Did unsupported citations about forest fungal networks double? It depends on the weighting

A preregistered reread of the claim in Karst, Jones and Hoeksema (2023) on the authors' own deposited table, with planted calibration and year-shuffled controls

Zain Dana Harper

Version 1.0, 9 November 2026. Finding F149 in the findings ledger. Strong signal (tier B*).

Abstract

Karst, Jones and Hoeksema reported in 2023 that citations crediting forest fungal network studies with more than their results support had doubled in the past 25 years. That sentence is now part of a public dispute. We reran the claim on the table the authors deposited, under a plan hash-locked before the table was analysed. The table holds 268 study-year rates for 11 focal studies on network function and 7 on network structure, and no denominators, which we rebuilt from the paper's supplement. On the authors' negative binomial model, unsupported citations rose by a factor of 2.19 for network function (0.86 to 4.14) and 2.84 for network structure (1.18 to 6.62). Weighted by citations, the rise was 1.77 for function (0.52 to 2.99) and 1.54 for structure (0.85 to 5.81), and neither interval excludes 1. Inside individual focal studies the rate rose for both: 3.0 times for function (p 0.001) and 3.5 times for structure (p 0.009). With this many focal studies, a true doubling gives an interval clear of 1 only 0.425 of the time, so an interval that includes 1 is no evidence against doubling. "Doubled" fairly describes the authors' fitted curve. Counted as a share of all citations the rise is smaller and less certain.

1 What this means for a reader

If you repeat that misleading citations about forest fungal networks doubled, say "rose, and about doubled on the authors' model", with the interval beside it. The rise is real inside the studies that were cited. How large it is depends on whether a year with one citation counts as much as a year with thirty, and with 18 focal studies the uncertainty stays wide either way. Nothing here bears on whether the networks exist or move resources between trees.

2 Introduction and motivation

A common mycorrhizal network is a set of fungal threads that colonise the roots of two or more plants at once. Popular accounts have described such networks as channels through which trees share carbon and warnings. Karst, Jones and Hoeksema [1] reviewed the field evidence and the way it is cited. They read the later citations of 18 field studies, which we call focal studies, and coded each citation as supported or not by what the focal study found. In their abstract they report that unsupported citations "have doubled in the past 25 years". An Author Correction followed [2].

The paper drew replies. Klein and colleagues [4] accepted that some of its criticism is justified and argued that evidence for carbon transfer between trees through these networks is solid and growing. Simard, Ryan and Perry [5] replied to three of its critiques; among other points they called its rationale for counting citations by year vague. Neither reply, as far as we read them at abstract and summary level, reanalysed the deposited numbers.

The doubling figure is a statistic, and the authors deposited the table behind it under CC0. That makes it

checkable at low cost. Our question was narrow: on that table, does the rate of unsupported citations double over the study period for both kinds of claim, and does the answer hold when each year is weighted by how many citations it holds? The decision the reread informs is how our own work may quote the figure. Searches run on 9 November 2026 for a numeric reanalysis of the deposit returned none. That is what the searches returned; it does not show that none exists.

3 Preregistration and design

The preregistration was hash-locked at 06:07:12 UTC on 31 October 2026, before the table was analysed. It was not blind, and it lists what had been seen: the deposit's metadata in full, the header row of the table, and one data row printed by accident while checking the header. From an earlier reading of ours we also knew the per-year numerators in the paper's supplementary Note 3 and the method in its Note 2: a negative binomial regression of the percentage on indexed year. The article body and its main figure were behind a paywall and unread, so the authors' fitted curve and their own definition of doubled were not available to us.

The plan fixed a definition. Doubled means: the ratio of the fitted rate at the end of each group's span to the fitted rate at its start has a 95 percent interval, from resampling focal studies, that excludes 1 and covers 2, and the point estimate beats year-shuffled rates. Function and structure are read separately.

The plan named five measures, set out in the methods, with a verdict for each and a calibration on planted data. Kill conditions, without discretion:

- If the calibration fails, no doubling verdict stands.
- If the authors' model fails in both groups, the fitted curve does not support doubling on this table.
- If the authors' model passes and the within-study model fails, the rise comes from a change in which studies are cited.
- If the citation-weighted model fails beside a pass of the authors' model, the rise depends on thin early years.

The plan also recorded our own forecasts, set against the outcomes in Table 2.

4 Data and provenance

The data are the authors' Dataverse deposit [3], released under CC0. Its table gives, for each focal study and each citing year, the study-year rate of unsupported citations. A "Count" column holds 100 times the rate, a percentage. The file holds 272 rows; 4 had a blank rate and were skipped and counted, leaving 268 rates: 158 for 11 function studies over 1997 to 2022 and 110 for 7 structure studies over 1999 to 2022. The year 2022 is incomplete.

The table holds no denominators. We built each rebuilt denominator from the per-year numerators printed in the paper's supplementary Note 3: for a nonzero rate and a known numerator k, the whole number n with rate k/n. Where no whole number fits, the reduced fraction's denominator was used as a lower bound. Years with a rate of zero carry no numerator information and shared what remained of the study's reviewed total, the weakest step in the rebuild. The supplement is copyrighted, so this paper reports only derived values from it.

5 Methods

Integrity and rebuild

Two checks came first. The integrity check required the Count column to equal 100 times the rate, the expected numbers of studies, and group means of the yearly rate within 1.5 points of the means stated by the authors, 32.2 and 14.8 percent. The rebuild check required at least 0.7 of nonzero rates to give a whole-number denominator and a median gap of at most 0.2 between rebuilt numerators and the totals in the paper's tables. A failed rebuild would void only the citation-weighted measure.

The authors' model as deposited

Following Note 2, a negative binomial model with a log link was fitted to the percentage in each study-year on the indexed year t, using every row including the incomplete 2022:

$$\ln \mu_t = a + \beta t \tag{1}$$

The ratio over a group's span T (25 years for function, 23 for structure) is

$$R = e^{\beta T} \tag{2}$$

The same fit without 2022 is reported beside it. This model gives every study-year equal weight, a year with one citation as much as a year with thirty.

Within study

New focal studies enter the pool over time, so a pooled rise could come from which studies are cited when. The within-study model gives each focal study s its own level:

$$\ln \mu_{st} = a_s + \beta t \tag{3}$$

fitted as a quasi-Poisson fit. It passes if the slope is above 0 and above the 95th percentile of its year-shuffled control.

Citation-weighted

The citation-weighted share model fits the rebuilt count k of unsupported citations out of n in each study-year by a binomial logistic regression on year:

$$\ln \frac{p_t}{1 - p_t} = a + \beta t \tag{4}$$

and reports the ratio of the fitted share at the last year to the share at the first:

$$R_w = \frac{p_{end}}{p_{start}} \tag{5}$$

Intervals and reported quantities

Intervals for the ratios come from study resampling, because citations of one focal study move together. Reported without a verdict: the pooled share in each group's first ten years against 2012 to 2021, the rate in the first five indexed years against the last five up to 2021, and growth in total citations over the last five years against the five before.

6 Controls and planted checks

The year-shuffled control permutes the rates across years inside each focal study, 1,000 times, and refits the authors' model and the within-study model. A real ratio or slope must beat the control's 95th percentile. Its p value is a p value from draws, so with 1,000 draws the smallest possible value is 1 in 1,001.

The planted check kept the real studies, years and rebuilt denominators and drew new counts from a known truth: a study effect (lognormal, standard deviation 0.4) and a true ratio of 1 or 2 over the span. It ran 60 simulations per group per truth, 120 per truth in all, with 150 study resamples each. The calibration passed if a flat truth was called a rise at most 0.15 of the time, the doubled truth gave a median ratio between 1.6 and 2.4, and its interval covered the truth at least 0.8 of the time.

The authors' stated means were recovered. Row means came out at 33.2 and 15.3 percent against the stated 32.2 and 14.8, within the 1.5 point allowance. The mean of study means gave 32.3 and 14.7, which identifies the definition the authors used.

The rebuild passed: 143 of 169 nonzero rates, 0.846, gave a whole-number denominator, and the rebuilt numerators matched the paper's table totals with a median gap of 0. The per-study totals are in Appendix B.

The calibration passed. With a flat truth the pipeline called a rise 0.092 of the time. With a doubled truth the median ratio was 2.11 and the interval covered the truth 0.89 of the time. The calibration also measured statistical power: with a true doubling, the interval excluded 1 only 0.425 of the time.

7 Results

Table 1 gives the ratios from all three models and Figure 1 draws them.

Table 1. Ratio of the fitted rate at the end of each group's span to the rate at its start. Intervals from study resampling; the shuffle cut is the 95th percentile of 1,000 year-shuffled draws.

Model and group	Ratio	Interval	Shuffle cut	p
Authors' model, function	2.19	0.86 to 4.14	1.31	0.002
Authors' model, structure	2.84	1.18 to 6.62	2.09	0.013
Authors' model without 2022, function	2.26	not reported		
Authors' model without 2022, structure	2.43	not reported		
Within study, function	3.02		slope 0.022	0.001
Within study, structure	3.46		slope 0.036	0.009
Citation-weighted, function	1.77	0.52 to 2.99		
Citation-weighted, structure	1.54	0.85 to 5.81		

On the authors' negative binomial model, unsupported citations rose by a factor of 2.19 for network function (0.86 to 4.14) and 2.84 for network structure (1.18 to 6.62). Both beat their year-shuffled controls. By the letter of the plan, function fails because its interval reaches below 1, and structure passes. Dropping the incomplete 2022 gave 2.26 and 2.43.

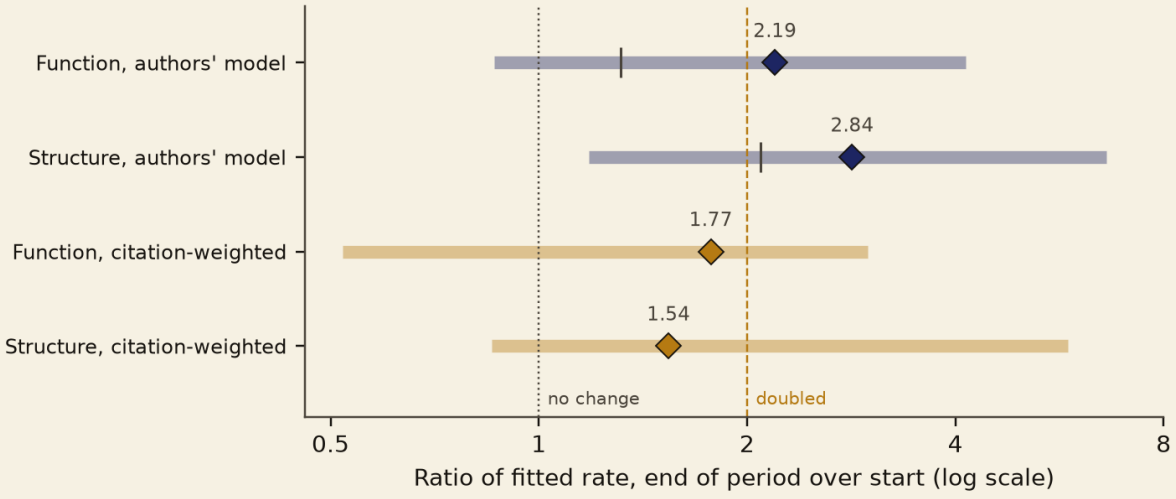


Figure 1. Ratios of fitted end rate to start rate with their study-resampling intervals. Blue: the authors' count model, with ticks at the year-shuffled 95th percentile. Ochre: the citation-weighted share model. Drawn from the run's result file.

Inside individual focal studies the rate rose for both: 3.0 times for function (p 0.001) and 3.5 times for structure (p 0.009). The slopes, 0.044 and 0.054 per year, beat their shuffled cuts of 0.022 and 0.036. The rise does not come from which focal studies happen to be cited early and late.

Weighted by citations, the rise was 1.77 for function (0.52 to 2.99) and 1.54 for structure (0.85 to 5.81), and neither interval excludes 1. The fitted share went from 0.205 to 0.363 for function and from 0.113 to 0.175 for structure. In pooled shares, function went from 21.5 percent in its first ten years to 34.9 percent in 2012 to 2021, about 1.6 times; structure went from 13.6 to 14.7 percent, about 1.1 times. Figure 2 shows why the weighting matters: the early structure years hold few citations each.

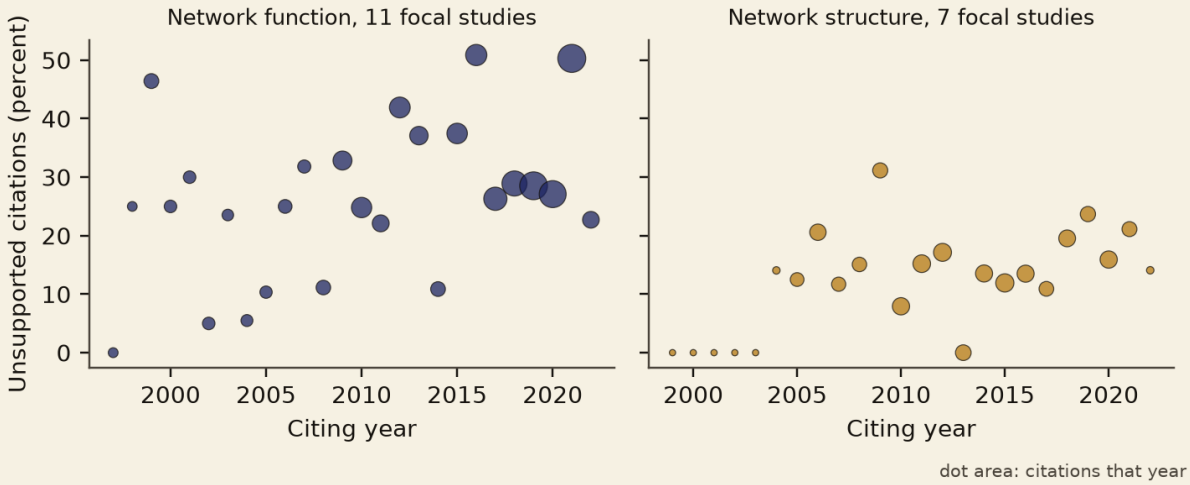


Figure 2. Pooled share of unsupported citations by citing year, from the rebuilt yearly totals, with dot area proportional to that year's citations. The first five structure years each hold about 3 citations and none unsupported.

The preregistered kill text for structure applies. Its pass on the authors' model leans on thin early years: the first five structure years, 1999 to 2003, each hold about 3.1 rebuilt citations and a rate of zero. On five-year windows, function's rate went from 0.253 to 0.377, a ratio of 1.49; structure's first five years are all zero, so its ratio has no finite value.

Total citations also grew. Over the last five years against the five before, function's citations grew 1.86 times and structure's fell to 0.85 times.

With this many focal studies, a true doubling gives an interval clear of 1 only 0.425 of the time, so an interval that includes 1 is no evidence against doubling. The failures of the citation-weighted verdicts show that the data cannot confirm a doubling on that weighting. They do not show that the rise was smaller than double.

Table 2. Forecasts fixed in the preregistration against what the run gave.

Quantity	Forecast	Outcome
Recovered means within 0.5 points	yes	0.96 and 0.50 on row means
Whole-number share of nonzero rates	0.75	0.846
Authors' model, function	2.2, passes	2.19, fails on its interval
Authors' model, structure	2.0, fails	2.84, passes
Without 2022	lower by 0.2	2.26 and 2.43
Within study, function	p about 0.02, passes	p 0.001, passes
Within study, structure	p about 0.10, fails	p 0.009, passes
Citation-weighted	1.6 and 1.5	1.77 and 1.54
Growth in citations	1.5	1.86 and 0.85
Flat truth called a rise	0.08	0.092

8 Discussion

The reread supports the direction of the original claim and narrows its size. On the authors' own model,

"doubled" is a fair description of the fitted curve, and the within-study fits show the rise inside the cited studies, the strongest part of their case. The size depends on a choice of unit. The deposited Count column is a percentage fitted as a count, so each study-year carries equal weight. A year with one citation coded unsupported counts as much as a busy year. Weighted by citations, function rose by about 1.6 to 1.8 times and structure by 1.1 to 1.5 times.

The power figure governs how to read the failures. With 11 and 7 focal studies, even a true doubling usually fails to give an interval clear of 1. The reread therefore does not refute a doubling. It shows that the deposit cannot pin the size down closely on either weighting.

For citing practice we take one rule from this: when a rate is modelled as a count, report the citation-weighted ratio beside it. The sentence we use for the claim is "rose, about doubled on the authors' model", with the interval.

9 Limitations

- The analysis was not blind. Before the lock we had seen the metadata, the header, one data row and the numerators from an earlier reading; the preregistration lists them.
- Denominators were rebuilt, not read. Years with a rate of zero share a remainder and are the least certain.
- Every rate is the original authors' coding. Their Note 2 calls the coding subjective and names an element of presentism. We did not recode any citation.
- Our definition of doubled may differ from the authors'. Their main figure and article body were not read.
- The calibration measures this pipeline on planted data shaped like the real table. It does not test the coding.
- Nothing here bears on whether common mycorrhizal networks exist or move resources between trees.
- The literature on the dispute was read at abstract and summary level.

10 Verified

Verified: on 9 November 2026 the 268 rates and 4 skipped rows, the group row counts, study counts and spans, the rebuild counts (143 whole-number, 26 fraction-only, 99 zero-rate years), the 0.846 whole-number share and zero median gap, the per-study totals, the recovered means, the calibration rates including the 0.425 power figure, every ratio with its interval, shuffled percentile and p value, the ratios without 2022, the within-study slopes and ratios, the fitted shares, the pooled and five-year windows, the citation growth figures and all ten verdicts were re-read from the run's result file; they match the findings ledger. The preregistration and result files match their recorded SHA-256 hashes. Both figures were drawn by a script that reads only the result file. The study was not re-run for this paper. The cited papers were found by searches run on 9 November 2026 and read at the level of abstracts, search summaries and publisher pages; the full texts were not read.

11 Corrections

- 31 October 2026, before any result: the first run stopped on rows with an empty rate cell. Such rows are now skipped and counted. The preregistration and plan were unchanged; the code hash changed.
- 9 November 2026: first published version.

12 Reproduction

- Data: the CC0 rate table from the Dataverse deposit for Karst, Jones and Hoeksema (2023), and the per-year numerators in the paper's supplementary Note 3.
- Rebuild each study-year's citation count as the whole number n with rate k/n; spread the remainder of each study's reviewed total over its zero-rate years; check that rebuilt numerators match the table totals.

- Fit eq. 1 to the percentages and report eq. 2; fit eq. 3 by quasi-Poisson; fit eq. 4 to the rebuilt counts and report eq. 5. Resample focal studies for intervals.
- Shuffle rates across years inside each study 1,000 times for the controls. Calibrate with planted flat and doubled truths of the same shape.
- Preregistration SHA-256: 55a3742673c0367dccf6c7ef22ead2435aae431199a620e13753f5df5924294e. Plan SHA-256: 7ffc684f107f7ca0963792dcabe7f1d9005158cae6ad78cfef9636fbf1fe6309. Run 20261031T060723Z. Result file SHA-256: 85c71a54d2d135758e9bba36dcee72af5c1955b93d8c1c47 51194942edb34902.
- Cost: the analysis took 237 seconds on one desktop computer.

13 Appendix A. The preregistered verdicts

Each verdict is a set of rules written before the run. A verdict passes only when every rule holds. Values are read from the run's result file.

Table 3. Preregistered verdicts and the rules that decided each one.

Verdict	Deciding rule	Value	Holds
Integrity	Count equals 100 times rate in at least 0.99 of rows; at least 11 and 7 studies	1.0; 11; 7	yes
Means	gap to stated means at most 1.5 points, function and structure	0.96; 0.50	yes
Rebuild	whole-number share at least 0.7; median gap at most 0.2	0.846; 0	yes
Calibration	flat truth called a rise at most 0.15; doubled median 1.6 to 2.4; coverage at least 0.8	0.092; 2.11; 0.89	yes
Doubled, function	interval low end above 1 and at most 2; high end at least 2; ratio above shuffle cut (1.31)	0.86; 4.14; 2.19	no
Doubled, structure	the same rules; shuffle cut 2.09	1.18; 6.62; 2.84	yes
Within study, function	slope above 0 and above shuffle cut (0.022)	0.044	yes
Within study, structure	slope above 0 and above shuffle cut (0.036)	0.054	yes
Weighted, function	interval low end above 1 and at most 2; high end at least 2	0.52; 2.99	no
Weighted, structure	the same rules	0.85; 5.81	no

14 Appendix B. Rebuilt totals by focal study

Totals over all years for each focal study: the citations the authors reviewed; the rebuilt citations in years with a nonzero rate, where a numerator sets the count (the remainder was spread over the zero-rate years); the sum of rebuilt unsupported citations; and the unsupported total in the paper's table. Derived values only.

15 Glossary

Each term is defined where it is first used in the text. The same definitions, with the field each comes from and what each excludes, are collected here.

- **common mycorrhizal network**. A set of fungal threads that colonise the roots of two or more plants at once and so may link them below ground. Field: forest ecology and mycology. Excludes claims about what such a network does, such as moving carbon or signals between trees, which are the matters in dispute.
- **focal study**. One of the 18 field studies on these networks whose later citations the original authors read and coded, 11 on network function and 7 on network structure. Field: our usage, following Karst and colleagues. Excludes the citing papers, which are the units counted inside each focal study.
- **unsupported citation**. A citation that credits a focal study with a claim that the original authors judged the

Table 4. Rebuilt totals by focal study, read from the run's result file.

Focal study	Reviewed	In nonzero years	Rebuilt unsupported	Table unsupported
Beiler 2010	83	48	14	14
Booth 2004	73	50	16	16
Booth 2010	52	39	17	19
Corrales 2016	75	28	5	5
Diedhiou 2010	41	8	3	3
Grelet 2010	56	29	5	5
Horton 1999	142	114	24	24
Horton 2005	49	23	11	10
Kennedy 2003	120	106	26	25
Klein 2016	91	90	27	27
Lerat 2002	49	45	25	25
Lian 2006	102	18	5	6
McGuire 2007	118	96	25	25
Simard 1997	382	345	118	116
Teste 2008	58	53	26	26
Teste 2009	80	68	27	28
Teste 2010	40	33	20	21
Warren 2008	65	49	24	25

study's results do not support. Field: citation analysis, as coded by Karst and colleagues. Excludes any judgement of ours: every coding in this paper is the original authors'.

- **preregistration**. A dated written plan for a study, fixed and hash-locked before any result is seen: the question, the data, the statistic, the controls, the forecasts and the kill conditions. Field: research method (metascience). Excludes a plan revised after looking at the outcome, and a protocol written only for the record after the run.
- **verdict**. The pass or fail of one preregistered rule set against the run's numbers, read off by code. Field: research method (our usage). Excludes an interpretation written after reading the numbers.
- **kill condition**. A rule written in the preregistration that, when met, decides against the claim with no room for judgement after the run. Field: research method (our usage, after Popper's falsification). Excludes reasons found after the run for setting a result aside.
- **study-year rate**. The share of one focal study's citations in one calendar year that were coded unsupported, as stored in the deposit without its denominator. Field: our usage for the deposited table. Excludes a count of citations, which the deposit does not hold.
- **rebuilt denominator**. The number of citations a focal study received in a year, recovered by finding the whole number n for which a published count k gives the deposited rate k over n. Field: our usage in this study. Excludes a count read directly from any source; years with a rate of zero share what remains of the study's total and are the least certain.
- **negative binomial model**. A regression for counts whose variance can exceed their mean, fitted here with a log link so that the fitted rate changes by a fixed factor per year. Field: statistics (count regression). Excludes models that weight each year by how many citations it holds.
- **within-study model**. A regression with its own baseline level for each focal study, so the year slope is estimated only from change inside studies. Field: statistics (fixed effects). Excludes change that comes from which focal studies happen to be cited in early and late years.
- **quasi-Poisson fit**. A log-linear count regression whose standard errors are scaled by the observed spread, used here for the within-study model. Field: statistics (count regression). Excludes a full probability model for the counts.

- **citation-weighted share model**. A regression of the number of unsupported citations out of all citations in each study-year on the year, so a year with thirty citations counts thirty times as much as a year with one. Field: statistics (logistic regression). Excludes the equal weight per study-year that a model of the deposited percentages gives.
- **study resampling**. Drawing focal studies with replacement, refitting the model each time, and taking the middle 95 percent of the refitted ratios as the interval. Field: statistics (cluster bootstrap). Excludes resampling single study-years, which would ignore that citations of one study move together.
- **control**. A version of the data with the structure under test removed (for example shuffled or pooled), scored the same way as the real data, so the real score has something to beat. Field: statistics and experimental design. Excludes a comparison with a different dataset or with a published number.
- **p value (from draws)**. The share of control draws scoring at or above the real score, with one added to the count and to the number of draws, so it is never zero. Field: statistics. Excludes a p value from a textbook formula that assumes a distribution.
- **planted check**. Data built so that the right answer is known, run through the same analysis before the real data, to show that the analysis can see the effect when it is there. Field: research method; called a positive control in experimental science. Excludes any check made on the real data, which cannot show what the analysis misses.
- **statistical power**. The chance that a test detects an effect of a given size when that effect is real, measured here on planted data with a known doubling. Field: statistics. Excludes the chance that an effect is real given that the test detected it.

16 References

[1] Karst, J., Jones, M. D. and Hoeksema, J. D. (2023). Positive citation bias and overinterpreted results lead to misinformation on common mycorrhizal networks in forests. Nature Ecology and Evolution 7, 501-511. doi:10.1038/s41559-023-01986-1

[2] Karst, J., Jones, M. D. and Hoeksema, J. D. (2023). Author Correction: Positive citation bias and overinterpreted results lead to misinformation on common mycorrhizal networks in forests. Nature Ecology and Evolution 7, 623. doi:10.1038/s41559-023-02035-7

[3] Karst, J., Jones, M. D. and Hoeksema, J. D. Deposited rate table (Data_submit.tab) and metadata accompanying Karst et al. (2023). Dataverse, CC0. doi:10.5683/SP3/88MZYX

[4] Klein, T., Rog, I., Livne-Luzon, S., van der Heijden, M. G. A. and Körner, C. (2023). Belowground carbon transfer across mycorrhizal networks among trees: facts, not fantasy. Open Research Europe 3, 168. doi:10.12688/openreseurope.16594.1

[5] Simard, S. W., Ryan, T. S. L. and Perry, D. A. (2025). Opinion: Response to questions about common mycorrhizal networks. Frontiers in Forests and Global Change 7, 1512518. doi:10.3389/ffgc.2024.1512518