Risk, Reward and Innovation in Peer Review

Kyle Siler
Rotman School of Management
University of Toronto
June 16, 2015
PEERE – Split, Croatia

MAIN SOURCES

Siler, Kyle and David Strang. “How Peer Review Chooses and Changes Published Science in *Administrative Science Quarterly*.” Working Paper.

Siler, Kyle, Kirby Lee and Lisa Bero. 2015. “Measuring the Effectiveness of Scientific Gatekeeping.” *Proceedings of the National Academy of Sciences*, Vol. 112(2), pp. 360-365.

STUDY ONE:

"Does Peer Review Have a Pro- or Anti-Innovation Influence?: Criticism and Changes to *Administrative Science Quarterly* Articles, 2005-2009"

The Importance of Peer Review

- Academic reward structures and hierarchies are ensconced in peer review.

The Importance of Peer Review

- Academic reward structures and hierarchies are ensconced in peer review.
- Quality exists, even if it is often difficult to define.

The Importance of Peer Review

- Academic reward structures and hierarchies are ensconced in peer review.
- Quality exists, even if it is often difficult to define.
- Serves *gatekeeping* and *gestational* functions.

Peer Review and Innovation

- “Courageous risk-takers” vs. “lazy conformists.”

Peer Review and Innovation

- “Courageous risk-takers” vs. “lazy conformists.”
- Innovation/*priority*: Being the first to claim a new idea or empirical turf is the coin of realm in science.

Conservativism in Scientific Organizations

- Even though many people in organizations desire innovation, their behaviors often belie this outwardly stated preference.

Conservativism in Scientific Organizations

- Even though many people in organizations desire innovation, their behaviors often belie this outwardly stated preference.
- Scholarly gatekeepers harbor conservative preferences and reinforce the status quo (Bourdieu, 1988; Lamont, 2009).

Conservativism in Scientific Organizations

- Even though many people in organizations desire innovation, their behaviors often belie this outwardly stated preference.
- Scholarly gatekeepers harbor conservative preferences and reinforce the status quo (Bourdieu, 1988; Lamont, 2009).
- Normal science (Kuhn, 1962) is productive!

Intellectual conservativism in peer review

- Survivor Biases

Intellectual conservativism in peer review

- Survivor Biases
- Psychological Resistance to Uncertainty

Intellectual conservativism in peer review

- Survivor Biases
- Psychological Resistance to Uncertainty
- Gatekeeper risk-aversion (Errors of omission vs. Errors of commission).

Intellectual conservativism in peer review

- Survivor Biases
- Psychological Resistance to Uncertainty
- Gatekeeper risk-aversion (Errors of omission vs. Errors of commission).
- Core Generalist Journals (like *ASQ*) have a mission to reach disparate audiences.

Intellectual conservativism in peer review

- Survivor Biases
- Psychological Resistance to Uncertainty
- Gatekeeper risk-aversion (Errors of omission vs. Errors of commission).
- Core Generalist Journals (like *ASQ*) have a mission to reach disparate audiences.
- Many scientific articles are ghostwritten by editors/reviewers.

Orienting Questions for Research

- Are different kinds of research treated and changed differently via the peer review process?

Orienting Questions for Research

- Are different kinds of research treated and changed differently via the peer review process?
- Do less conventional contributions face a more arduous peer review process?

Methods

- Survey first authors of *ASQ* articles 2005-2009.

Methods

- Survey first authors of *ASQ* articles 2005-2009.
- 52 Responses (67% Response Rate)

Methods

- Survey first authors of *ASQ* articles 2005-2009.
- 52 Responses (67% Response Rate)
- 37 also sent the first draft of their eventually published paper.

Methods

- Survey first authors of *ASQ* articles 2005-2009.
- 52 Responses (67% Response Rate)
- 37 also sent the first draft of their eventually published paper.
- Bibliometric and Textual Comparisons of Initial Submissions to Published Articles.

Methods

- Main independent variable is the *orthodoxy* of a manuscript. This is measured in 3 ways:

Methods

- Main independent variable is the *orthodoxy* of a manuscript. This is measured in 3 ways:
- Author Characterization

Methods

- Main independent variable is the *orthodoxy* of a manuscript. This is measured in 3 ways:
- Author Characterization
- Bibliographic Orthodoxy

Methods

- Main independent variable is the *orthodoxy* of a manuscript. This is measured in 3 ways:
- Author Characterization
- Bibliographic Orthodoxy
- Textual Orthodoxy

Methods

- Main independent variable is the *orthodoxy* of a manuscript. This is measured in 3 ways:
- Author Characterization
- Bibliographic Orthodoxy
- Textual Orthodoxy
- More common words/citations in the *ASQ* canon denote higher degrees of orthodoxy.

Self-Reports of Conventionality

T-Tests of Article Conventionality and Peer Review Arduousness			
	***ASQ* Revisions**	**Reported Criticisms**	**Reported Changes**
Test/Combine Existing Perspectives(s) (N=33)	1.88	3.42	3.41
New/Challenge Perspective (N=16)	2.25*	3.81*	3.94*

+ p < .10; * p <.05; ** p <.01; *** p <.001 (two-tailed tests).

Self-Reports of Conventionality

T-Tests of Article Conventionality and Peer Review Arduousness			
	***ASQ* Revisions**	**Reported Criticisms**	**Reported Changes**
Test/Combine Existing Perspectives(s) (N=33)	1.88	3.42	3.41
New/Challenge Perspective (N=16)	2.25*	3.81*	3.94*

) <.05; ** p <.01; *** p <.001 (two-tailed tests).

More subversive articles faced more arduous peer review.

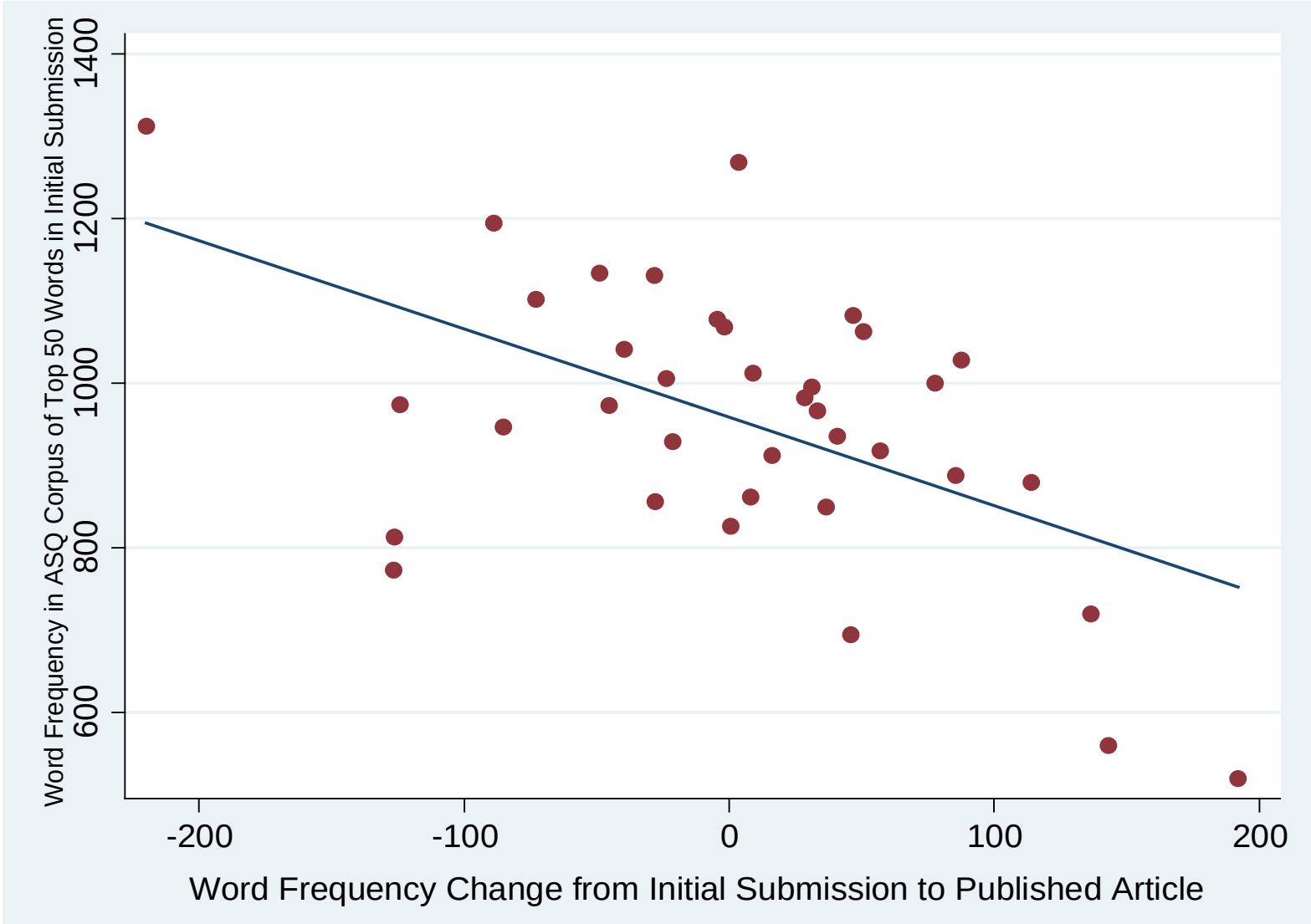
Text Orthodoxy and Arduousness

Correlations between Word Orthodoxy in Initial Submissions and Peer Review Arduousness			
	ASQ Revisions	Reported Criticisms	Reported Changes
Initial Frequency in *ASQ* Corpus of Top 10 Words	-0.18	0.13	0.02
Initial Frequency in *ASQ* Corpus of Top 50 Words	-0.39	-0.01	-0.16
Initial Frequency in *ASQ* Corpus of Top 100 Words	-0.33	0.04	-0.09

Text Orthodoxy and Arduousness

Correlations between Word Orthodoxy in Initial Submissions and Peer Review Arduousness			
	ASQ Revisions	Reported Criticisms	Reported Changes
Initial Frequency in *ASQ* Corpus of Top 10 Words	-0.18	0.13	0.02
Initial Frequency in *ASQ* Corpus of Top 50 Words	-0.39	-0.01	-0.16
Initial Frequency in *ASQ* Corpus of Top 100 Words	-0.33	0.04	-0.09

Relationship Between Word Orthodoxy in Initial Submission and Manuscript Changes Prior to Publication (r =-0.53)



Revisions, Criticisms and Changes in *Administrative Science Quarterly* Articles by Author and Paper Characteristics

	Revisions				Criticisms				Changes			
	Model 1	Model 2	Model 4	Model 5	Model 1	Model 2	Model 4	Model 5	Model 1	Model 2	Model 4	Model 5
AUTHOR CHARACTERISTICS												
Institutional Rank of Highest Ranked Author	-.005 (.006)	-.009 (.006)	-.006 (.009)	-.002 (.008)	-.002 (.005)	-.004 (.005)	.005 (.007)	.005 (.007)	.008 (.006)	.004 (.007)	.014+ (.008)	.015+ (.008)
Past *ASQ* Publications of Most Experienced Author	-.048 (.041)	-.078+ (.046)	-.056 (.048)	-.020 (.047)	-.022 (.037)	-.028 (.038)	.004 (.037)	.009 (.038)	-.015 (.043)	-.038 (.044)	.007 (.043)	.026 (.043)
ARTICLE CHARACTERISTICS												
Article Challenges Existing Paradigm		omitted				omitted				omitted		
Article Offers New Paradigm		.045 (.501)				-.759+ (.416)				-1.236* (.481)		
Article Combines Existing Paradigms		-.520 (.450)				-.821* (.378)				-1.243** (.438)		
Article Tests/Extends Existing Paradigm		-.570 (.458)				-1.201** (.383)				-1.417** (.439)		
Bibliographic Orthodoxy (Initial Submission)			-.034 (.040)				.000 (.031)				-.008 (.036)	
Text Orthodoxy (Top 50 Words - Initial Submission)				-.002* (.001)				-.000 (.001)				-.001 (.001)
Constant	2.251*** (.211)	2.716*** (.0441)	2.576*** (.404)	4.141*** (.862)	3.642*** (.196)	4.525*** (.366)	3.493*** (.344)	3.662*** (.724)	3.422*** (.234)	4.628*** (.424)	3.406*** (.397)	4.344*** (.786)
R-Squared	.033	.117	.058	.163	.009	.203	.020	.017	.053	.241	.099	.135
N	52	49	37	38	47	47	34	35	48	48	35	36

Revisions, Criticisms and Changes in *Administrative Science Quarterly* Articles by Author and Paper Characteristics

	Revisions				Criticisms				Changes			
	Model 1	Model 2	Model 4	Model 5	Model 1	Model 2	Model 4	Model 5	Model 1	Model 2	Model 4	Model 5
AUTHOR CHARACTERISTICS												
Institutional Rank of Highest Ranked Author	-.005 (.006)	-.009 (.006)	-.006 (.009)	-.002 (.008)	-.002 (.005)	-.004 (.005)	.005 (.007)	.005 (.007)	.008 (.006)	.004 (.007)	.014+ (.008)	.015+ (.008)
Past *ASQ* Publications of Most Experienced Author	-.048 (.041)	-.078+ (.046)	-.056 (.048)	-.020 (.047)	-.022 (.037)	-.028 (.038)	.004 (.037)	.009 (.038)	-.015 (.043)	-.038 (.044)	.007 (.043)	.026 (.043)
ARTICLE CHARACTERISTICS												
Article Challenges Existing Paradigm		omitted								omitted		
Article Offers New Paradigm		.045 (.501)				-.759+ (.416)				-1.236* (.481)		
Article Combines Existing Paradigms		-.520 (.450)				-.821* (.378)				-1.243** (.438)		
Article Tests/Extends Existing Paradigm		-.570 (.458)				-1.201** (.383)				-1.417** (.439)		
Bibliographic Orthodoxy (Initial Submission)			-.034 (.040)				.000 (.031)				-.008 (.036)	
Text Orthodoxy (Top 50 Words - Initial Submission)				-.002* (.001)				-.000 (.001)				-.001 (.001)
Constant	2.251*** (.211)	2.716*** (.0441)	2.576*** (.404)	4.141*** (.862)	3.642*** (.196)	4.525*** (.366)	3.493*** (.344)	3.662*** (.724)	3.422*** (.234)	4.628*** (.424)	3.406*** (.397)	4.344*** (.786)
R-Squared	.033	.117	.058	.163	.009	.203	.020	.017	.053	.241	.099	.135
N	52	49	37	38	47	47	34	35	48	48	35	36

No major author effects.

Revisions, Criticisms and Changes in *Administrative Science Quarterly* Articles by Author and Paper Characteristics

	Revisions				Criticisms				Changes			
	Model 1	Model 2	Model 4	Model 5	Model 1	Model 2	Model 4	Model 5	Model 1	Model 2	Model 4	Model 5
AUTHOR CHARACTERISTICS												
Institutional Rank of Highest Ranked Author	-.005 (.006)	-.009 (.006)	-.006 (.009)	-.002 (.008)	-.002 (.005)	-.004 (.005)	.005 (.007)	.005 (.007)	.008 (.006)	.004 (.007)	.014+ (.008)	.015+ (.008)
Past *ASQ* Publications of Most Experienced Author	-.048 (.041)	-.078+ (.046)	-.056 (.048)	-.020 (.047)	-.022 (.037)	-.028 (.038)	.004 (.037)	.009 (.038)	-.015 (.043)	-.038 (.044)	.007 (.043)	.026 (.043)
ARTICLE CHARACTERISTICS												
Article Challenges Existing Paradigm		omitted				omitted				omitted		
Article Offers New Paradigm		.045 (.501)				-.759+ (.416)				-1.236* (.481)		
Article Combines Existing Paradigms		-.520 (.450)				-.821* (.378)				-1.243** (.438)		
Article Tests/Extends Existing Paradigm		-.570 (.458)				-1.201** (.383)				-1.417** (.439)		
Bibliographic Orthodoxy (Initial Submission)												
Text Orthodoxy (Top 50 Words - Initial Submission)				-.002* (.001)				-.000 (.001)				-.001 (.001)
Constant	2.251*** (.211)	2.716*** (.0441)	2.576*** (.404)	4.141*** (.862)	3.642*** (.196)	4.525*** (.366)	3.493*** (.344)	3.662*** (.724)	3.422*** (.234)	4.628*** (.424)	3.406*** (.397)	4.344*** (.786)
R-Squared	.033	.117	.058	.163	.009	.203	.020	.017	.053	.241	.099	.135
N	52	49	37	38	47	47	34	35	48	48	35	36

More "subversive" articles reported more criticisms and changes.

Revisions, Criticisms and Changes in *Administrative Science Quarterly* Articles by Author and Paper Characteristics

	Revisions				Criticisms				Changes			
	Model 1	Model 2	Model 4	Model 5	Model 1	Model 2	Model 4	Model 5	Model 1	Model 2	Model 4	Model 5
AUTHOR CHARACTERISTICS												
Institutional Rank of Highest Ranked Author	-.005 (.006)	-.009 (.006)	-.006 (.009)	-.002 (.008)	-.002 (.005)	-.004 (.005)	.005 (.007)	.005 (.007)	.008 (.006)	.004 (.007)	.014+ (.008)	.015+ (.008)
Past *ASQ* Publications of Most Experienced Author	-.048 (.041)	-.078+ (.046)	-.056 (.048)	-.020 (.047)	-.022 (.037)	-.028 (.038)	.004 (.037)	.009 (.038)	-.015 (.043)	-.038 (.044)	.007 (.043)	.026 (.043)
ARTICLE CHARACTERISTICS												
Article Challenges Existing Paradigm		omitted				omitted				omitted		
Article Offers New Paradigm		.045 (.501)				-.759+ (.416)				-1.236* (.481)		
Article Combines Existing Paradigms		-.520 (.450)				-.821* (.378)				-1.243** (.438)		
Article Tests/Extends Existing Paradigm		-.570 (.458)				-1.201** (.383)				-1.417** (.439)		
Bibliographic Orthodoxy (Initial Submission)			-.034 (.040)				.000 (.031)				-.008 (.036)	
Text Orthodoxy (Top 50 Words - Initial Submission)				-.002* (.001)				-.000 (.001)				-.001 (.001)
Co	(.211)	* (.0441)	* (.404)	41** * (.862)	3.642*** (.196)	4.525*** (.366)	3.493** * (.344)	3.662*** (.724)	3.422*** (.234)	4.628*** (.424)	3.406** * (.397)	4.344** * (.786)
R-Squared	.033	.117	.058	.163	.009	.203	.020	.017	.053	.241	.099	.135
N	52	49	37	38	47	47	34	35	48	48	35	36

Less orthodox text begets more revisions.

How do different types of articles fare?

Post-Publication Citations/year by Paradigm Identification (Relative to "Offers new perspective")		
	Citations/year	Std. Dev.
Tests or extends a perspective (N = 15)	-7.46*	3.18
Combines two or more perspectives (N = 18)	-6.74*	3.97
Challenges a perspective (N = 5)	-3.20	4.32
Constant	13.38***	2.42
R-Squared	.131	

How do different types of articles fare?

Post-Publication Citations/year by Paradigm Identification (Relative to "Offers new perspective")		
	Citations/year	Std. Dev.
Tests or extends a perspective (N = 15)	-7.46*	3.18
Combines two or more perspectives (N = 18)	-6.74*	3.97
Challenges a perspective (N = 5)	-3.20	4.32
Constant		42
R-Squared	.131	

More conventional articles received fewer citations.

How do different types of articles fare?

Post-Publication Citations/year by Paradigm Identification (Relative to "Offers new perspective")		
	Citations/year	Std. Dev.
Tests or extends a perspective (N = 15)	-7.46*	3.18
Combines two or more perspectives (N = 18)	-6.74*	3.97
Challenges a perspective (N = 5)	-3.20	4.32
Constant		
R-Squared		

More conventional articles have lower variance in citation outcomes.

Results Summary

- Less conventional articles faced more revisions, criticisms and changes.

Results Summary

- Less conventional articles faced more revisions, criticisms and changes.
- Articles tend to be revised in a *conventional* manner; not necessarily conservatively.

Implications

- Is academia risk-averse?

Implications

- Is academia risk-averse?
- Incentives for pursuing different types of work.

Implications

- Is academia risk-averse?
- Incentives for pursuing different types of work.
- How effective are the gestational functions of peer review? Are they worth the costs?

Implications

- Is academia risk-averse?
- Incentives for pursuing different types of work.
- How effective are the gestational functions of peer review? Are they worth the costs?
- Are most new ideas bad ideas?

Implications

- Is academia risk-averse?
- Incentives for pursuing different types of work.
- How effective are the gestational functions of peer review? Are they worth the costs?
- Are most new ideas bad ideas?
- Is homogenizing manuscripts a good thing?

Future Work

- Study of rejected articles.

STUDY TWO:

Measuring the Effectiveness of Peer Review in Elite Medical Journals

Methods and Data

- 1,008 submissions to *Annals of Internal Medicine*, *British Medical Journal* and *The Lancet*.

Methods and Data

- 1,008 submissions to *Annals of Internal Medicine*, *British Medical Journal* and *The Lancet*.
- 772 Desk-rejected, 174 rejected after peer review, 62 published.

Methods and Data

- 1,008 submissions to *Annals of Internal Medicine*, *British Medical Journal* and *The Lancet*.
- 772 Desk-rejected, 174 rejected after peer review, 62 published.
- 819 submissions were eventually published somewhere.

Methods and Data

- Important to analyze rejections and acceptances separately to avoid Matthew Effects.

Methods and Data

- Important to analyze rejections and acceptances separately to avoid Matthew Effects.
- Journal impact factor correlated 0.54 with citations received.

Methods and Data

- Important to analyze rejections and acceptances separately to avoid Matthew Effects.
- Journal impact factor correlated 0.54 with citations received.
- Dependent Variable: Citations received.

Methods and Data

- Important to analyze rejections and acceptances separately to avoid Matthew Effects.
- Journal impact factor correlated 0.54 with citations received.
- Dependent Variable: Citations received.
- 2 quality control filters: Editorial desk-rejections and peer reviewers.

Rejections

- Desk-rejections (N=571) averaged 69.8 citations vs. 94.7 for articles sent for peer review prior to rejection (N=187; $p < .01$).

Rejections

- Desk-rejections (N=571) averaged 69.8 citations vs. 94.7 for articles sent for peer review prior to rejection (N=187; $p < .01$).
- Logged citations: Desk rejections averaged 3.44; 3.92 for peer reviewed rejections ($p < .001$).

Rejections

- Desk-rejections (N=571) averaged 69.8 citations vs. 94.7 for articles sent for peer review prior to rejection (N=187; $p < .01$).
- Logged citations: Desk rejections averaged 3.44; 3.92 for peer reviewed rejections ($p < .001$).
- 12 of the 15 most highly-cited manuscripts were desk-rejected.

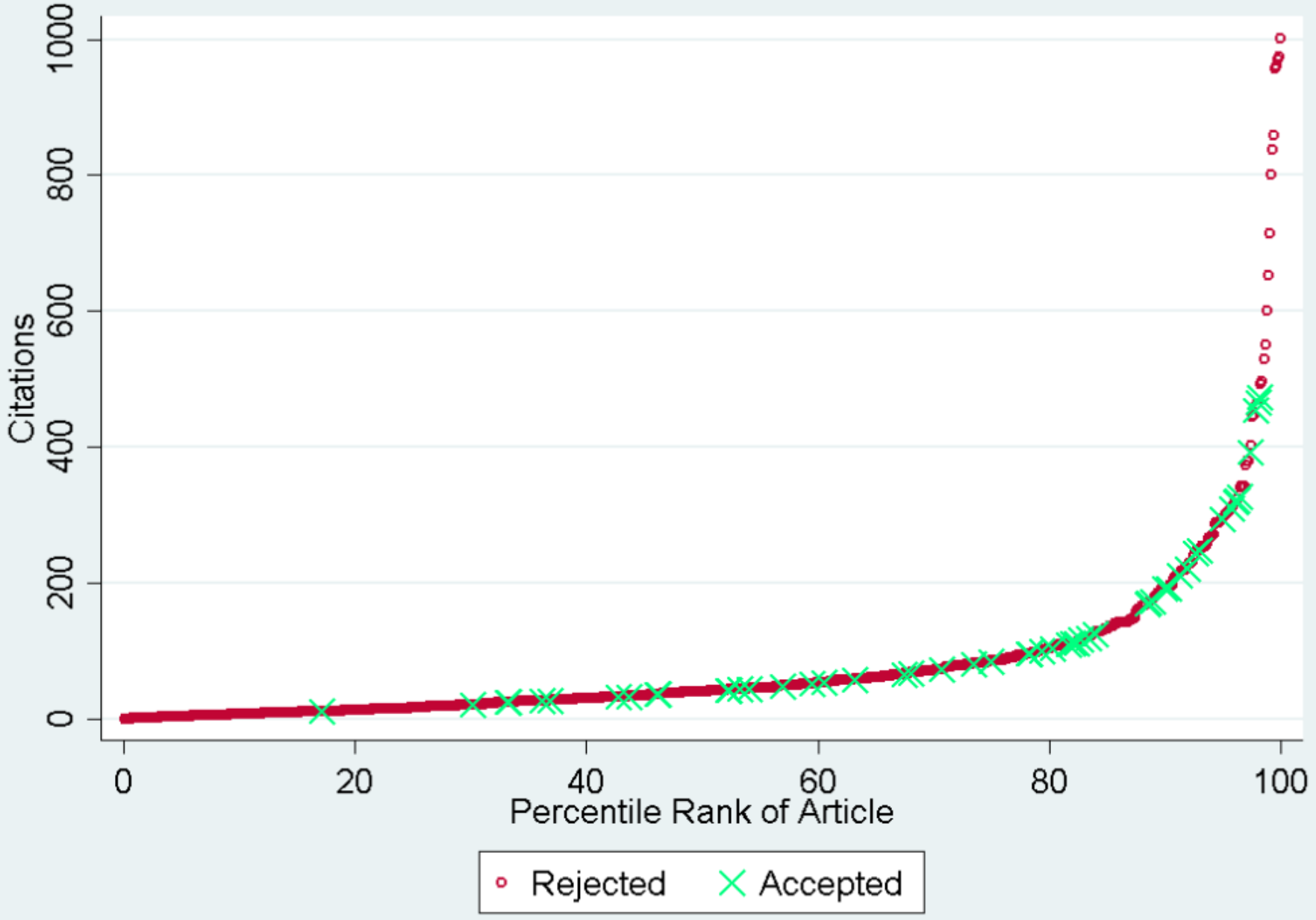
Accepted Manuscripts

- Published articles without a “reject” recommendation (N=30) averaged 162.8 citations, compared to 115.2 for articles with one “reject” recommendation (N=21; $p < .10$).

Accepted Manuscripts

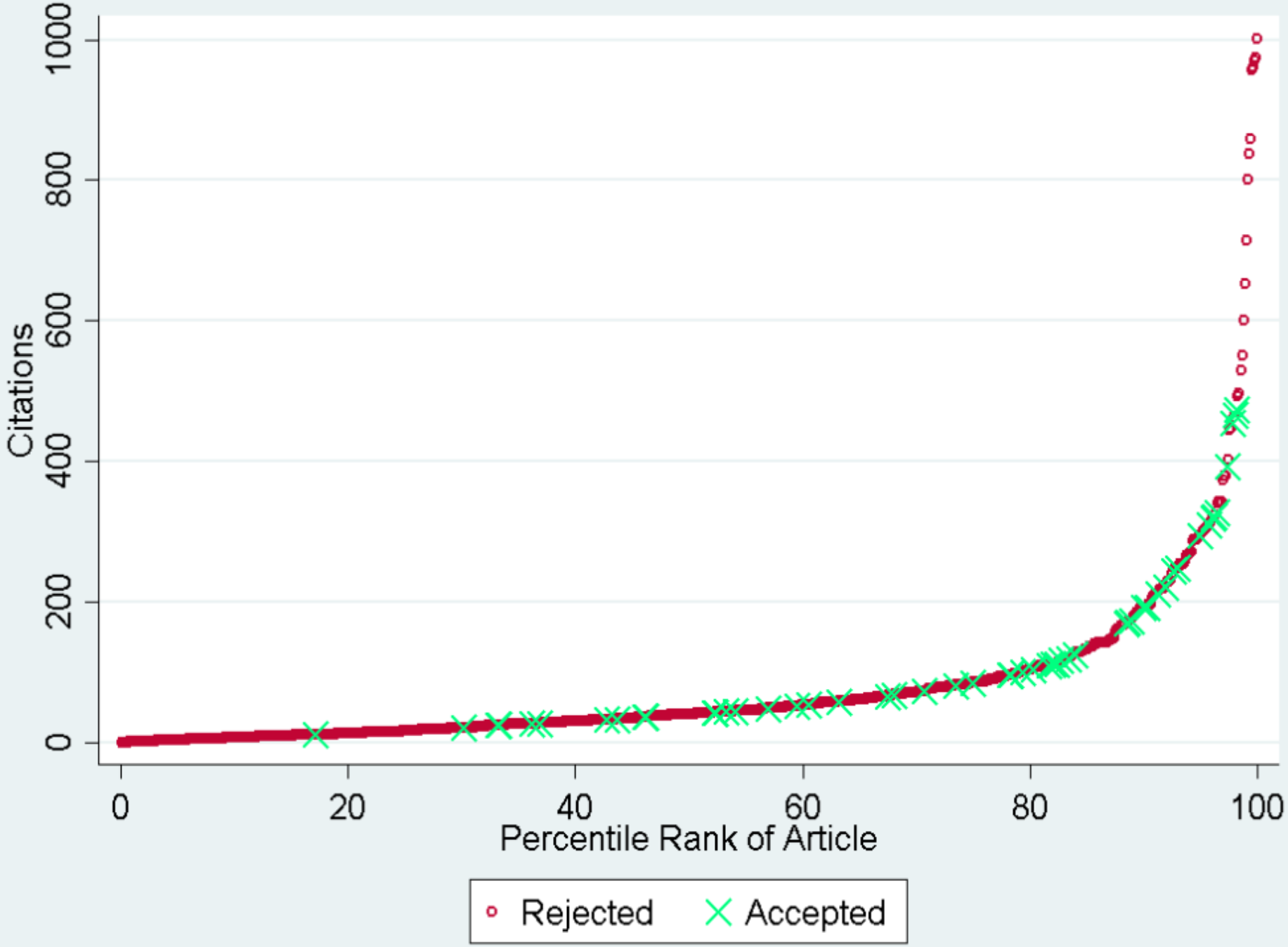
- Published articles without a “reject” recommendation (N=30) averaged 162.8 citations, compared to 115.2 for articles with one “reject” recommendation (N=21; $p < .10$).
- Small N; second filter level.

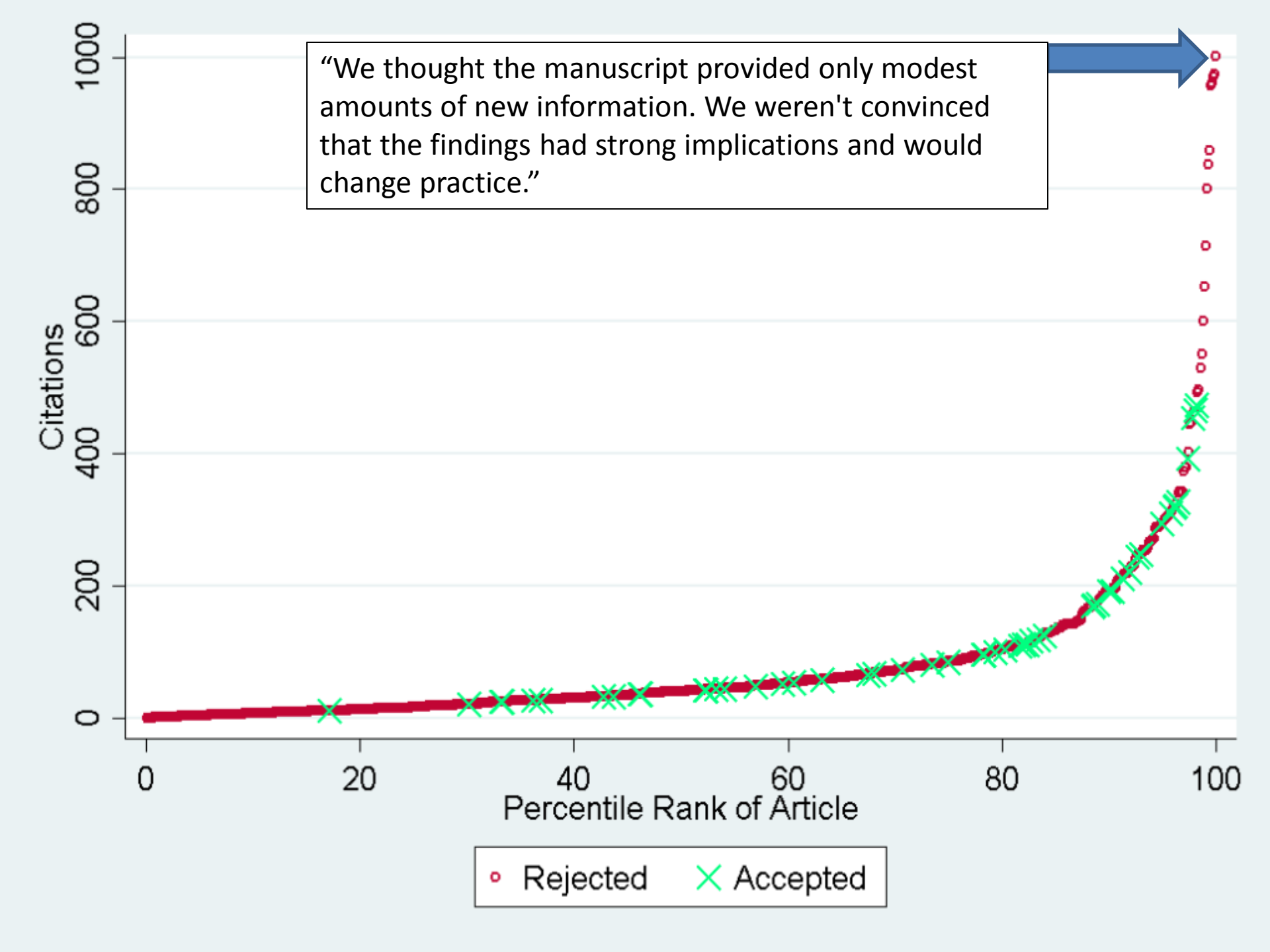
Fates of Accepted vs. Rejected Manuscripts

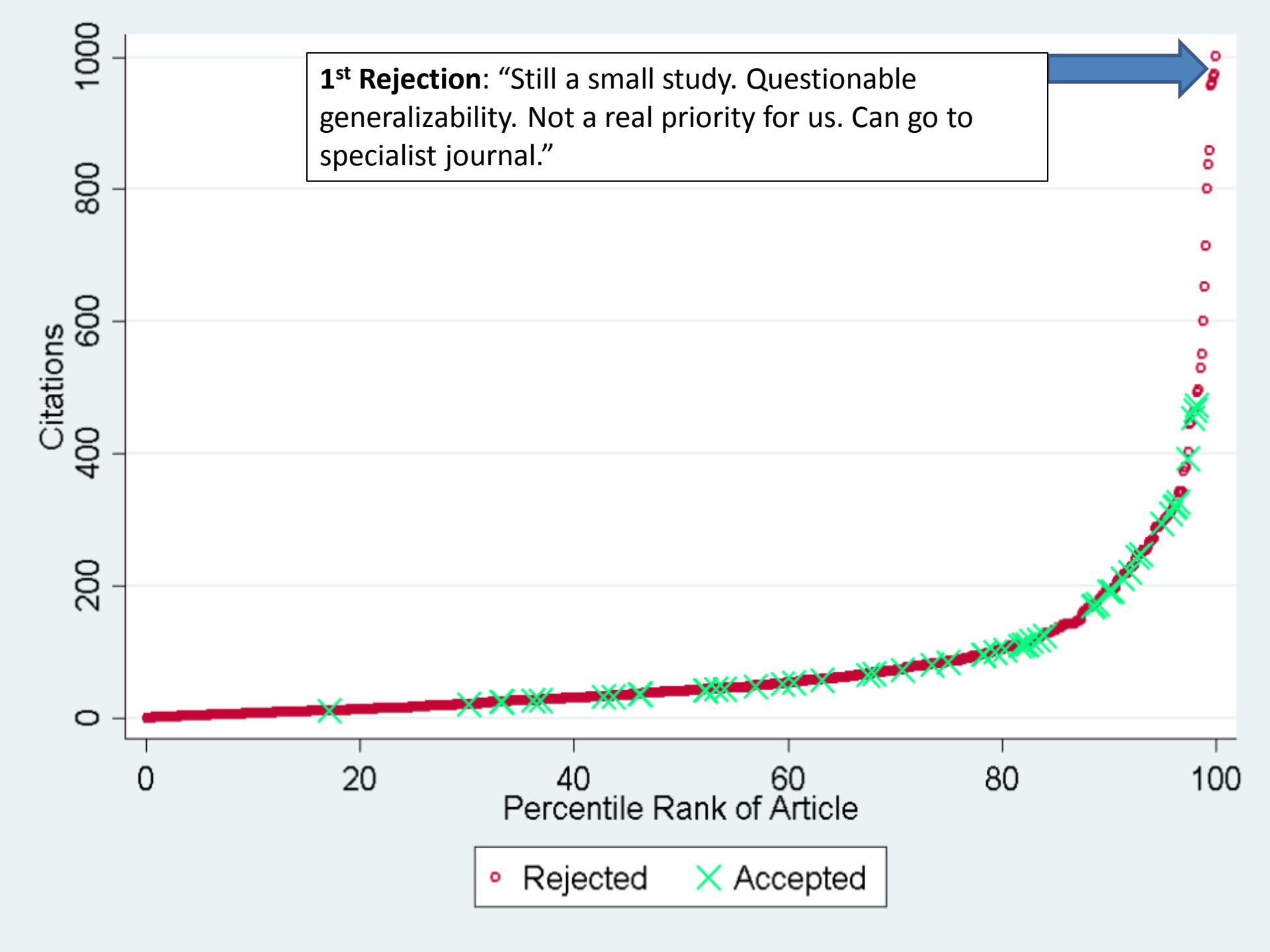


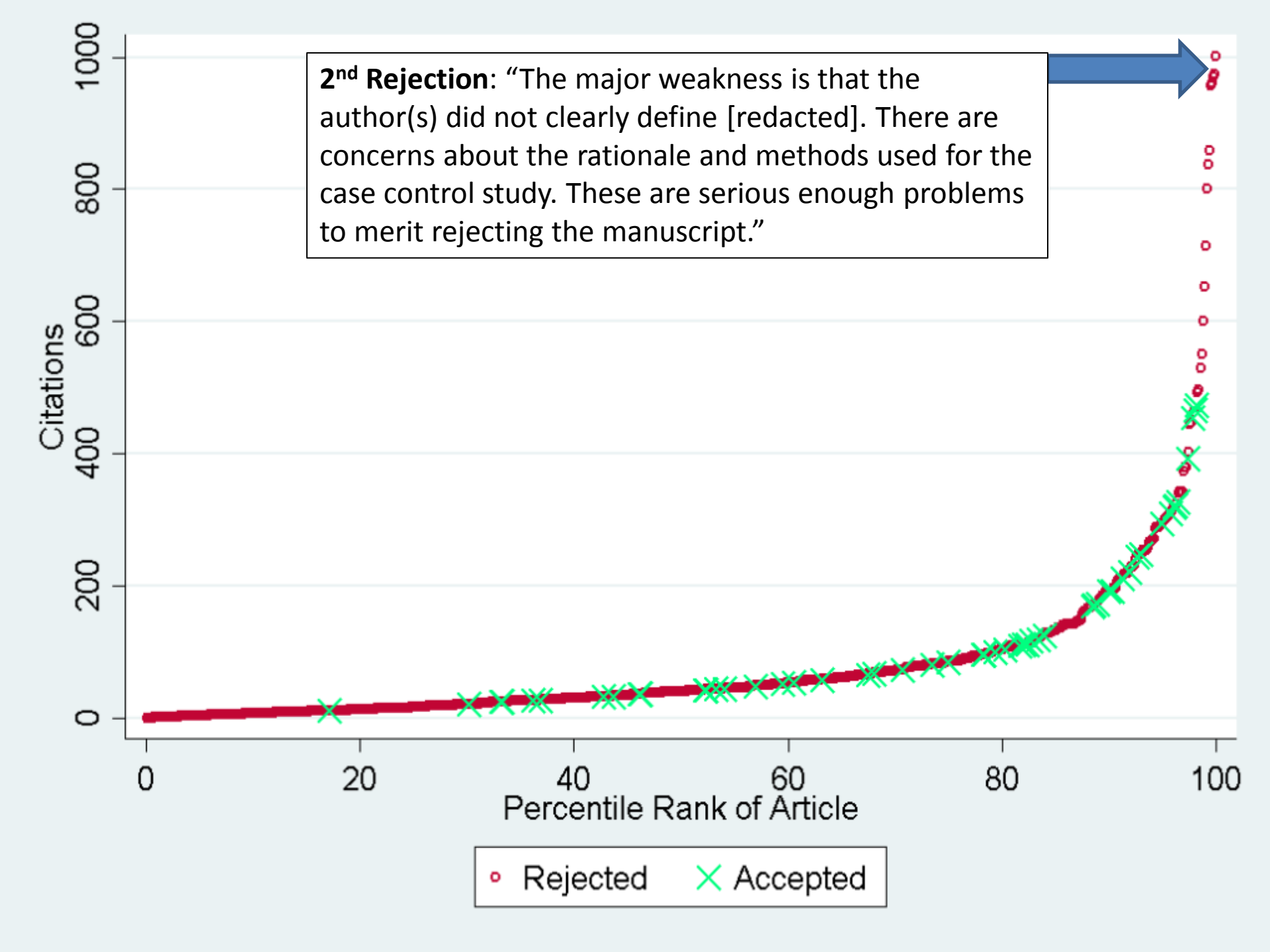
Fates of Accepted vs. Rejected Manuscripts

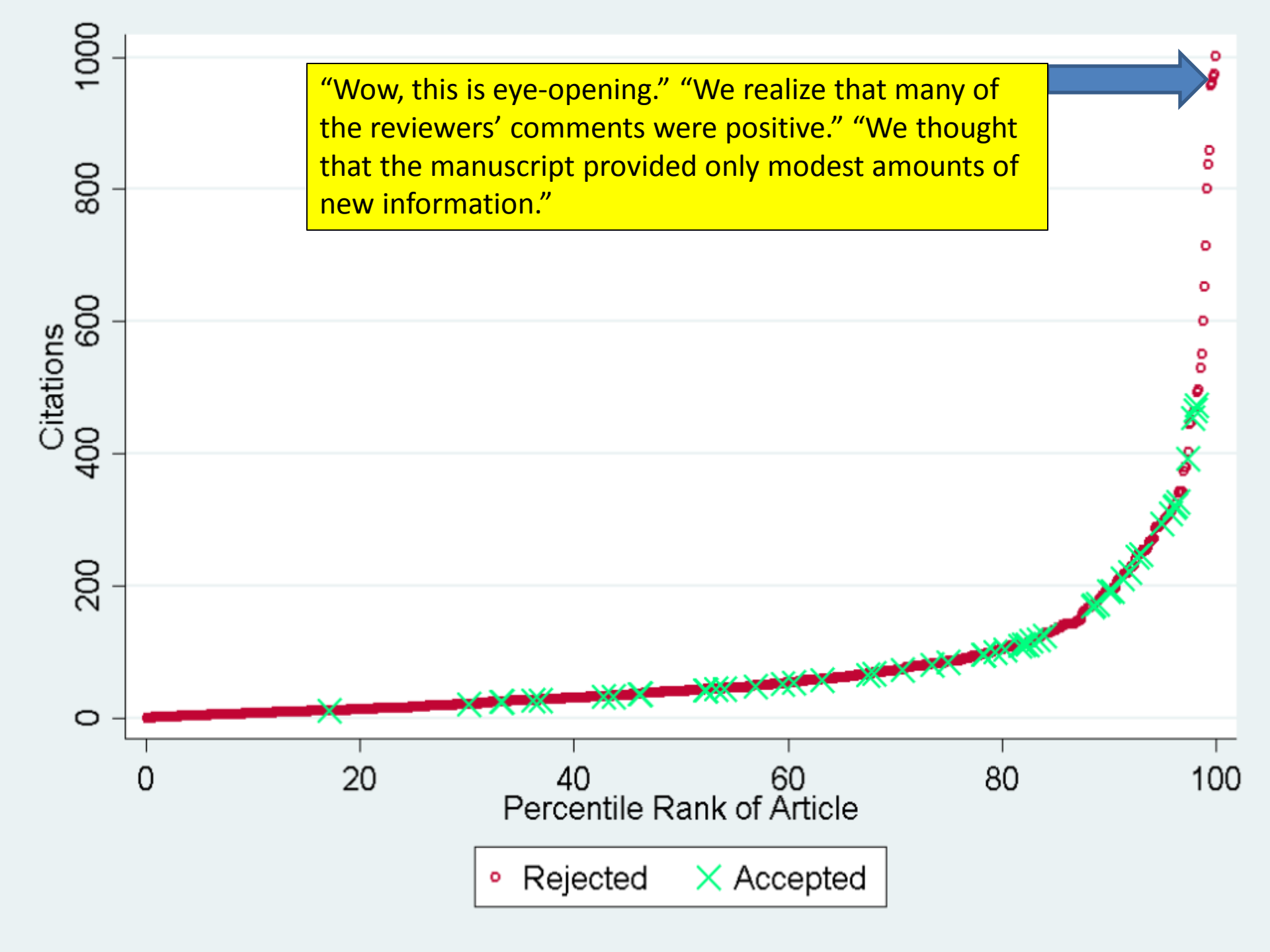
- Mean percentile of acceptances: 71.9
- Median percentile of acceptances: 79.4
- 25^{th} percentile: 53.3
- 75^{th} percentile: 91.2
- Least-cited acceptance: 17.1 percentile
- Most-cited acceptance: 98.2 percentile

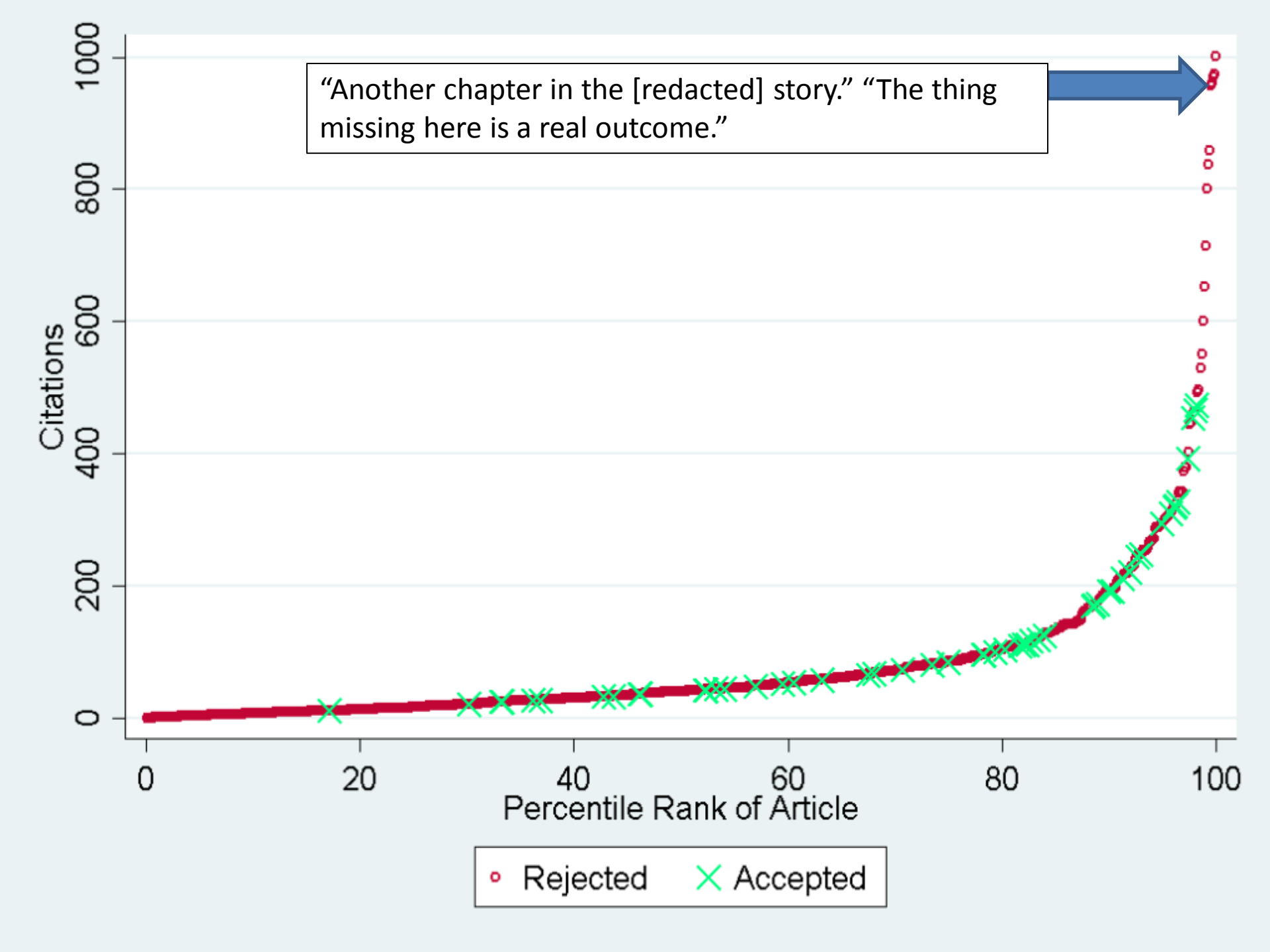

Citations
0
200
400
600
800
1000
0
20
40
60
80
100
Percentile Rank of Article
Rejected
Accepted

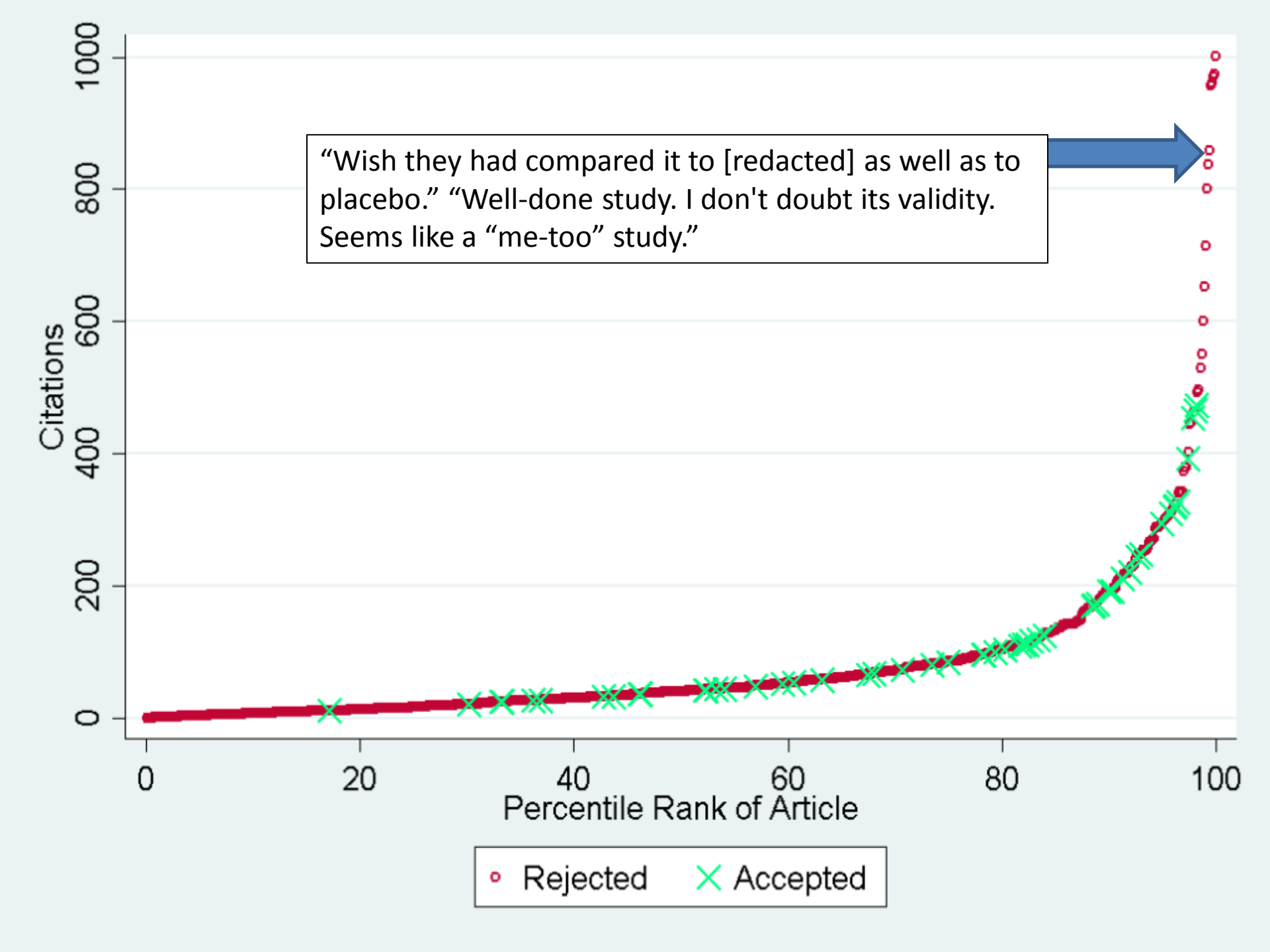

"We thought the manuscript provided only modest amounts of new information. We weren't convinced that the findings had strong implications and would change practice."
Citations
0
200
400
600
800
1000
0
20
40
60
80
100
Percentile Rank of Article
Rejected
Accepted

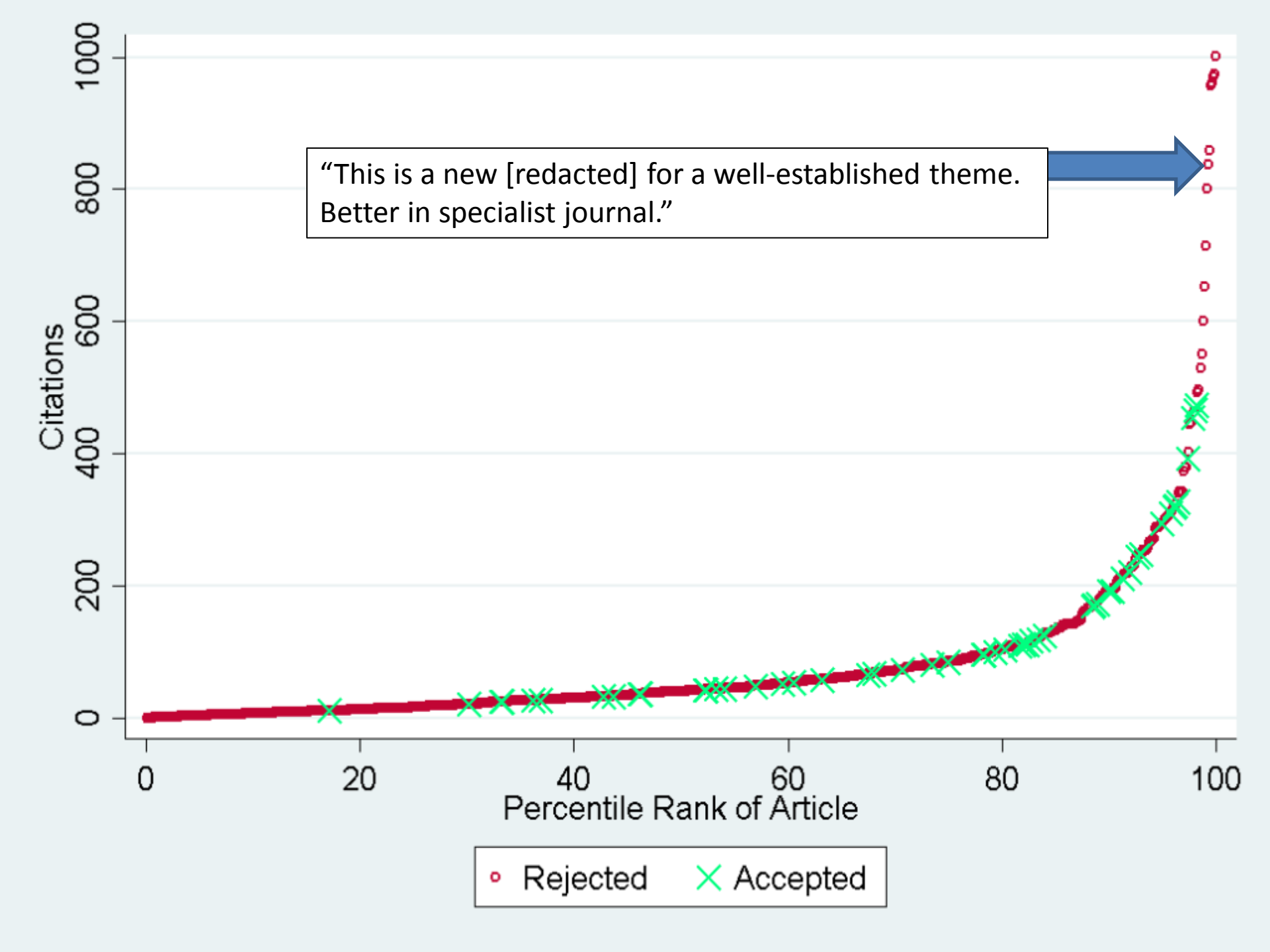

1st Rejection: "Still a small study. Questionable generalizability. Not a real priority for us. Can go to specialist journal."
Citations
0
200
400
600
800
1000
0
20
40
60
80
100
Percentile Rank of Article
Rejected
Accepted

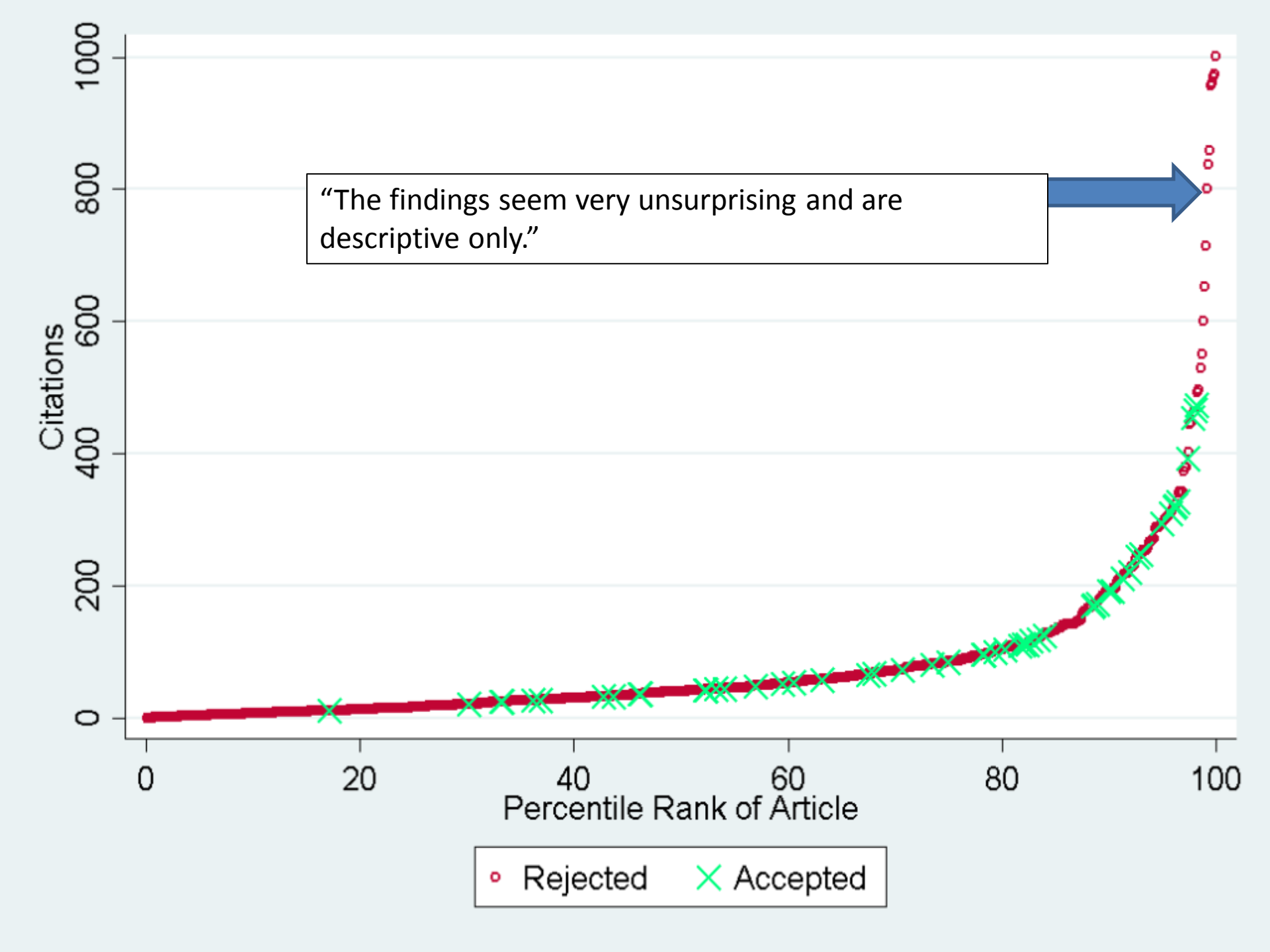

2nd Rejection: "The major weakness is that the author(s) did not clearly define [redacted]. There are concerns about the rationale and methods used for the case control study. These are serious enough problems to merit rejecting the manuscript."
Citations
0
200
400
600
800
1000
0
20
40
60
80
100
Percentile Rank of Article
Rejected
Accepted

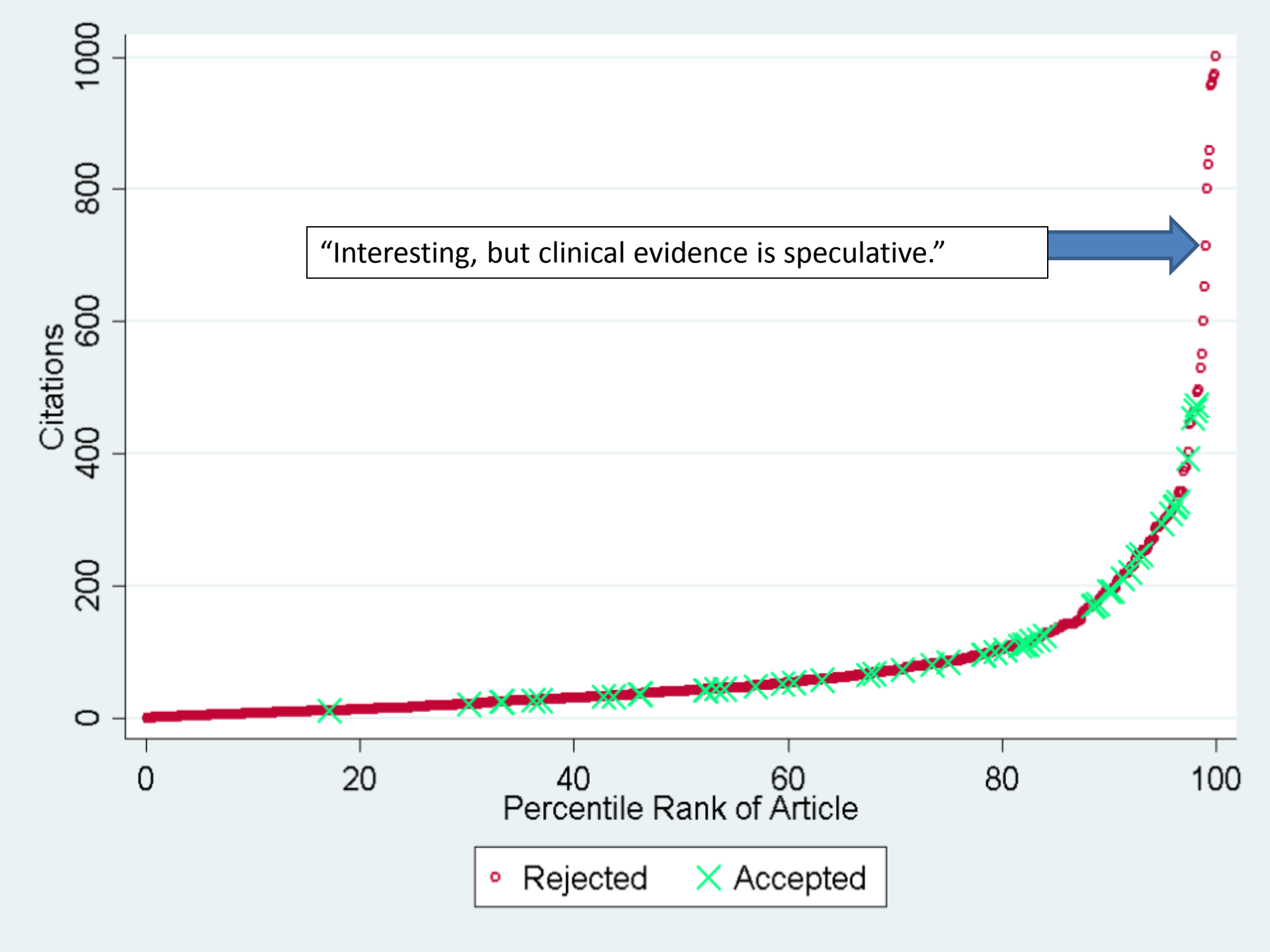

"Wow, this is eye-opening." "We realize that many of the reviewers' comments were positive." "We thought that the manuscript provided only modest amounts of new information."
Citations
0
200
400
600
800
1000
0
20
40
60
80
100
Percentile Rank of Article
Rejected
Accepted

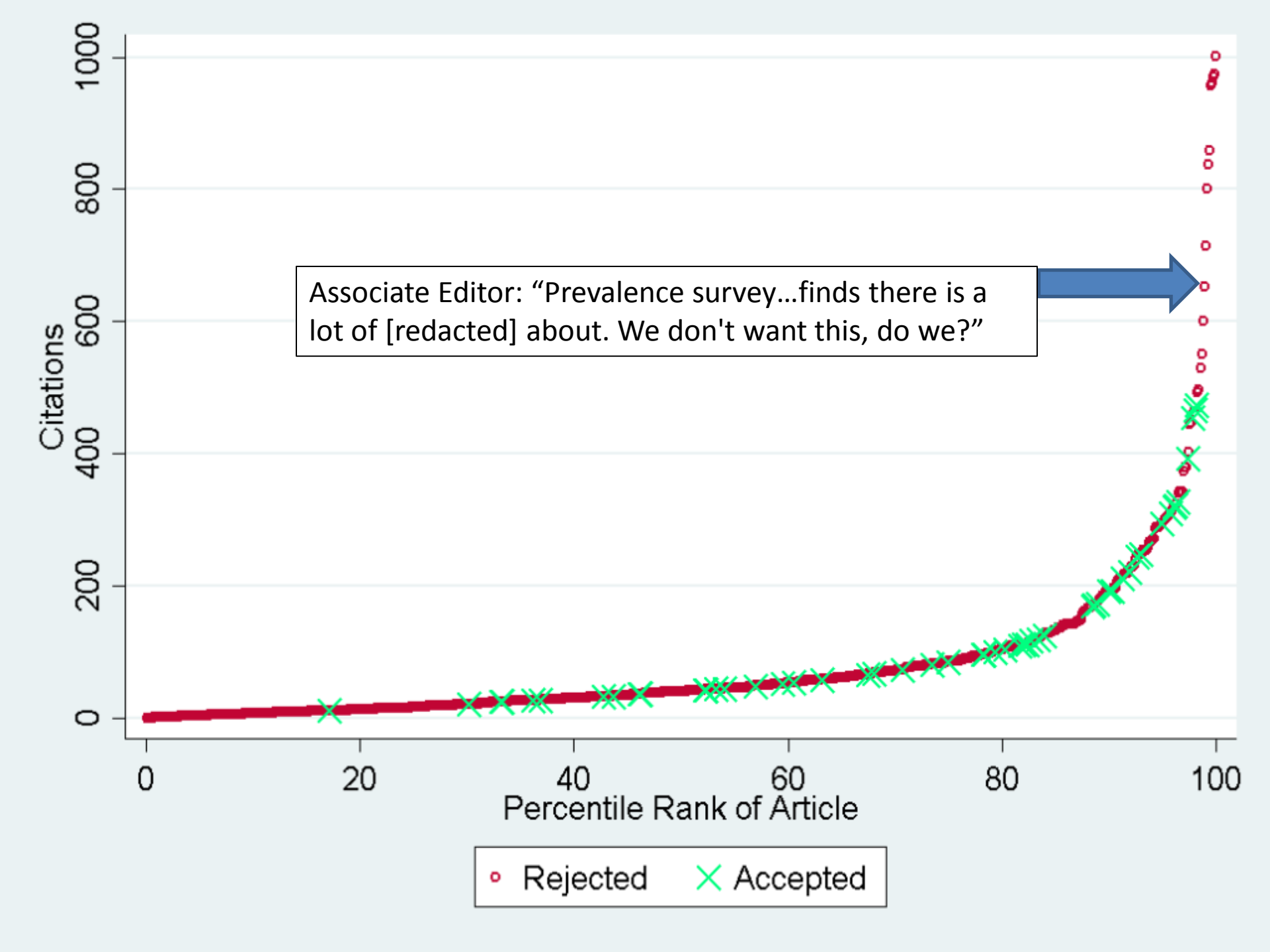

"Another chapter in the [redacted] story." "The thing missing here is a real outcome."
Citations
0
200
400
600
800
1000
0
20
40
60
80
100
Percentile Rank of Article
Rejected
Accepted

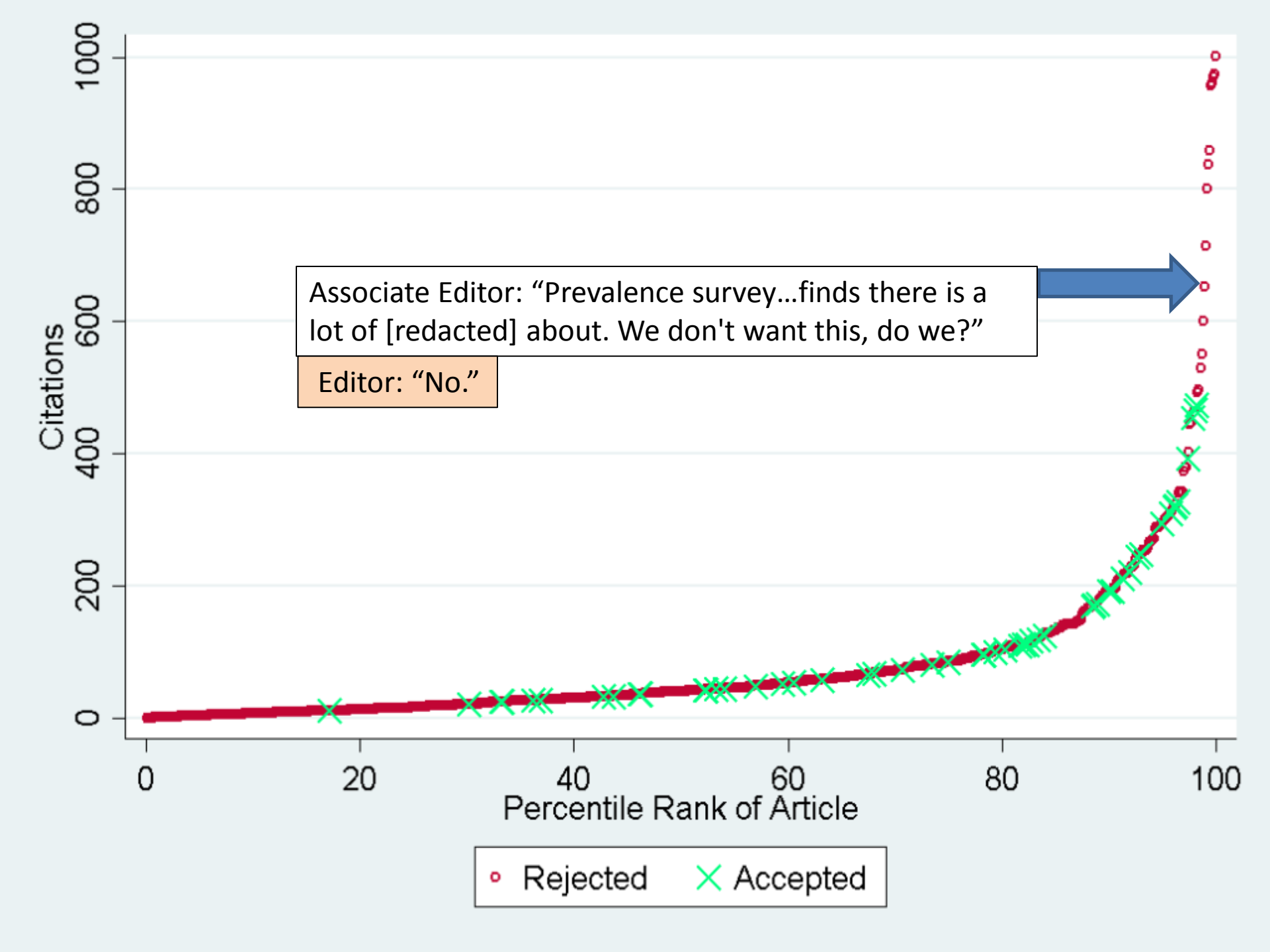

"Wish they had compared it to [redacted] as well as to placebo." "Well-done study. I don't doubt its validity. Seems like a "me-too" study."
Citations
0
200
400
600
800
1000
0
20
40
60
80
100
Percentile Rank of Article
Rejected
Accepted

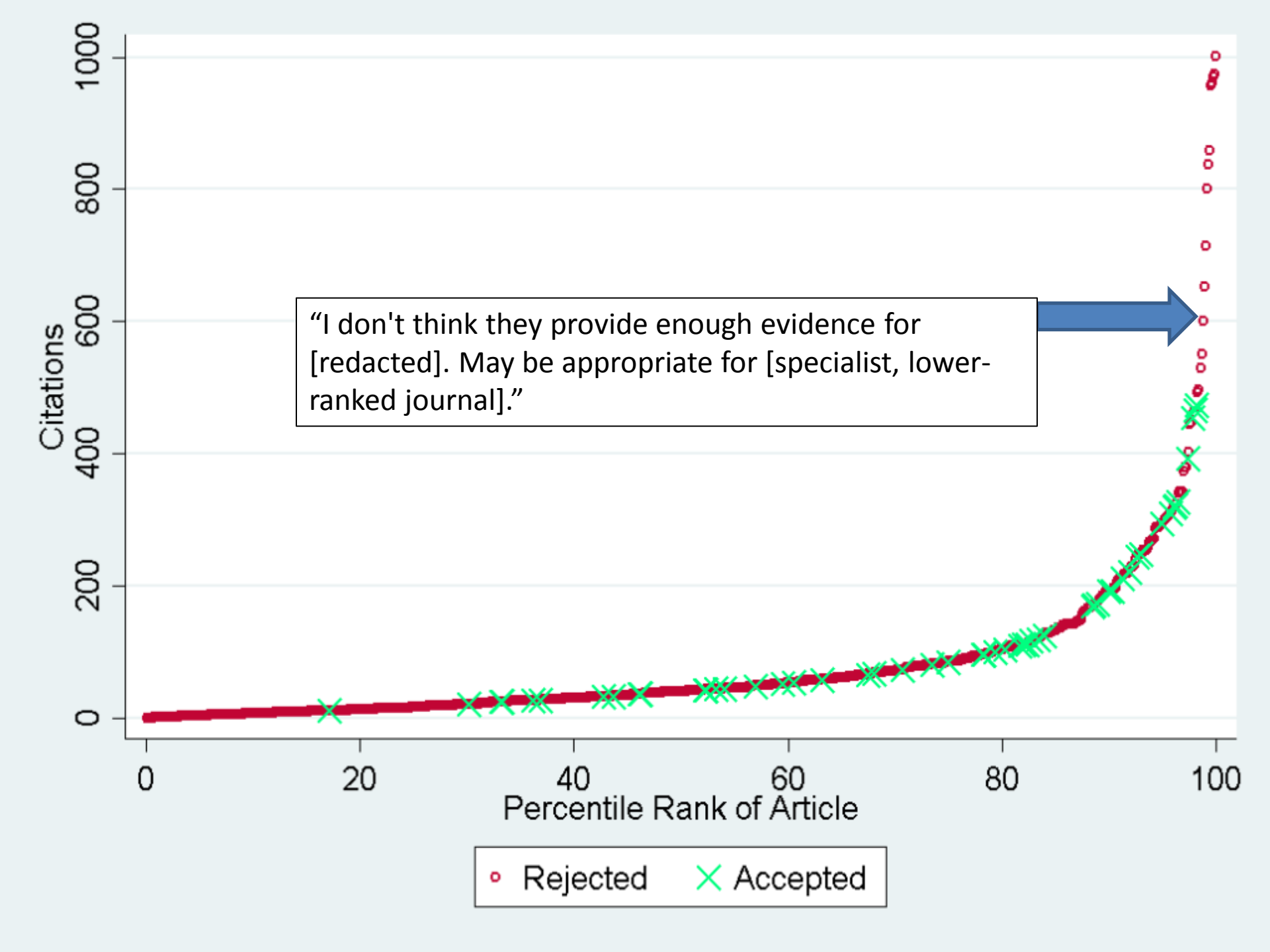

"This is a new [redacted] for a well-established theme. Better in specialist journal."
Citations
0
200
400
600
800
1000
0
20
40
60
80
100
Percentile Rank of Article
Rejected
Accepted

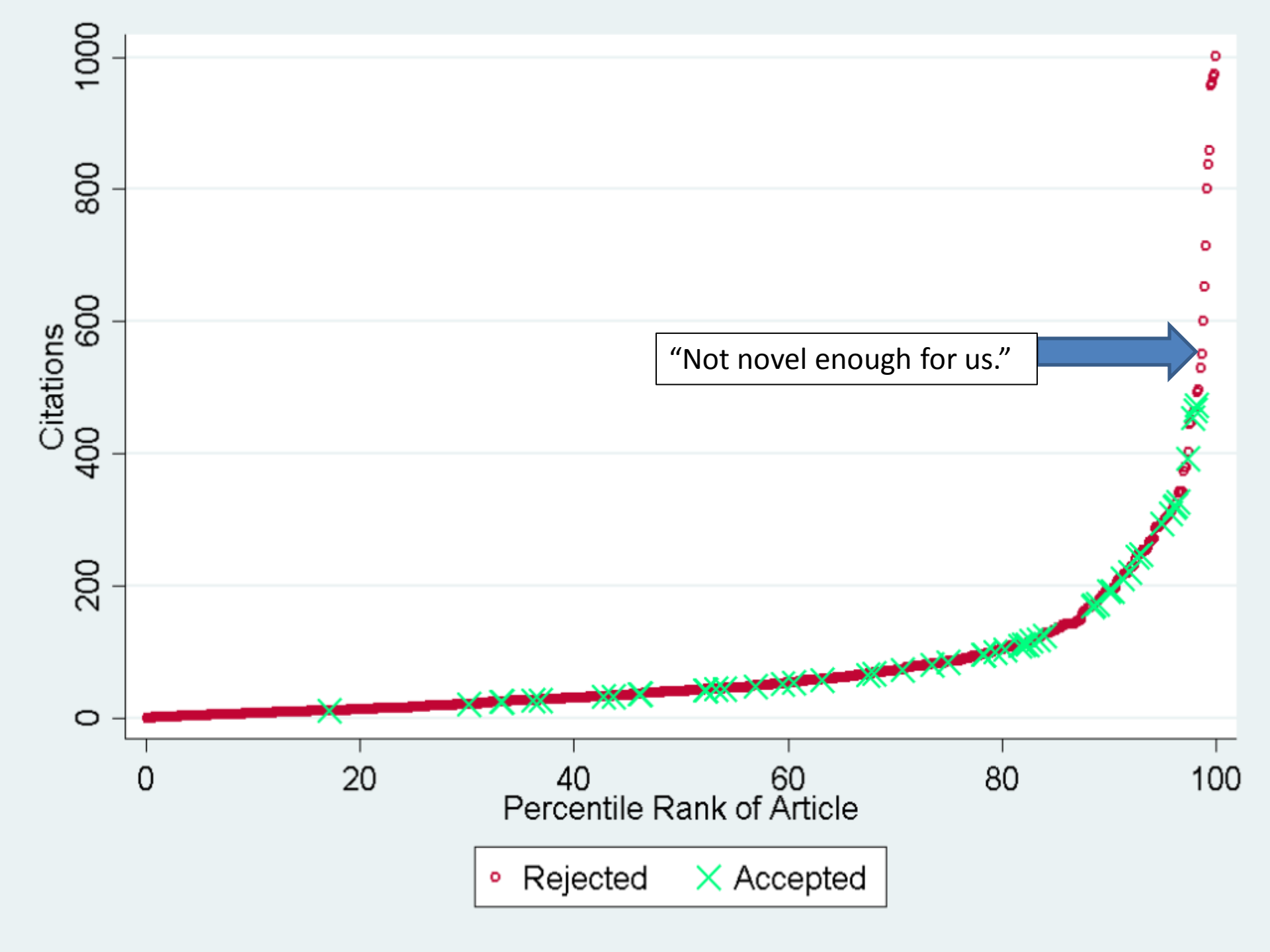

"The findings seem very unsurprising and are descriptive only."
Citations
0
200
400
600
800
1000
0
20
40
60
80
100
Percentile Rank of Article
Rejected
Accepted

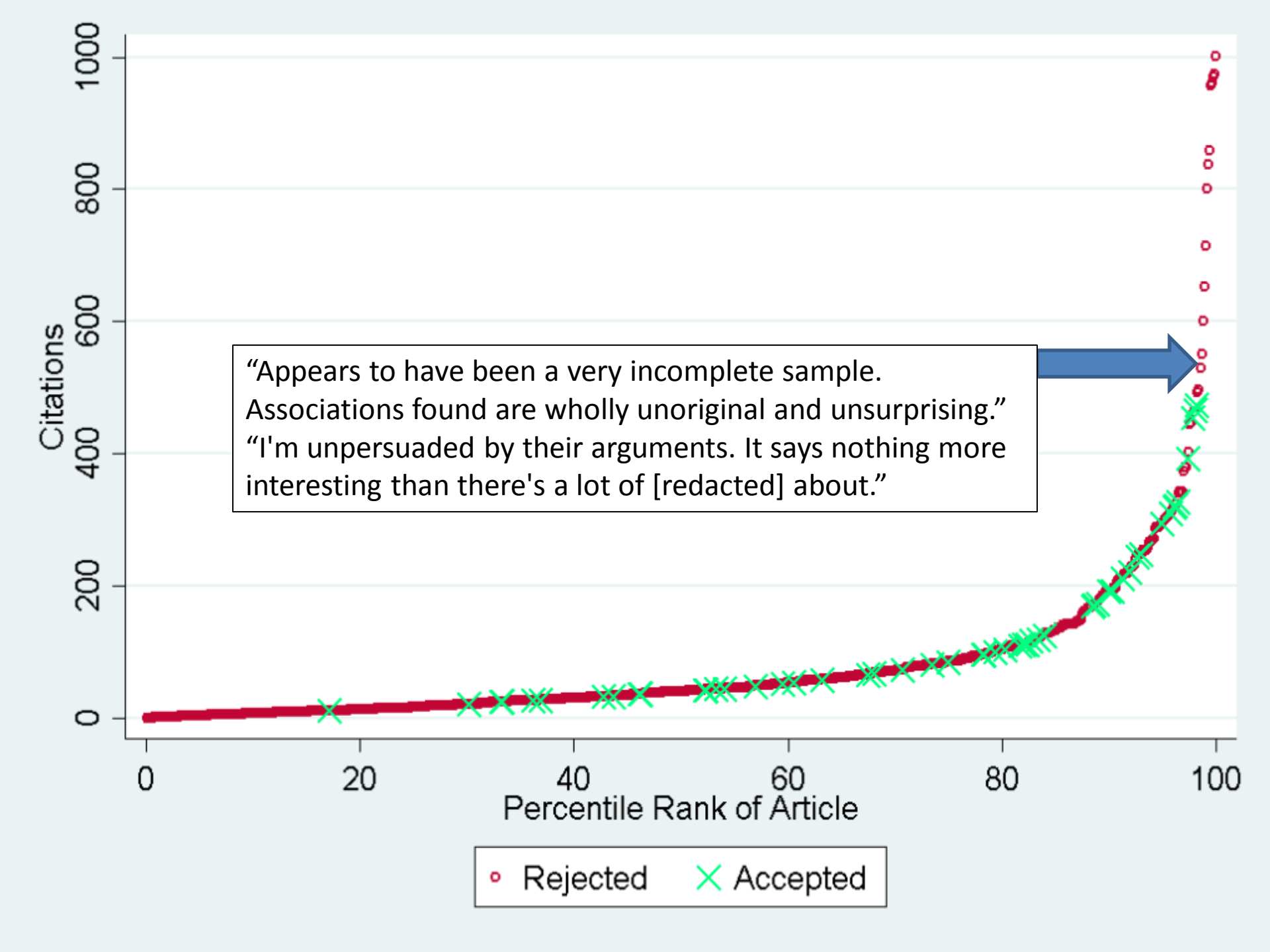

"Interesting, but clinical evidence is speculative."
Citations
0
200
400
600
800
1000
0
20
40
60
80
100
Percentile Rank of Article
Rejected
Accepted

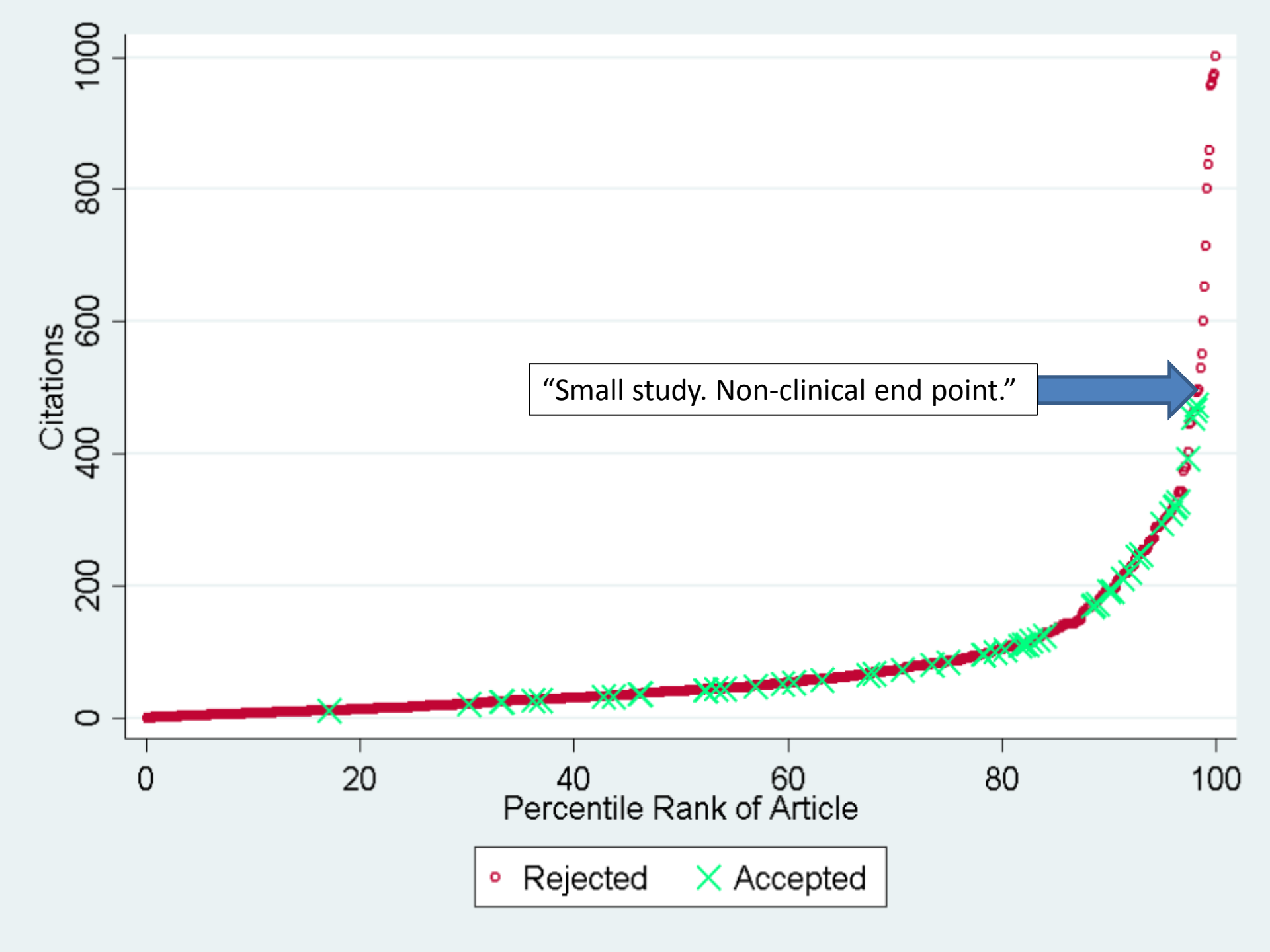

Associate Editor: “Prevalence survey…finds there is a lot of [redacted] about. We don't want this, do we?”
Citations
0
200
400
600
800
1000
0
20
40
60
80
100
Percentile Rank of Article
Rejected
Accepted

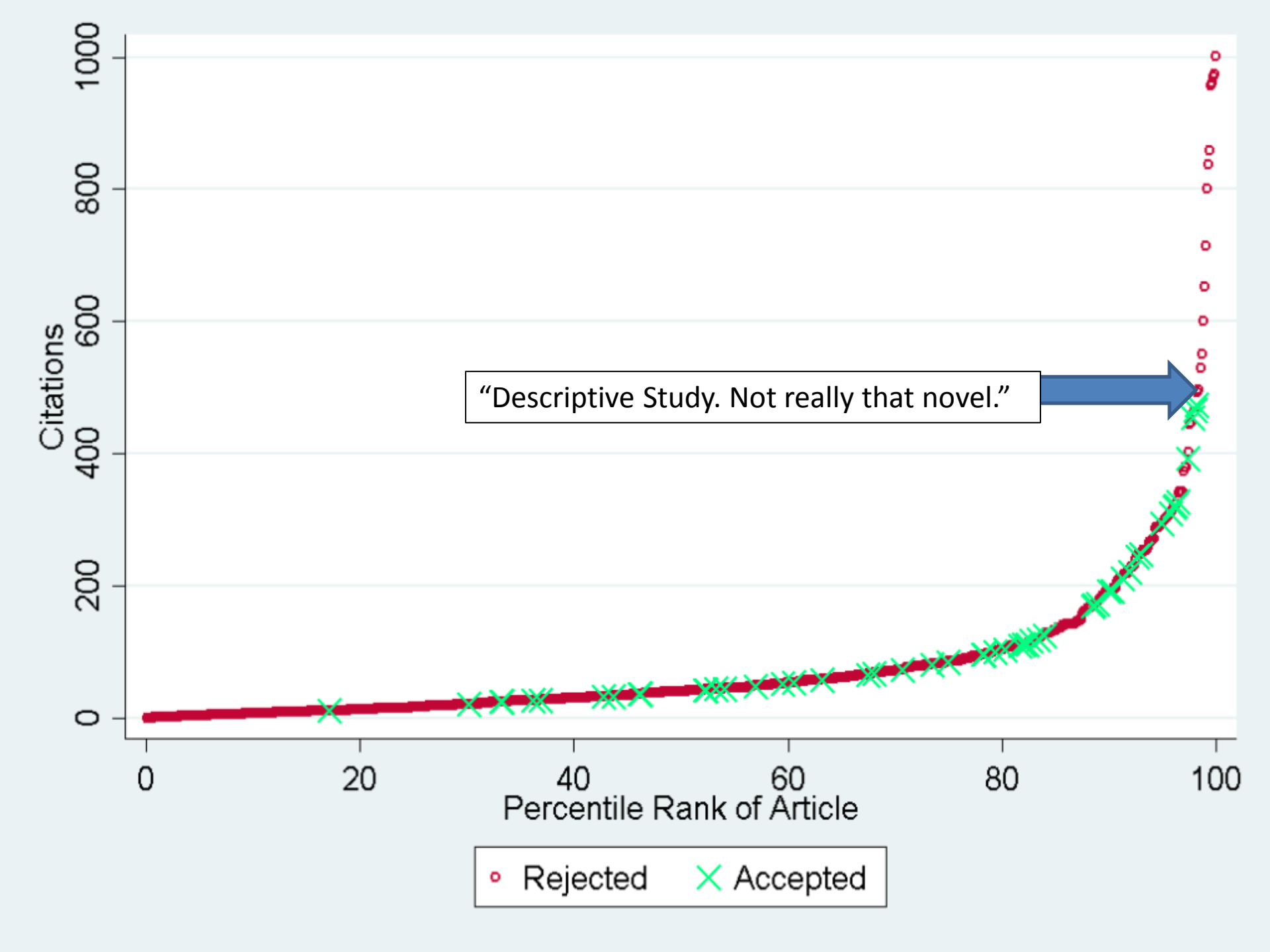

Associate Editor: “Prevalence survey…finds there is a lot of [redacted] about. We don't want this, do we?”
Editor: “No.”
Citations
0
200
400
600
800
1000
0
20
40
60
80
100
Percentile Rank of Article
Rejected
Accepted


Citations
0
200
400
600
800
1000
0
20
40
60
80
100
Percentile Rank of Article
Rejected
Accepted
"I don't think they provide enough evidence for [redacted]. May be appropriate for [specialist, lower-ranked journal]."


Citations
0
200
400
600
800
1000
0
20
40
60
80
100
Percentile Rank of Article
"Not novel enough for us."
Rejected
Accepted


Citations
0
200
400
600
800
1000
0
20
40
60
80
100
Percentile Rank of Article
Rejected
Accepted
"Appears to have been a very incomplete sample. Associations found are wholly unoriginal and unsurprising."
"I'm unpersuaded by their arguments. It says nothing more interesting than there's a lot of [redacted] about."


Citations
0
200
400
600
800
1000
0
20
40
60
80
100
Percentile Rank of Article
"Small study. Non-clinical end point."
Rejected
Accepted


Citations
0
200
400
600
800
1000
0
20
40
60
80
100
Percentile Rank of Article
"Descriptive Study. Not really that novel."
Rejected
Accepted

Most Common Justifications for Rejection in Top-15 Articles	
Lacking Novelty	7
Methodological Problems	4
Magnitude of Results Too Small	4
No reason given	3
Insufficient Data/Evidence	2
Speculative Results/Questionable Validity	2

Justifying Rejections

- Importance of perceived novelty, theoretical impact.

Justifying Rejections

- Importance of perceived novelty, theoretical impact.
- Akerlof, Granovetter, Yalow, J.K. Rowling…are rejections of outstanding work more the rule than the exception?

Justifying Rejections

- Importance of perceived novelty, theoretical impact.
- Akerlof, Granovetter, Yalow, J.K. Rowling…are rejections of outstanding work more the rule than the exception?
- Or, is “exceptional” work more ordinary than often assumed?

What are publishing “mistakes”?

- Ionnadis (2005): Publication system tends to validate false positives, a tendency that highly selective journals may be particularly prone to (Brembs et al., 2013).

What are publishing “mistakes”?

- Ionnadis (2005): Publication system tends to validate false positives, a tendency that highly selective journals may be particularly prone to (Brembs et al., 2013).
- Heightened scrutiny may be needed to counterbalance incentives favoring publication of results that `no one has seen before.’

What are publishing “mistakes”?

- Ionnadis (2005): Publication system tends to validate false positives, a tendency that highly selective journals may be particularly prone to (Brembs et al., 2013).
- Heightened scrutiny may be needed to counterbalance incentives favoring publication of results that `no one has seen before.’
- Low inter-rater reliability: Should editors tap disparate reviewers?

Two Recent Examples of Mistakes

Experimental evidence of massive-scale emotional contagion through social networks

Adam D. I. Kramer[a,1], Jamie E. Guillory[b,2], and Jeffrey T. Hancock[b,c]

[a]Core Data Science Team, Facebook, Inc., Menlo Park, CA 94025; and Departments of [b]Communication and [c]Information Science, Cornell University, Ithaca, NY 14853

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved March 25, 2014 (received for review October 23, 2013)

Emotional states can be transferred to others via emotional contagion, leading people to experience the same emotions without their awareness. Emotional contagion is well established in laboratory experiments, with people transferring positive and negative emotions to others. Data from a large real-world social network, collected over a 20-y period suggests that longer-lasting moods (e.g., depression, happiness) can be transferred through networks [Fowler JH, Christakis NA (2008) *BMJ* 337:a2338], although the results are controversial. In an experiment with people who use Facebook, we test whether emotional contagion occurs outside of in-person interaction between individuals by reducing the amount of emotional content in the News Feed. When positive expressions were reduced, people produced fewer positive posts and more negative posts; when negative expressions were reduced, the opposite pattern occurred. These results indicate that emotions expressed by others on Facebook influence our own emotions, constituting experimental evidence for massive-scale contagion via social networks. This work also suggests that, in contrast to prevailing assumptions, in-person interaction and nonverbal cues are not strictly necessary for emotional contagion, and that the observation of others' positive experiences constitutes a positive experience for people.

computer-mediated communication | social media | big data

Emotional states can be transferred to others via emotional contagion, leading them to experience the same emotions as demonstrated that (*i*) emotional contagion occurs via text-based computer-mediated communication (7); (*ii*) contagion of psychological and physiological qualities has been suggested based on correlational data for social networks generally (7, 8); and (*iii*) people's emotional expressions on Facebook predict friends' emotional expressions, even days later (7) (although some shared experiences may in fact last several days). To date, however, there is no experimental evidence that emotions or moods are contagious in the absence of direct interaction between experiencer and target.

On Facebook, people frequently express emotions, which are later seen by their friends via Facebook's "News Feed" product (8). Because people's friends frequently produce much more content than one person can view, the News Feed filters posts, stories, and activities undertaken by friends. News Feed is the primary manner by which people see content that friends share. Which content is shown or omitted in the News Feed is determined via a ranking algorithm that Facebook continually develops and tests in the interest of showing viewers the content they will find most relevant and engaging. One such test is reported in this study: A test of whether posts with emotional content are more engaging.

The experiment manipulated the extent to which people (N = 689,003) were exposed to emotional expressions in their News Feed. This tested whether exposure to emotions led people to change their own posting behaviors, in particular whether exposure to emotional content led people to post content that was consistent with the exposure—thereby testing whether exposure

Female hurricanes are deadlier than male hurricanes

Kiju Jung[a,1], Sharon Shavitt[a,b,1], Madhu Viswanathan[a,c], and Joseph M. Hilbe[d]

[a]Department of Business Administration and [b]Department of Psychology, Institute of Communications Research, and Survey Research Laboratory, and [c]Women and Gender in Global Perspectives, University of Illinois at Urbana–Champaign, Champaign, IL 61820; and [d]Department of Statistics, T. Denny Sanford School of Social and Family Dynamics, Arizona State University, Tempe, AZ 85287-3701

Edited* by Susan T. Fiske, Princeton University, Princeton, NJ, and approved May 14, 2014 (received for review February 13, 2014)

Do people judge hurricane risks in the context of gender-based expectations? We use more than six decades of death rates from US hurricanes to show that feminine-named hurricanes cause significantly more deaths than do masculine-named hurricanes. Laboratory experiments indicate that this is because hurricane names lead to gender-based expectations about severity and this, in turn, guides respondents' preparedness to take protective action. This finding indicates an unfortunate and unintended consequence of the gendered naming of hurricanes, with important implications for policymakers, media practitioners, and the general public concerning hurricane communication and preparedness.

gender stereotypes | implicit bias | risk perception | natural hazard communication | bounded rationality

Estimates suggest that hurricanes kill more than 200 people in the United States annually, and severe hurricanes can cause fatalities in the thousands (1). As the global climate changes, the frequency and severity of such storms is expected to increase (2). However, motivating hurricane preparedness remains a major challenge for local and state authorities (3). Although natural hazards such as hurricanes represent both physical and social phenomena (4, 5), meteorologists and geoscientists point out violence and destruction (23, 24). We extend these findings to hypothesize that the anticipated severity of a hurricane with a masculine name (Victor) will be greater than that of a hurricane with a feminine name (Victoria). This expectation, in turn, will affect the protective actions that people take. As a result, a hurricane with a feminine vs. masculine name will lead to less protective action and more fatalities.

Archival Study

To test this hypothesis, we used archival data on actual fatalities caused by hurricanes in the United States (1950–2012). Ninety-four Atlantic hurricanes made landfall in the United States during this period (25). Nine independent coders who were blind to the hypothesis rated the masculinity vs. femininity of historical hurricane names on two items (1 = very masculine, 11 = very feminine, and 1 = very man-like, 11 = very woman-like), which were averaged to compute a masculinity-femininity index (MFI). A series of negative binomial regression analyses (26, 27) were performed to investigate effects of perceived masculinity-femininity of hurricane names (MFI), minimum pressure, normalized damage (NDAM) (28), and the interactions among them on the number of deaths caused by the hurricanes (see *Materials and Methods* for complete descriptions of models tested. Table S1 for

Backlash to Publishing Mistakes

The Washington Post

Search

Capital Weather Gang

Disbelief, shock and skepticism: Hurricane gender study faces blowback

By Jason Samenow June 3, 2014 Follow @capitalweather

The study "Female hurricanes are deadlier than male hurricanes" has generated a stormy response from academics and the public alike. Its extraordinary conclusion that female-named hurricanes have been about twice as deadly compared to male hurricanes due to gender bias, to many, seems far-fetched.

Let's examine the three most common reactions to its conclusions that people (implicitly) think female storms are less risky....

1) Disbelief. This study isn't for real. It's from The Onion, right?

Most Read Local

1 'Everybody loved her': Woman's life ended with gunshots in a parking lot

2 College students remain deeply divided over what consent actually means

3 Principal: 'I cannot be part of reforms that eat away at the moral fabric of our schools'

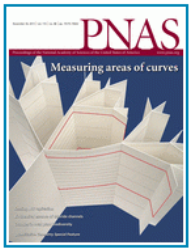
4 Dangerous flood threat

Retraction Watch

Tracking retractions as

Rapid mood swing: PNAS issues Expression of Concern for controversial Facebook study

with 58 comments



The *Proceedings of the National Academy of Sciences* (PNAS) is subjecting a much-criticized study involving Facebook that it published just two weeks ago to an Expression of Concern.

From the abstract of the original study:

> In an experiment with people who use Facebook, we test whether emotional contagion occurs outside of in-person interaction between individuals by reducing the amount of emotional content in the News Feed. When positive expressions were reduced, people produced fewer positive posts and more negative posts; when negative expressions were reduced, the opposite pattern occurred. These results indicate that emotions expressed by others on Facebook influence our own emotions, constituting experimental evidence for massive-scale contagion via social networks. This work also suggests that, in contrast to prevailing assumptions, in-person interaction and non-verbal cues are not strictly necessary for emotional contagion, and that the observation of others' positive experiences constitutes a positive experience for people.

In other words, the researchers manipulated hundreds of thousands of Facebook feeds to see what effect it would have.

Critics — and there were many online — said the study violated ethical norms because it did not alert

Nicholas Christakis
@NAChristakis

Follow

I feel bad 4 Susan Fiske who was PNAS editor of both hurricane study pnas.org/content/early/… & #FacebookExperiment pnas.org/content/111/24… 1/2

RETWEETS 2 FAVORITES 5

1:20 PM - 4 Jul 2014

Nicholas Christakis
@NAChristakis

Following

…because I think that #PNAS is publishing some of the best social science these days, and cool, innovative stuff has risks too. 2/2

RETWEETS 2 FAVORITES 3

1:21 PM - 4 Jul 2014

Reply to @NAChristakis

Marcel Salathe @marcelsalathe · Jul 4
@NAChristakis It's great that she had the courage to take these risks. Hopefully the NAS members will have her back

1

Nicholas Christakis @NAChristakis · Jul 4
.@marcelsalathe yes, exactly! @PNASNews has been publishing creative & original social science (including these studies) which has risks too

1

LaCour Scandal

Article Views
› Abstract
› Full Text
› Full Text (PDF)
› Figures Only
› Supplementary Materials
Article Tools
› Save to My Folders
› Download Citation
› Alert Me When Article is Cited
› Post to CiteULike
› Article Usage Statistics
› E-mail This Page
› Rights & Permissions
› Commercial Reprints and E-Prints

This article has been retracted

An Expression of Concern has been published for this article

‹ Prev | Table of Contents | Next ›
Read Full Text to Comment (0)

Science 12 December 2014:
Vol. 346 no. 6215 pp. 1366-1369
DOI: 10.1126/science.1256151

REPORT

When contact changes minds: An experiment on transmission of support for gay equality

Michael J. LaCour[1], Donald P. Green[2]

+ Author Affiliations

ABSTRACT | EDITOR'S SUMMARY

Can a single conversation change minds on divisive social issues, such as same-sex marriage? A randomized placebo-controlled trial assessed whether gay (n = 22) or straight (n = 19) messengers were effective at encouraging voters (n = 972) to support same-sex marriage and whether attitude change persisted and spread to others in voters' social networks. The results, measured by an unrelated panel survey, show that both gay and straight canvassers produced large effects initially, but only gay canvassers' effects persisted in 3-week, 6-week, and 9-month follow-ups. We also find strong evidence of within-household transmission of opinion change, but only in the wake of conversations with gay canvassers. Contact with gay

National hiring experiments reveal 2:1 faculty preference for women on STEM tenure track

Wendy M. Williams[1] and Stephen J. Ceci

Department of Human Development, Cornell University, Ithaca, NY 14853

Edited* by Richard E. Nisbett, University of Michigan, Ann Arbor, MI, and approved March 5, 2015 (received for review September 30, 2014)

National randomized experiments and validation studies were conducted on 873 tenure-track faculty (439 male, 434 female) from biology, engineering, economics, and psychology at 371 universities/colleges from 50 US states and the District of Columbia. In the main experiment, 363 faculty members evaluated narrative summaries describing hypothetical female and male applicants for tenure-track assistant professorships who shared the same lifestyle (e.g., single without children, married with children). Applicants' profiles were systematically varied to disguise identically rated scholarship; profiles were counterbalanced by gender across faculty to enable between-faculty comparisons of hiring preferences for identically qualified women versus men. Results revealed a 2:1 preference for women by faculty of both genders across both math-intensive and non–math-intensive fields, with the single exception of male economists, who showed no gender preference. Results were replicated using weighted analyses to control for national sample characteristics. In follow-up experiments, 144 faculty evaluated competing applicants with differing lifestyles (e.g., divorced mother vs. married father), and 204 faculty compared same-gender candidates with children, but differing in whether they took 1-y-parental leaves in graduate school. Women preferred divorced mothers to married fathers; men preferred mothers who took leaves to mothers who did not. In two validation studies, 35 engineering faculty provided rankings using full curricula vitae instead of narratives, and 127 faculty rated one applicant rather than choosing from a mixed-gender group; the same preference for women was shown by faculty of both genders. These results suggest it is a propitious time for women launching careers in academic science. Messages to the contrary may discourage women from applying for STEM (science, technology, engineering, mathematics) tenure-track assistant professorships.

gender bias | hiring bias | underrepresentation of women | faculty hiring | women in science

controlling for demographics, degree characteristics, and field (15). [This winnowing of women in the STEM (science, technology, engineering, mathematics) tenure-track pipeline is a result of women Ph.D.s being far less likely than men to apply for tenure-track jobs, rather than to women applying but being rejected at higher rates than men (14).] Against this bleak backdrop, it is perhaps no surprise that talented young women opt out of the STEM tenure track either by not applying for assistant professorships at the same rate as men or, in some fields, by not majoring in them in college in the first place (14).

The point at which scientists choose to apply for tenure-track assistant professorships is a key juncture in understanding the problem of women's underrepresentation. Once hired, women prosper in the STEM professoriate (14, 16–18): They are remunerated, persist, and are promoted at rates roughly comparable to men's (14) after controlling for observable characteristics, including academic productivity. However, to be hired and eventually tenured, women must first apply. Unfortunately, despite their success once hired, women apply for tenure-track positions in far smaller percentages than their male graduate student counterparts (14, 16, 18). Why might this be?

One reason may be omnipresent discouraging messages about sexism in hiring, but does current evidence support such messages? Despite this question's centrality to any informed discussion about women's underrepresentation in academic science, only one experimental study (7) contrasted faculty ratings of the relative "hirability" of hypothetical identically qualified women and men. Results showed that both female and male psychology faculty members downgraded a hypothetical woman's academic record compared with an identical man's. However, this study

Significance

The underrepresentation of women in academic science is typically attributed, both in scientific literature and in the me-

- Are polarizing or controversial articles mistakes or virtues for scientific journals and communities?

Decision-Making Under Scientific Uncertainty

- Desk-rejections common on right-tail of quality.

Decision-Making Under Scientific Uncertainty

- Desk-rejections common on right-tail of quality.
- *"The very nature of "speed reviewing" means people tap into previously formed cognitive categories and cannot deal with any deviations or novel ideas. It seems doomed toward very slow incrementalism (at best)."*

Decision-Making Under Scientific Uncertainty

- Desk-rejections common on right-tail of quality.
- *"The very nature of "speed reviewing" means people tap into previously formed cognitive categories and cannot deal with any deviations or novel ideas. It seems doomed toward very slow incrementalism (at best)."*
- Despite prizing novelty, peer review is often conservative!

Decision-Making Under Scientific Uncertainty

- McAfee (2014): “Journals that do not occasionally reject classic articles will publish manuscripts of lower mean quality.”

Decision-Making Under Scientific Uncertainty

- McAfee (2014): “Journals that do not occasionally reject classic articles will publish manuscripts of lower mean quality.”
- Schrodt (2010): “With a few exceptions, peer review keeps junk out of the journals. But does so at the cost of biasing the system to publishing safe, predictable, incremental, lowest common-denominator articles.”

The Value of Citation Analysis

- Are citations always an indicator of quality?

The Value of Citation Analysis

- Are citations always an indicator of quality?
- Contagion often more driven by social forces than quality.

The Value of Citation Analysis

- Are citations always an indicator of quality?
- Contagion often more driven by social forces than quality.
- “Low Church” niches outside of elite journals?

The Value of Citation Analysis

- Are citations always an indicator of quality?
- Contagion often more driven by social forces than quality.
- “Low Church” niches outside of elite journals?
- Still, citations matter: Impact factors, status, journal revenues linked to receiving cites.

CONCLUDING THOUGHTS

- Peer review adds value, but…

CONCLUDING THOUGHTS

- Peer review adds value, but…
  - Tendency towards conservativism in gestation, if not also gatekeeping.

CONCLUDING THOUGHTS

- Peer review adds value, but…
  - Tendency towards conservativism in gestation, if not also gatekeeping.
    - Is this necessarily a bad thing?

CONCLUDING THOUGHTS

- Peer review adds value, but…
  - Tendency towards conservativism in gestation, if not also gatekeeping.
    - Is this necessarily a bad thing?
  - Many successful articles are rejected.

CONCLUDING THOUGHTS

- Peer review adds value, but…
  - Tendency towards conservativism in gestation, if not also gatekeeping.
    - Is this necessarily a bad thing?
  - Many successful articles are rejected.
    - Is this due to value-maximizing tradeoffs?

CONCLUDING THOUGHTS

- Peer review adds value, but…
  - Tendency towards conservativism in gestation, if not also gatekeeping.
    - Is this necessarily a bad thing?
  - Many successful articles are rejected.
    - Is this due to value-maximizing tradeoffs?
  - Possible problems working at the right tail of quality.

FURTHER ISSUES

- What are the tradeoffs between errors of omission and commission?

FURTHER ISSUES

- What are the tradeoffs between errors of omission and commission?
- Polarizing versus less controversial articles.

FURTHER ISSUES

- What are the tradeoffs between errors of omission and commission?
- Polarizing versus less controversial articles.
  - Are publishing the Hurricane *PNAS* piece, or rejecting "The Market for Lemons" necessary evils?

FURTHER ISSUES

- What are the tradeoffs between errors of omission and commission?
- Polarizing versus less controversial articles.
  - Are publishing the Hurricane *PNAS* piece, or rejecting "The Market for Lemons" necessary evils?
- Do low acceptance rates squeeze out innovation? (Alberts et al., 2014)

FURTHER ISSUES

- What are the tradeoffs between errors of omission and commission?
- Polarizing versus less controversial articles.
  - Are publishing the Hurricane *PNAS* piece, or rejecting "The Market for Lemons" necessary evils?
- Do low acceptance rates squeeze out innovation? (Alberts et al., 2014)
- Elite journals in a schizophrenic position.

Good peer review will limit mistakes and improve published works. However, given the uncertainty of science, optimizing decisions must be made regarding publishing errors of omission and commission. These strategic/evaluative cultures determine the science that is published.

Thank you! ☺

Questions, Comments:
ksiler@gmail.com

Twitter: @KyleSiler