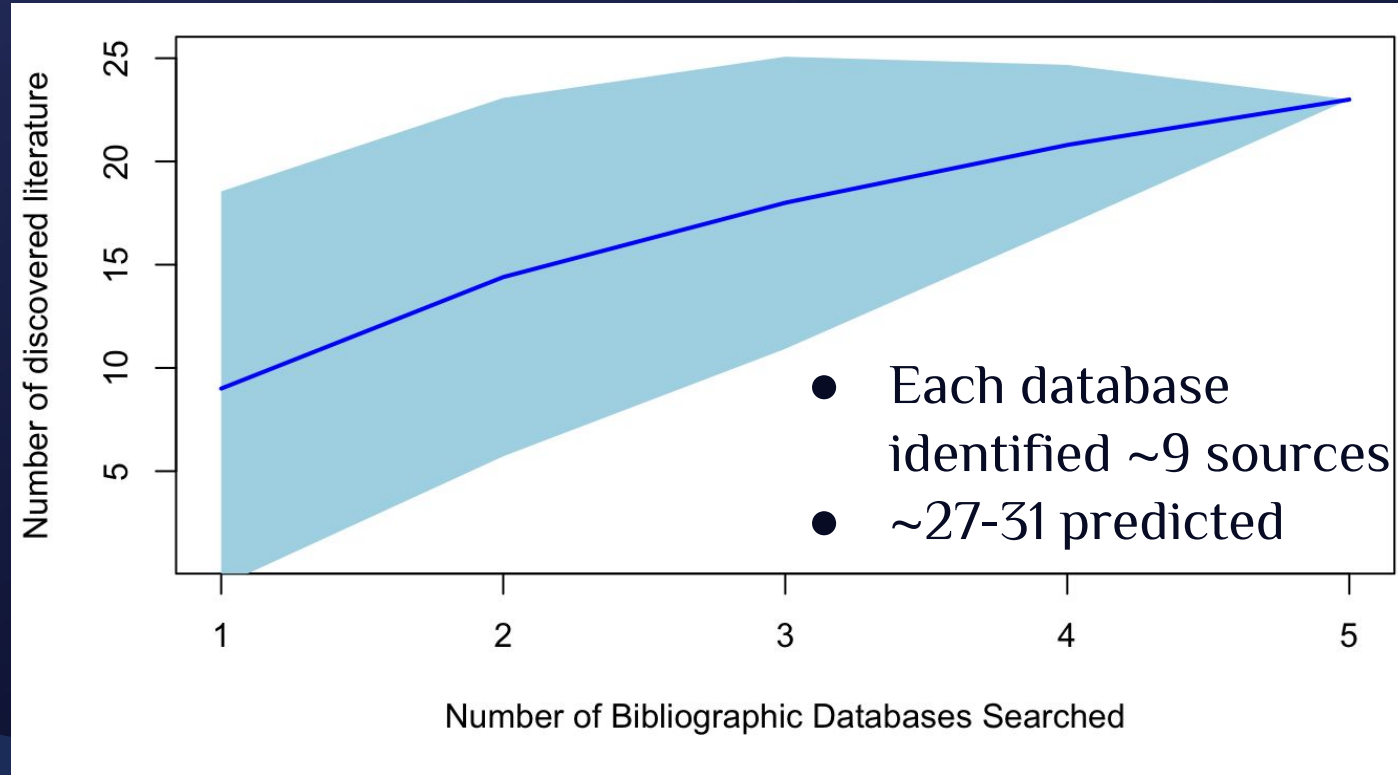
68 *sources*

26 *journals*

2 *languages*

56% *open access*

Was five bibliographic databases enough?



Approaches collate different publishers



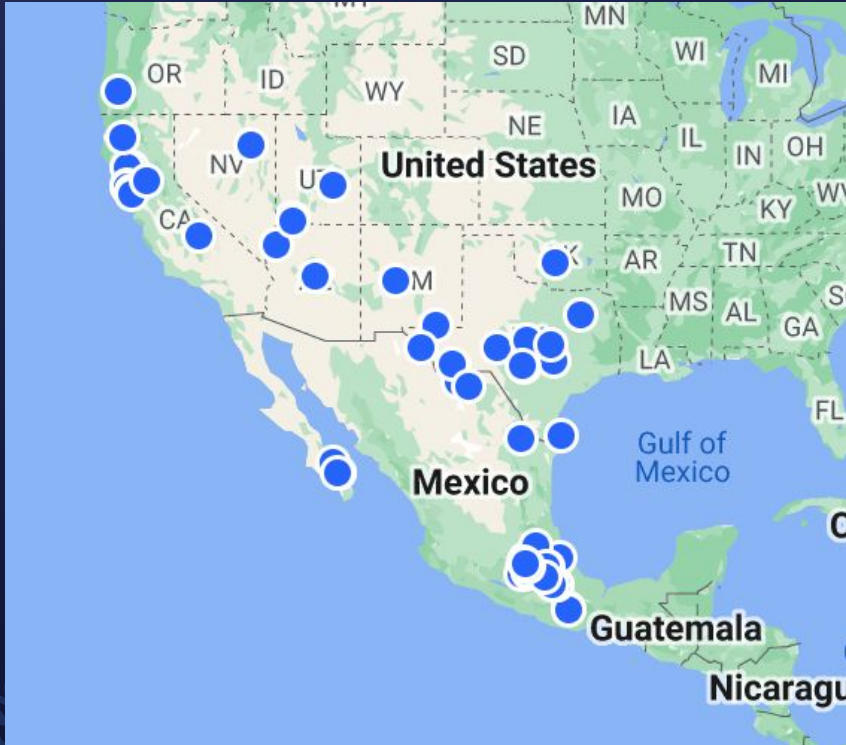
Systematic Search
discovered
literature from
global publishing
houses in 11
countries

Approaches collate different publishers



Snowballing
focused on local
research,
particularly
collegiate
products from
Mexico (27) and
the US (32)

How does this affect the data?



81	Total
53	Snowball Unique
8	Systematic Unique
20	Both

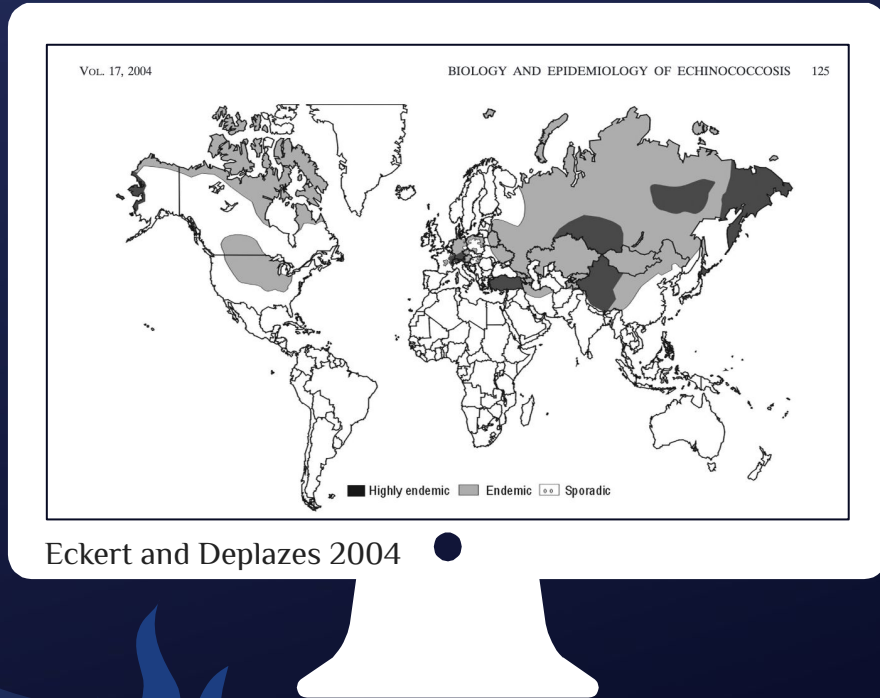
Wrap up thoughts

- For a species limited to two countries, place-based searches are achievable and fruitful
- Additional bibliographic databases are not always needed, weigh the cost:benefits before more work
- Give space for citation chaining and requesting hard-copy literature

Comments or Questions?



Case Study 2: Where is the fox tapeworm?



Echinococcus multilocularis



Helminth parasite that infects wild mammals and people

Native to Europe, but spreading?

What data do we need?

Laurimaa et al. 2015b

LETTERS

Noninvasive Detection of *Echinococcus multilocularis* Tapeworm in Urban Area, Estonia

Leidi Laurimaa, John Davison, Liivi Plumer, Karmen Süld, Ragne Oja, Epp Moks, Marju Keis, Maris Hindrikson, Liina Kinkar, Teivi Laurimäe, Jaana Abner, Jaanus Remm, Peeter Anijalg, and Urmas Saarma

Author affiliation: University of Tartu, Tartu, Estonia

DOI: <http://dx.doi.org/10.3201/eid2101.140136>

To the Editor: Alveolar echinococcosis, which is caused by the fox tapeworm *Echinococcus multilocularis*, is an emerging disease in Europe that shows a high mortality rate (1). Humans can become infected after ingesting parasite eggs (e.g., through direct contact with dogs and red foxes [*Vulpes vulpes*] or with their contaminated feces). *E. multilocularis* tapeworm eggs are extremely resistant and can remain viable in the environment for years (2).

infected with the *E. multilocularis* tapeworm in natural habitats in Estonia (4), it is essential to monitor parasite spillover into urban areas, where it could become a serious public health risk. Consequently, there is an acute need for methods that can effectively detect the parasite and thereby help prevent human infection.

Although immunologic (2) and genetic methods (5–7) are available for identifying *Echinococcus* spp. parasites, a sensitive molecular diagnostic method that detects tapeworms and identifies their host species from degraded fecal samples would be useful. The purposes of this study were to develop a sensitive, noninvasive, genetic method to identify the host species by discriminating between feces of red foxes and dogs; detect *E. multilocularis* tapeworms in feces and distinguish them from the related parasite *E. granulosus*; and collect carnivore feces in an urban area in Estonia to identify this tapeworm.

Fecal samples suspected to be from red foxes were collected during January–March 2012 and January–March 2013 from streets and grassy areas of Tartu, Estonia. Tartu is a relatively small city (area 39 km²) with 98,000 human inhabitants. We surveyed 14 transects, each ≈4 km in length, that included all major districts in the city (Figure).

What data do we need?

GPS points!

Laurimaa et al. 2015b

Noninvasive Detection of *Echinococcus multilocularis* Tapeworm in Urban Area, Estonia

Leidi Laurimaa, John Davison, Liivi Plumer, Karmen Süld, Ragne Oja, Epp Moks, Marju Keis, Maris Hindrikson, Liina Kinkar, Teivi Laurimäe, Jaana Abner, Jaanus Remm, Peeter Anijalg, and Urmas Saarma

Author affiliation: University of Tartu, Tartu, Estonia

DOI: <http://dx.doi.org/10.3201/eid2101.140136>

To the Editor: Alveolar echinococcosis, which is caused by the fox tapeworm *Echinococcus multilocularis*, is an emerging disease in Europe that shows a high mortality rate (1). Humans can become infected after ingesting parasite eggs (e.g., through direct contact with dogs and red foxes [*Vulpes vulpes*] or with their contaminated feces). *E. multilocularis* tapeworm eggs are extremely resistant and can remain viable in the environment for years (2).

infected with habitats in Estonia. spillover into urban areas is a public health risk. We used methods that can help prevent human infection.

Although data on the prevalence of *E. multilocularis* are available from rural areas, no data are available from a sensitive monitoring system in urban areas. We collected worms and identified them using molecular methods. Fecal samples were collected from streets and grassy areas of Tartu, Estonia. We were to develop a sensitive monitoring system to identify the location of red foxes and their feces in urban areas and distinguish them from the related parasite *E. granulosus*; and collect carnivore feces in an urban area in Estonia to identify this tapeworm.

Fecal samples suspected to be from red foxes were collected during January–March 2012 and January–March 2013 from streets and grassy areas of Tartu, Estonia. Tartu is a relatively small city (area 39 km²) with 98,000 human inhabitants. We surveyed 14 transects, each ≈4 km in length, that included all major districts in the city (Figure).

Northeastern Europe

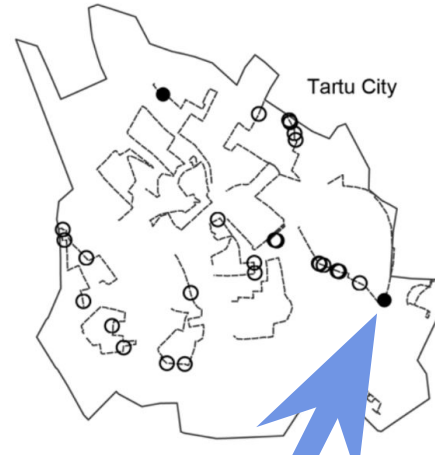


Figure. Location of Tartu in northeastern Europe, Estonia, and red fox feces sampling area in Tartu. The Tartu City boundary is indicated by a solid black line, survey transects are indicated by dashed lines, and fox fecal samples (n = 28) are indicated by circles. Filled circles (n = 2) indicate samples positive for *Echinococcus multilocularis* tapeworms.

Emerging Infectious Diseases • www.cdc.gov • Vol. 21, No. 1, January 2015

163

58°23'11.0"N 26°44'07.9"E



Hybrid Approach

- 1) Systematic Review + Person & Place Snowballing
performed by *Oksanen et al. 2016*
 - 2) Systematic Review + Publisher Snowballing
performed by *Wardeh et al. 2015* and *Stephens et al. 2017*
- 

1) Systematic Review + Person & Place Snowballing performed by *Oksanen et al. 2016*

A. Library Search (two rounds: 2013 & 2015)¹

Library databases: MEDLINE, EMBASE, SciSearch, BIOSIS, CABI, Google Scholar accessed through Documentation Service at Superiore di Sanità, Rome, Italy.

Search terms: collated keywords then performed database searches

[Echinococcus multilocularis OR (Echinococcus AND Multilocularis) OR E# Multilocularis OR Alveolar Echinococcosis OR A# Echinococcosis] AND (Dog OR Dogs OR Cat OR Cats OR Canis OR Felis OR Canid? OR Felid? OR Wolf OR Animal OR Animals OR Fox OR FOxes OR Vulpes OR Ferret OR Ferrets OR Rodent OR Rodents OR Rodentia OR Nutria# OR Muskrat# OR Jackal# OR Arvicolid? OR Arvicolinae OR Worm Burden OR Host OR Hosts OR Hosted) AND (Occurrence# OR Geographic? Distribut? OR Geographic? Diffus? OR Incidence# OR Frequency OR Epidemic Outbreak# OR Endemic Outbreak# OR Prevalence# OR Epidemiology)].



1) Systematic Review + Person & Place Snowballing performed by *Oksanen et al. 2016*

B. Targeted Academic Sources

a. Scientific conferences:

Search terms: “European Union report, “EU report”, “conference proceedings”, “Echinococcus multilocularis”, “E. multilocularis” and “alveolar echinococcosis”.

b. Theses and Dissertations

Search terms: “**Echinococcus multilocularis**” and “**alveolar echinococcosis**”.:




2) Systematic Review + Publisher Snowballing performed by *Stephens et al. 2017*


A. iterative Boolean searches (1999 - 2010/2015)¹


Library databases: Biological Abstracts, Web of Science, AGRICOLA, Medline, PrimateLit

Search terms: combination of authoritative (Wilson and Reeder 2005 & 1993; Corbet and Hill 1999)

primate, ungulate, and carnivore species binomials **AND**

("pathogen", "parasite" or "disease")






2) Systematic Review + Publisher Snowballing performed by *Wardeh et al. 2015*

B. Publisher-based Boolean searches

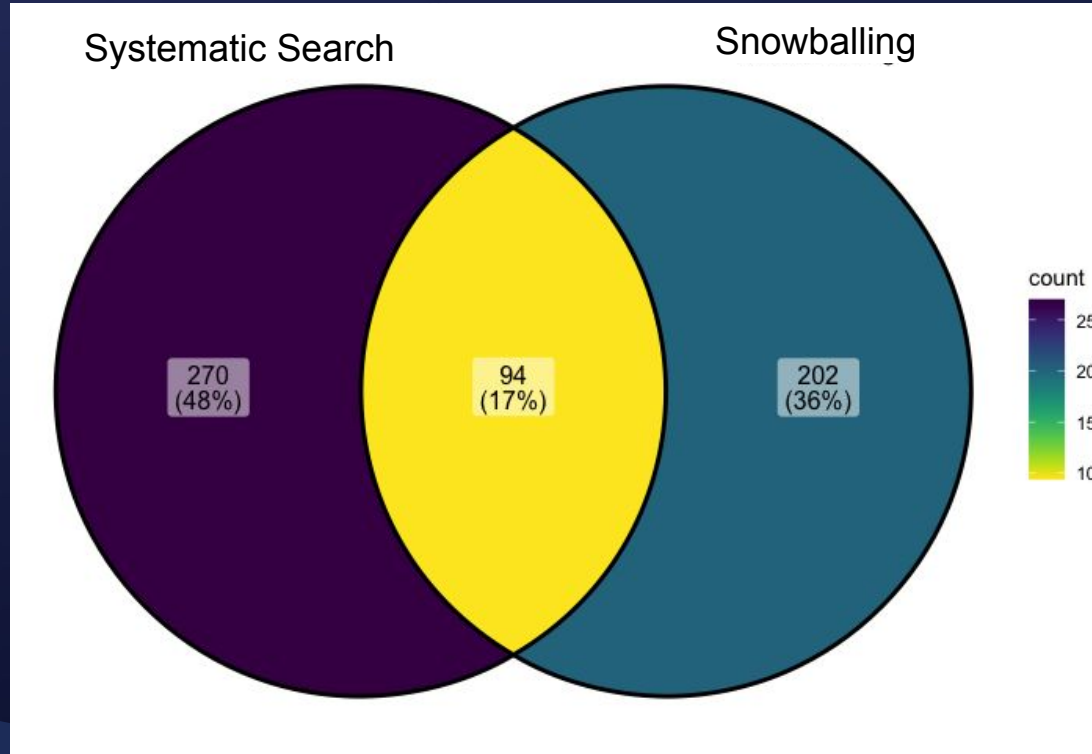
Public and Library databases: NCBI Taxonomy and Nucleotide Sequence, PubMed

Search terms: NCBI-detected species + synonyms

('classical swine fever virus' [Text Word] OR 'csfv' [Text Word] OR 'hog cholera virus' [Text Word] OR 'pestivirus type 2' [Text Word] OR 'swine fever virus'[Text Word]) NOT ' african swine fever' [Text Word]

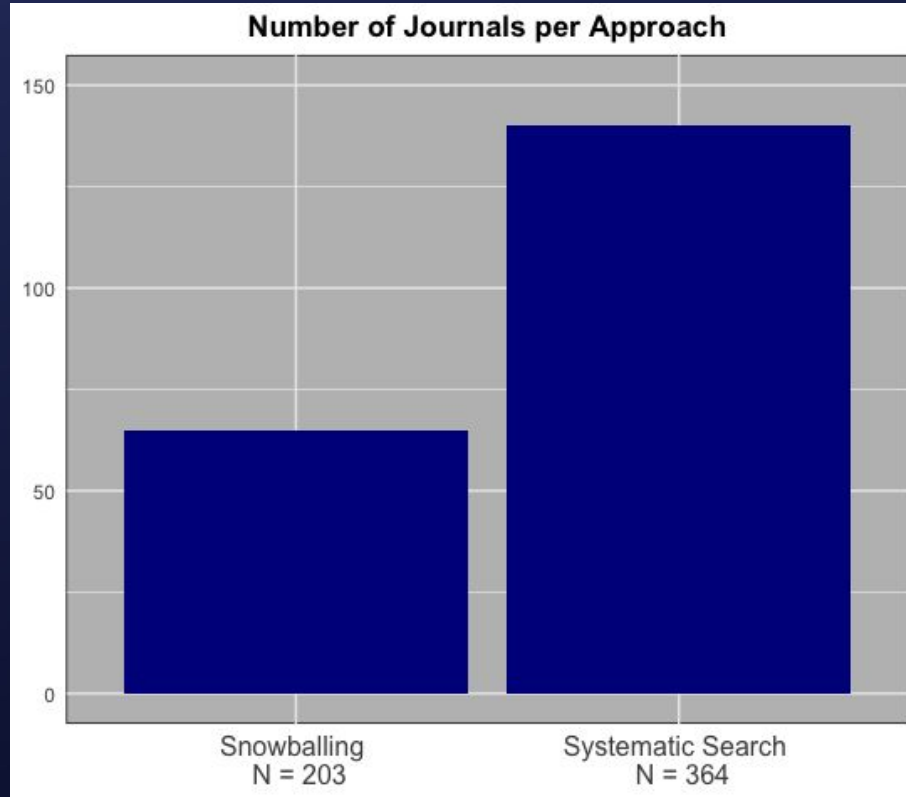


What literature did we find?



567 Papers

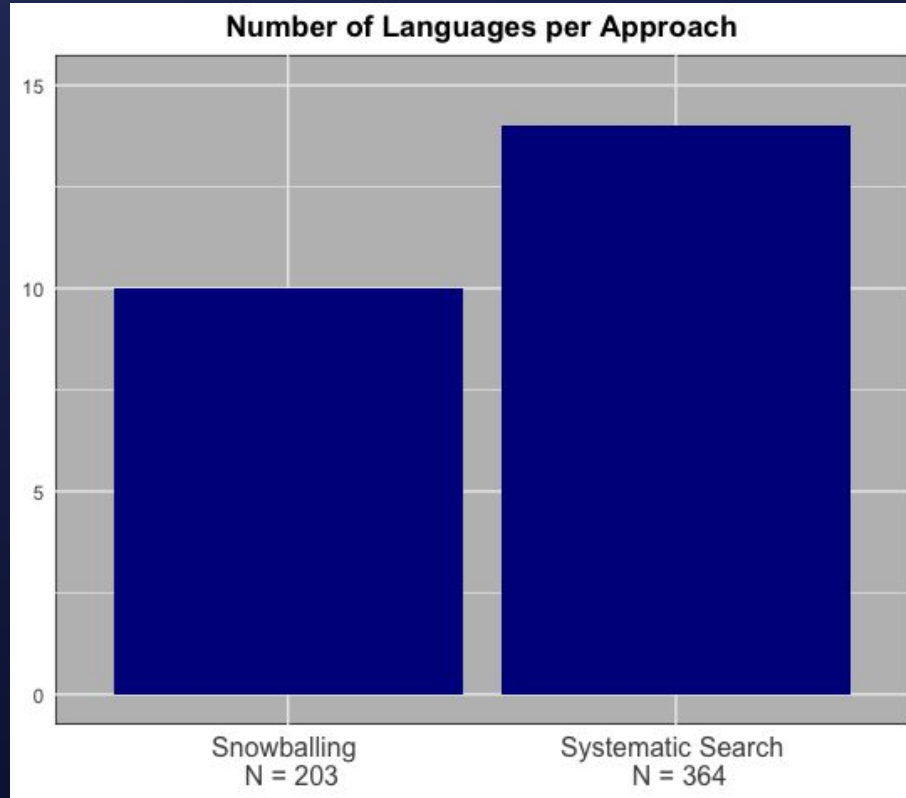
What literature did we find?



567 Papers

170 Journals

What literature did we find?

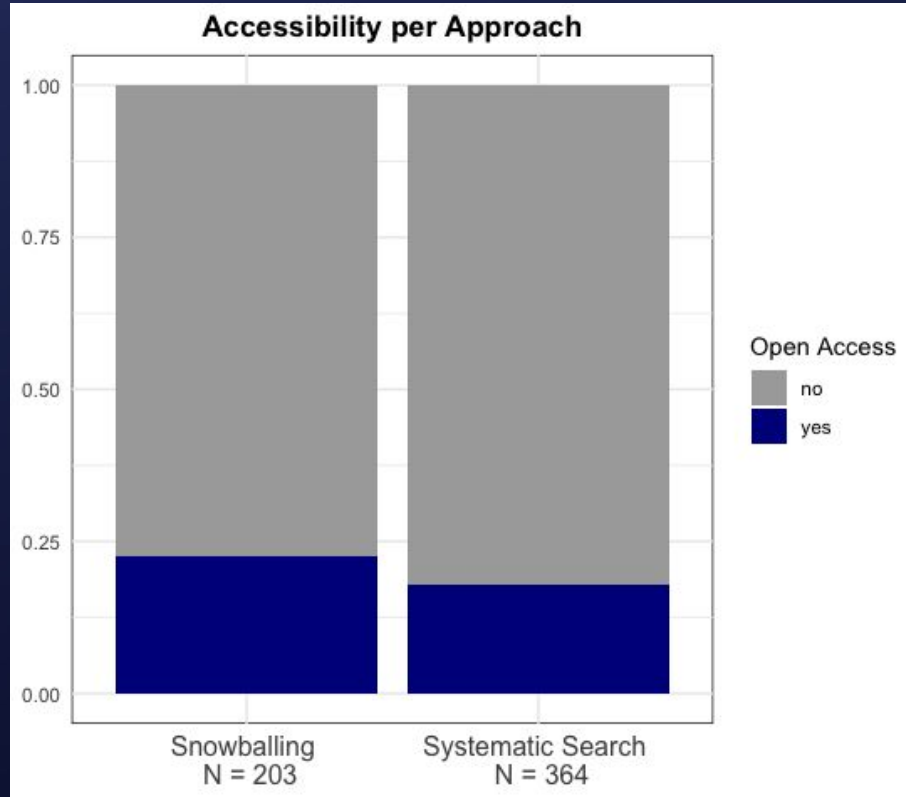


567 Papers

170 Journals

16 languages

What literature did we find?



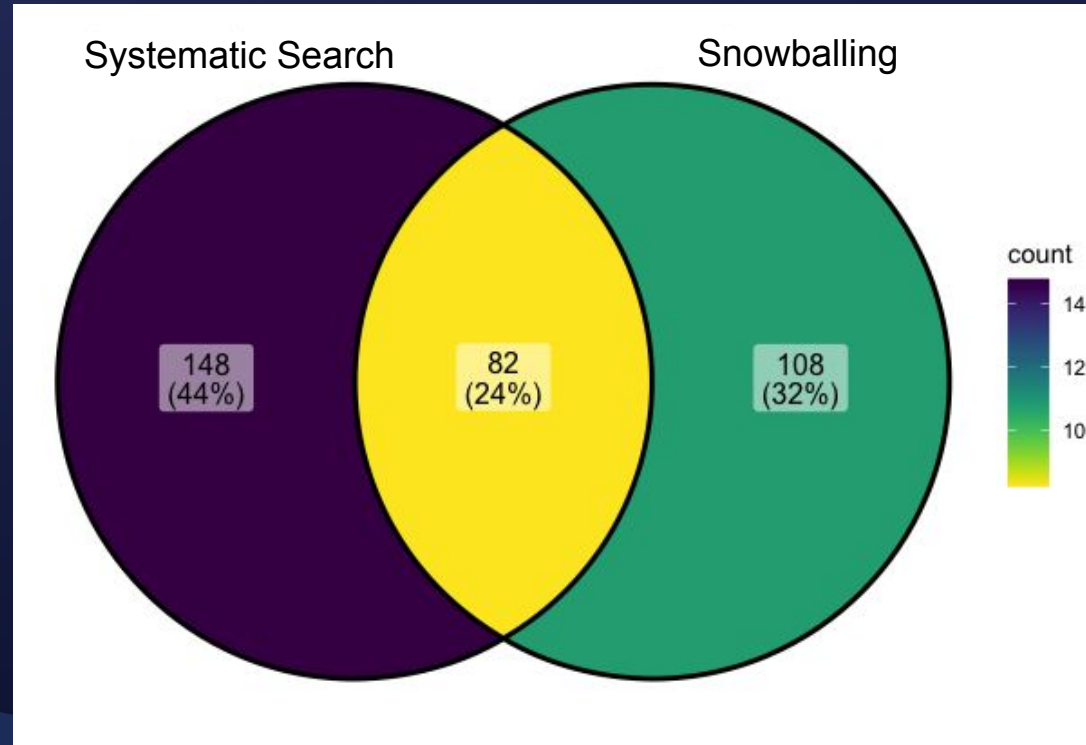
567 Papers

170 Journals

16 languages

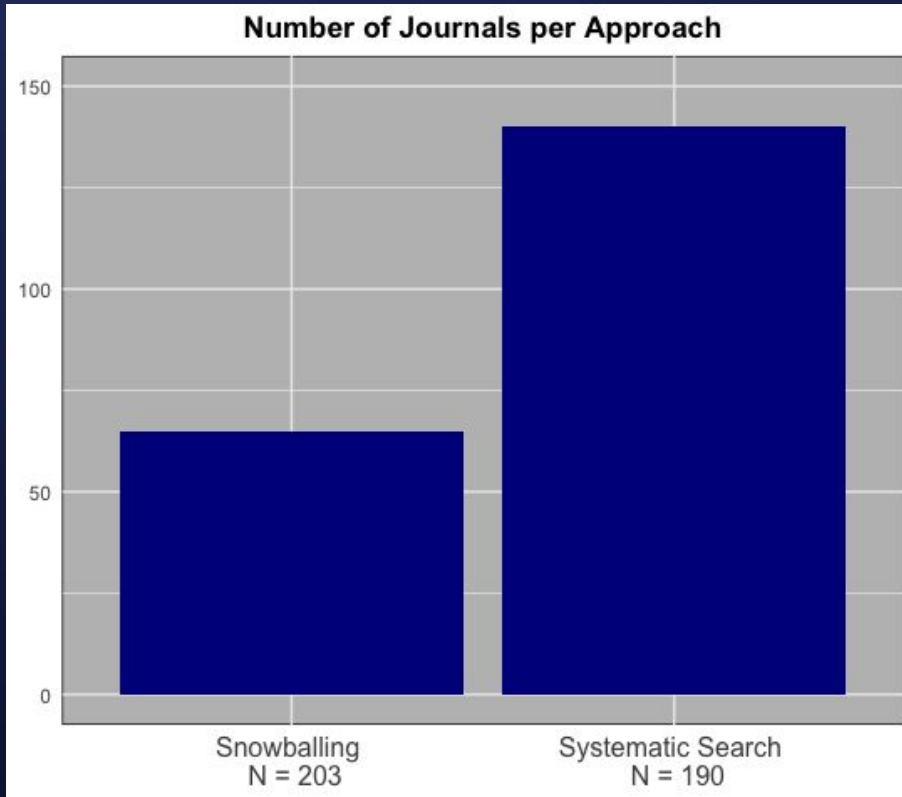
111 Open Access

What did relevant literature did we find?



338 Papers

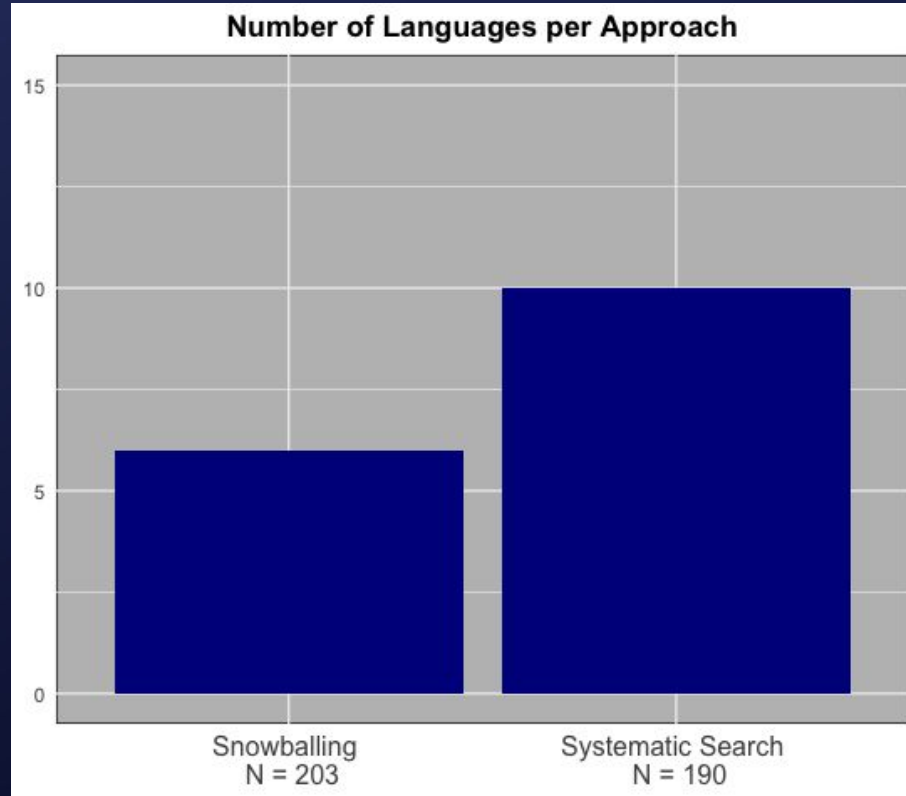
What did relevant literature did we find?



339 Papers

106 Journals

What did relevant literature did we find?

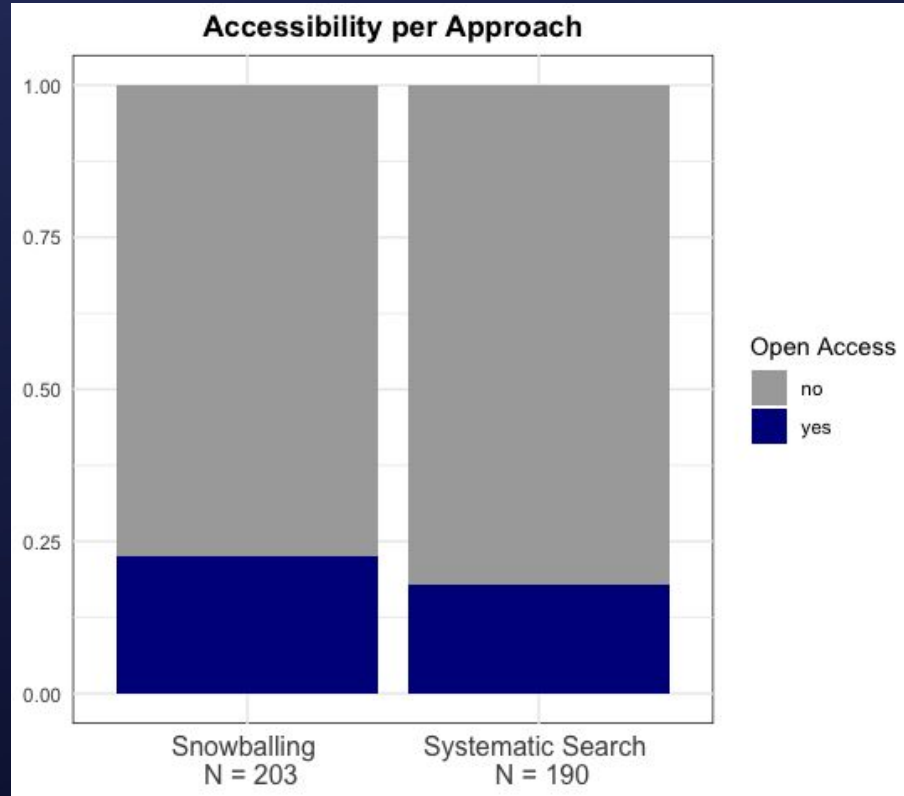


339 Papers

106 Journals

13 languages

What did relevant literature did we find?



339 Papers

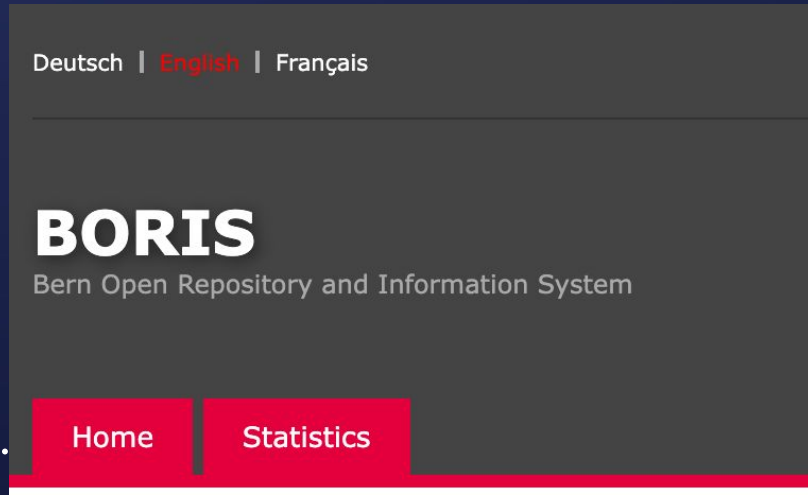
106 Journals

13 languages

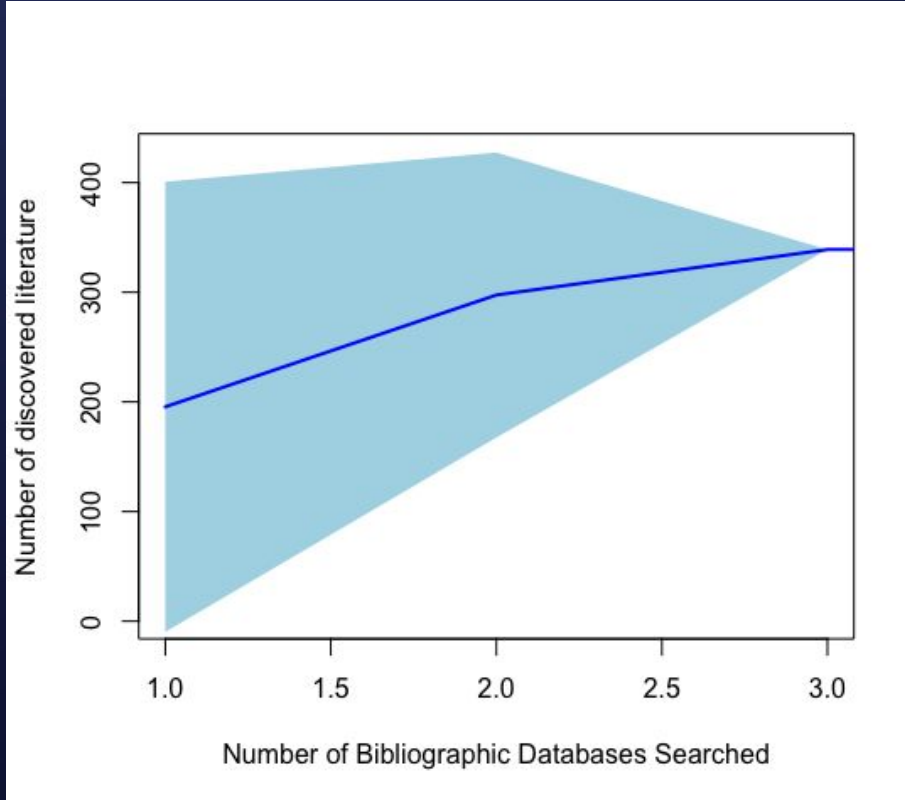
77 Open Access

(23%)

Authors are making their work available!

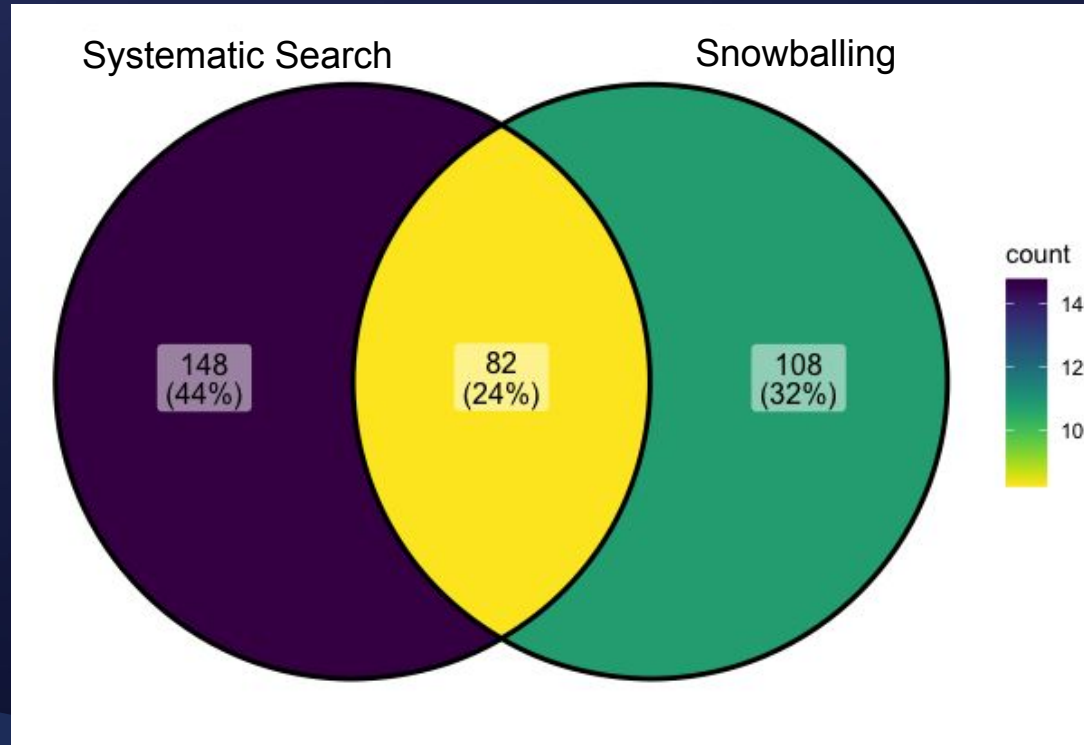


Was three bibliographic sources enough?



Each source
identified
~ 195 papers
(+/- 118)

How might this affect our results?



Wrap up thoughts

- Systematic reviews are often used by scientists to identify literature gaps and inform future research directions
- A combination of systematic searches and snowballing approaches can increase our knowledge of important public health data



Comments or Questions?



No Silver Bullet!

- Systematic Quantitative Review require immense data organization with various file types
- Hybrid approaches expand the geographic, languages, and freely accessible relevant literature
- New metrics are required to inform how effective review methods are

Thank you!

Do you have any questions?

Anna Willoughby
anna.willoughby@uga.edu

**Many thanks to UGA Libraries for
fulfilling an exorbitant amount of
inter-library loan requests**

CREDITS: This presentation template was
created by [Slidesgo](#), and includes icons by
[Flaticon](#), and infographics & images by [Freepik](#)

References

Dekkers, R., Carey, L. and Langhorne, P. (2022). Making Literature Reviews Work: A Multidisciplinary Guide to Systematic Approaches. Springer, London.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. International journal of surgery, 88, 105906.

Oksanen

CDC