

과편협 출판윤리워크숍 2016년 6월 21일

편집인이 바라보는 재현성 위기

조은희

조선대학교 생물교육과

과학학술지편집인협의회 출판윤리위원

차례

- 연구의 재현성에 대한 이해
- '재현성 위기'의 원인과 극복을 위한 전략
- 연구자, 연구지원기관, 학술단체의 대응
- 재현성 위기 극복을 위한 학술지의 역할

연구 결과의 재현성에 대한 문제 제기

- 과학 연구의 객관성을 담보하는 두 축:
투명성(transparency) & 재현성(reproducibility)
- 그러나 요즘의 과학 연구 환경
'새로운 발견' >> 연구 결과의 재현성
→ '재현성 위기' 대두
- 연구자들이 통계적 방법론을 제대로 이해하고 적절하게
적용하지 못하고 있다는 문제 제기에서 시작 (황승식 교
수님의 주제 발표 참조)
 - Ioannidis, John PA. "Why most published research findings
are false." *PLoS Med* 2.8 (2005): e124.
 - Button, Katherine S., et al. "Power failure: why small sample
size undermines the reliability of neuroscience." *Nature
Reviews Neuroscience* 14.5 (2013): 365-376.

다양한 분야에서

- 심리학 분야에서 가장 먼저 문제 제기
 - 통계적인 분석을 많이 활용하는 타 분야에서도
- 종양 생물학 재현성 위기의 대표적인 원인
 - Contaminated cell lines in cancer science
 - A breast-cancer cell line used in more than 1,000 published
studies turned out to have been a melanoma cell line.
 - 그 밖의 재료나 시약의 문제
 - 재현 실험의 방법이나 기술적 문제 등
- 신경질환 연구 등의 분야에서는
 - 동물 실험의 표본 크기
- That problem is likely to be worse in policy-relevant fields
such as nutrition, education, epidemiology and economics,
in which the science is often uncertain and the societal
stakes can be high
 - <http://science.sciencemag.org/content/348/6242/1411.full>

Reproducibility Crisis in Cancer Biology

Repeat failures

6 of 53

Cancer papers that Amgen could reproduce

14 of 67

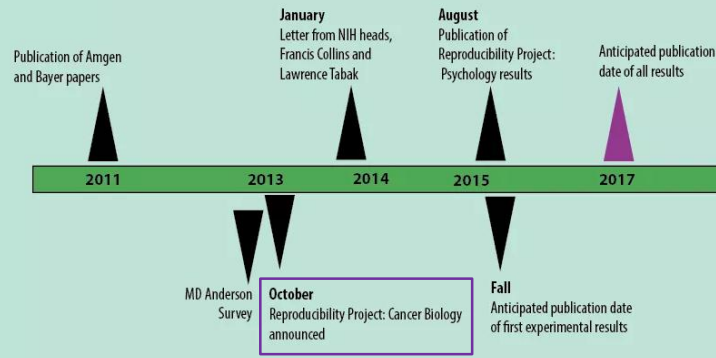
Biomedical papers that Bayer completely reproduced

55%

MD Anderson researchers who could not reproduce a published study

DATA: C. G. BEGLEY AND L. M. ELLIS, NATURE 483, 759 (20 MAR 2012); F. PRINZ ET AL., NATURE REVIEWS DRUG DISCOVERY 10, 712 (SEPTEMBER 2011); A. MOBLEY ET AL., PLOS ONE 8, 5 (15 MAY 2013)

Reproducibility Record: A timeline of events in the story of reproducibility

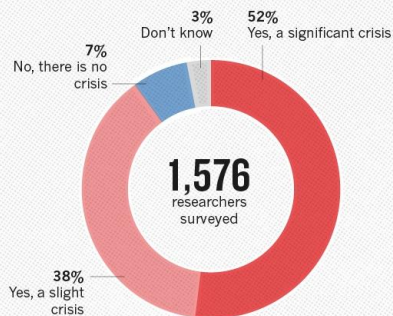


<http://sitn.hms.harvard.edu/flash/2015/reproduce-or-bust-bringing-reproducibility-back-to-center-stage/>

<http://science.sciencemag.org/content/348/6242/1411.full>

Nature 독자의견

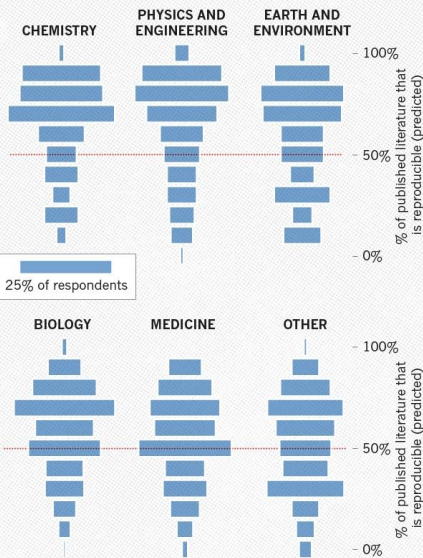
IS THERE A REPRODUCIBILITY CRISIS?



Baker (2016) Nature 533: 453-454

HOW MUCH PUBLISHED WORK IN YOUR FIELD IS REPRODUCIBLE?

Physicists and chemists were most confident in the literature.



Number of respondents from each discipline:
Biology 703, Chemistry 106, Earth and environmental 95,
Medicine 203, Physics and engineering 236, Other 233

©nature

재현성 위기에 대한 이해

- 연구부정행위와 무관
- 분야에 따라 다양한 원인
 - 객관성을 담보하지 못하는 방법론
 - p값에 의존, 충분하지 않은 표본 크기 ...
 - 표준화되지 않은 연구 방법
 - 시료나 재료의 차이
 - 연구자의 편향, 왜곡, 실수
 - 이해충돌에 따른 편향
 - 출판 편향

Does a failure to replicate mean that the original result was a false positive?

- No.
- the original being a false positive
- the replication being a false negative due to
 - insufficient power,
 - error in analysis,
 - differences in statistical methods,
 - error in the design and implementation of the study procedures -
 - reagent variability/lack of validation, unintentional selective reporting, lack of appropriate controls, lack of equipment calibration, or unrecognized experimental variables
- The difference may provide new insights
 - the original effect may be real but overestimated by the original study
 - the conditions necessary to obtain the particular result are not yet understood.

Errington, Timothy M., et al. "An open investigation of the reproducibility of cancer biology research." *Elife* 3 (2014): e04333.

Does a successful replication mean that the original interpretation is correct?

- No.
- can verify that a result can be obtained
- establishes some generalizability by showing that it can be obtained in different circumstances.
- Direct replication does not confirm the interpretation of the result.
 - For example, if an original design has an unidentified confounding influence, then the direct replication is also likely to be influenced by that confound.
- **Conceptual replication** will further clarify original findings.

Errington, Timothy M., et al. "An open investigation of the reproducibility of cancer biology research." *Elife* 3 (2014): e04333.

주의: 재연되지 않는 실험의 또 다른 과학적 의미

- There is room for improvement in how science is done and reported, but something can often be learned from irreproducible experiments.
- It is important to recognize that researchers cannot control for an unknown variable.
- There is an example of a useful resource that enhances our understanding of underlying biological phenomena and results from experiments that might otherwise be branded as irreproducible.
 - a web tool for identifying unwanted 'passenger mutations' that could confound analyses of transgenic mice (T.Vanden Berghe et al. *Immunity* 43, 200–209; 2015).

Nature 2016 532:177

재현성 위기의 원인



<http://www.acmedsci.ac.uk/viewFile/56314e40aac61.pdf>



Data dredging

Also known as p-hacking, this involves repeatedly searching a dataset or trying alternative analyses until a 'significant' result is found.



Omitting null results

When scientists or journals decide not to publish studies unless results are statistically significant.



Underpowered study

Statistical power is the ability of an analysis to detect an effect, if the effect exists – an underpowered study is too small to reliably indicate whether or not an effect exists.



Errors

Technical errors may exist within a study, such as misidentified reagents or computational errors.



Underspecified methods

A study may be very robust, but its methods not shared with other scientists in enough detail, so others cannot precisely replicate it.



Weak experimental design

A study may have one or more methodological flaws that mean it is unlikely to produce reliable or valid results.

Issues

재현성 위기 극복을 위한 전략



<http://www.acmedsci.ac.uk/viewFile/56314e40aac61.pdf>

Open data

Openly sharing results and the underlying data with other scientists.



Pre-registration

Publicly registering the protocol before a study is conducted.



Collaboration

Working with other research groups, both formally and informally.



Automation

Finding technological ways of standardising practices, thereby reducing the opportunity for human error.



Open methods

Publicly publishing the detail of a study protocol.



Post-publication review

Continuing discussion of a study in a public forum after it has been published (most are reviewed before publication).



Reporting guidelines

Guidelines and checklists that help researchers meet certain criteria when publishing studies.



Dr. Ioannidis의 제언

- Large-scale collaborative research
- **Adoption of replication culture**
- Registration (of studies, protocols, analysis codes, datasets, raw data, and results)
- Sharing (of data, protocols, materials, software, and other tools)
- Reproducibility practices
- **Containment of conflicted sponsors and authors**
- More appropriate statistical methods
- Standardization of definitions and analyses
- **More stringent thresholds for claiming discoveries or “successes”**
- Improvement of study design standards
- **Improvements in peer review, reporting, and dissemination of research**
- **Better training of scientific workforce in methods and statistical literacy**

Ioannidis JPA (2014) How to Make More Published Research True. PLoS Med 11(10): e1001747.
doi:10.1371/journal.pmed.1001747
<http://journals.plos.org/plosmedicine/article?id=info:doi/10.1371/journal.pmed.1001747>

연구자의 편향성 극복

HOW SCIENTISTS FOOL THEMSELVES — AND HOW THEY CAN STOP

Humans are remarkably good at self-deception. But growing concern about reproducibility is driving many researchers to seek ways to fight their own worst instincts.

COGNITIVE FALLACIES IN RESEARCH



HYPOTHESIS MYOPIA

Collecting evidence to support a hypothesis, not looking for evidence against it, and ignoring other explanations.



TEXAS SHARPSHOOTER

Seizing on random patterns in the data and mistaking them for interesting findings.



ASYMMETRIC ATTENTION

Rigorously checking unexpected results, but giving expected ones a free pass.



JUST-SO STORYTELLING

Finding stories after the fact to rationalize whatever the results turn out to be.

DEBIASING TECHNIQUES



DEVIL'S ADVOCACY

Explicitly consider alternative hypotheses — then test them out head-to-head.



PRE-COMMITMENT

Publicly declare a data collection and analysis plan before starting the study.



TEAM OF RIVALS

Invite your academic adversaries to collaborate with you on a study.



BLIND DATA ANALYSIS

Analyse data that look real but are not exactly what you collected — and then lift the blind.

재현 연구 시도 The Reproducibility Project: Cancer Biology

- Open-source replication project (<https://osf.io/e81xl>)
- 2013년 시작
- 2010~2012년에 발표된 전 임상 단계의 암 연구 논문 중 가장 영향력이 높은 50편을 선택
- 각 논문에서 연구 결과 일부를 재연하는 중
- 재현 실험의 종류, 방법, 결과는 모두 공개



Errington, Timothy M., et al. "An open investigation of the reproducibility of cancer biology research." *Elife* 3 (2014): e04333.

연구 지원기관의 대응 : 미국 국립보건원

- 2016년 5월 25일부터 사용되는 신청서 양식 개정
- 개정의 목적 : 연구의 신뢰도 향상
 - 연구의 전제에 대한 타당성 설명
 - 재연가능하고 편향 없는 결과를 얻을 수 있는 엄격한 실험 설계
 - 생물학적 변수를 충분히 고려하고 있음을 증명해야 (연구에 사용된 실험동물의 성별 명기하는 등)
 - 세포주나 항체 등의 실험 재료가 명기된 대로 정확하게 사용되었다는 확인을 할 수 있어야

출처: NIH, Implementing Rigor and Transparency in NIH & AHRQ Research Grant Applications, <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-16-011.html> (2016. 5. 25. 최종 확인)

학회의 대응: 지침 마련

- 신뢰도 낮은 시약이나 자료에 의해 소모되지 말자!
- 미국 실험생물학회 연합 : 항체와 동물 실험에 관한 표준안
- 신경과학회 : 전체 집단을 대상으로 하지 않는 경우에는 반드시 무작위 표본
- 미국 심리학회 : 자료 공유를 촉진하는 기반설비 및 정책
- 생물물리학회 : 실험 자료 공유 방안 상세화
- 미국 세포생물학회 : 분석방법을 표준화할 수 있는 분과 수립

Nature Editorial (2016) 529: 256

연구 보상 체계가 달라져야

	Different examples of reward systems		
	Current	Change 1	Change 2
CURRENCIES			
Publication (per unit)	Win 1	No value	No value
Replicated publication (per unit)	Win 1	Win 2	Win 2
Successfully translated publication (per unit)	Win 1	Win 5	Win 5
Refuted publication (per unit)	Win 1	Lose 1	Lose 1
Sharing data, protocols, analysis codes (per unit)	No value	Win 2	Win 2
Contribution to peer-review (per unit)	No value	Win 2	Win 2
Contribution to education/training (per unit)	No value	Win 1	Win 1
Grant funding (per one R01)	Win 5	Win 5	Lose 5
OTHER WEALTH ITEMS			
Assistant professor, title in good university	Win 3	Win 3	No value
Associate professor, title in good university	Win 10	Win 10	No value
Tenured professor, title in good university	Win 20	Win 20	No value
Team leader/director			
Per 1 doctoral student/post-doc	Win 2	Win 2	Lose 2
Administrative power, networking, lobbying	Win up to 200	No value	Lose up to 200

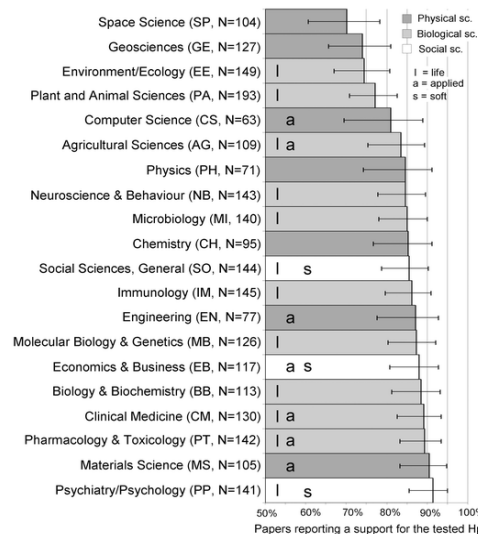
doi:10.1371/journal.pmed.1001747.t002

Ioannidis JPA (2014) How to Make More Published Research True. PLoS Med 11(10): e1001747.
doi:10.1371/journal.pmed.1001747

<http://journals.plos.org/plosmedicine/article?id=info:doi/10.1371/journal.pmed.1001747>

학술지 편집인과 재현성 위기

Positive Results by Discipline.

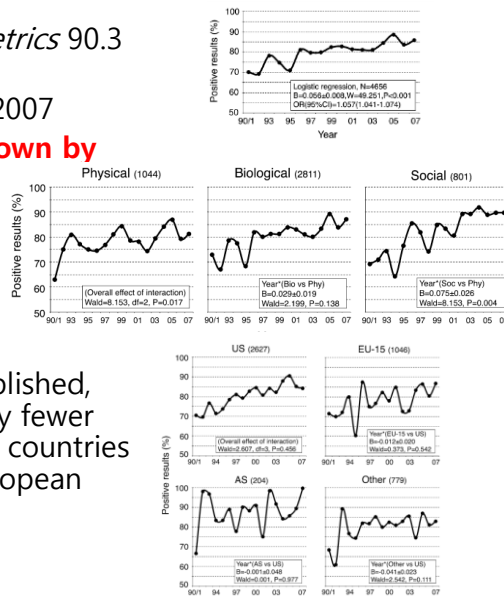


Fanelli D (2010) "Positive" Results Increase Down the Hierarchy of the Sciences. PLoS ONE 5(4): e10068.
doi:10.1371/journal.pone.0010068
<http://journals.plos.org/plosone/article?id=info:doi/10.1371/journal.pone.0010068>

2434 편의 가설 검증 논문 분석

Negative results are disappearing

- Fanelli, Daniele. *Scientometrics* 90.3 (2011): 891-904.
- over 4,600 papers, 1990~2007
- The positive supports grown by over 22%**, with significant differences between disciplines and countries.
- The increase was stronger in the social and some biomedical disciplines.
- The United States had published, over the years, significantly fewer positive results than Asian countries (Japan) but more than European countries.



eLife

- 2012년 창간, open access, 생의학 전문 학술지
- HHMI, Max Planck Society, Wellcome Trust 지원
- 세계 정상급 학자들이 편집위원으로 참여
- 편집위원장: Randy Schekman
 - 버클리대 세포생물학 교수, 2013년 노벨생리의학상 수상
 - CNS는 널리 인용될만한 논문을 주로 게재하여 과학계를 왜곡
 - “(The) pressure to publish in “luxury” journals encouraged researchers to cut corners and pursue trendy fields of science instead of doing more important work.”
 - ✓ Sample, Ian (9 December 2013). [“Nobel winner declares boycott of top science journals”](#). *The guardian.com*.
- 과학 연구의 투명성을 위해
 - 재현 연구 게재 (Reproducibility project: cancer biology)
 - 논문 평가자들은 최종 심사종료 전 심사 내용 토의
 - 논문 게재 이후의 주요 성과를 발표할 기회 제공 “The Research Advance”

Morrison, Sean J. "Time to do something about reproducibility." *eLife* 3 (2014): e03981.

Preclinical Reproducibility and Robustness Channel

- a platform for open and transparent publication of confirmatory and non-confirmatory studies in biomedical research.
- to improve the reproducibility of studies.
- Each publication will undergo a fully transparent post-publication peer review following the *F1000Research* publication model.
- *F1000Research* recommends that authors raise any concerns about previous studies directly with the original authors before publishing their replication attempts. The authors whose studies have been confirmed or not confirmed can provide signed comments on the article, or publish their own full Correspondence article

Channel Advisors



Bruce Alberts
University of California, San Francisco
USA



Alexander Kamb
Amgen Inc.
USA

<http://www.sciencemag.org/news/2016/02/if-you-fail-reproduce-another-scientist-s-results-journal-wants-know>

재현성 위기의 극복

- (연구부정행위에 의한 문제가 아니라) 정상적이라고 수용된 연구 및 분석 방법에 대한 반성
- 관행적으로 사용해 온 방법이나 기준에서 개선하고 강화할 부분을 찾고 발전시켜야 한다는 제안
- 이러한 과정을 거칠 때 과학적 방법론의 객관성이 높아질 것으로 기대
- 학술지 편집인 또한 연구자들의 편향, 이해충돌, 출판 편향 등에 대한 이해를 높이고 이에 적극적으로 대처함으로써 재현성 위기 극복을 통한 연구의 객관성 증진에 제 몫을 해야