'On the causal effect of gender: beyond the (many!) anecdotes a statistician's view on evidence of gender bias'

Royal Belgian Statistical Society, October 19, 2018

Els Goetghebeur Ghent University, Belgium

Florence Nightingale

Born: May 12, 1820, Florence, Italy
Died: August 13, 1910, Mayfair, London, UK

- Through careful observation in army hospitals: war wounds only accounted for one death in six.
- Diseases (typhus, cholera, dysentery…) main reasons for high death-rate
- Traced back to poor sanitation: a causal analysis that lead her to propose reforms of military hospitals

2

The Nightingale Lectures

The life of an applied statistician…

- Military officers and doctors interpreted her comments

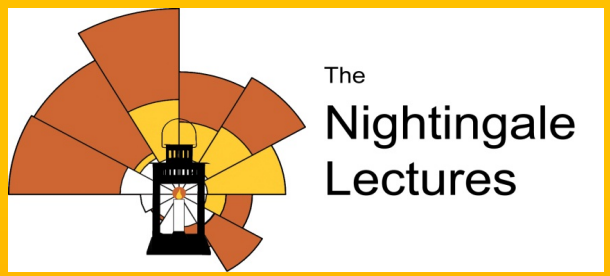
3

The Nightingale Lectures

The life of a statistician…

- Military officers and doctors interpreted her comments as **an attack on their professionalism.**

4



The life of a statistician…

- Military officers and doctors interpreted her comments as an attack on their professionalism .
- Little help from the military until *The Times* reported details...
- after a great deal of publicity, Nightingale was given the task of organizing the barracks of military hospitals
- She dramatically reduced the death-rate of her patients

5

At last, a degree of honour for 900 Cambridge women

A triumphant day in October 1948 saw the then Queen

receive the first woman's honorary degree in the Senate House,

the building where on 4 July more than 900 women who graduated from Girton and Newnham colleges up to 1948 will be honoured

From the 1998 paper

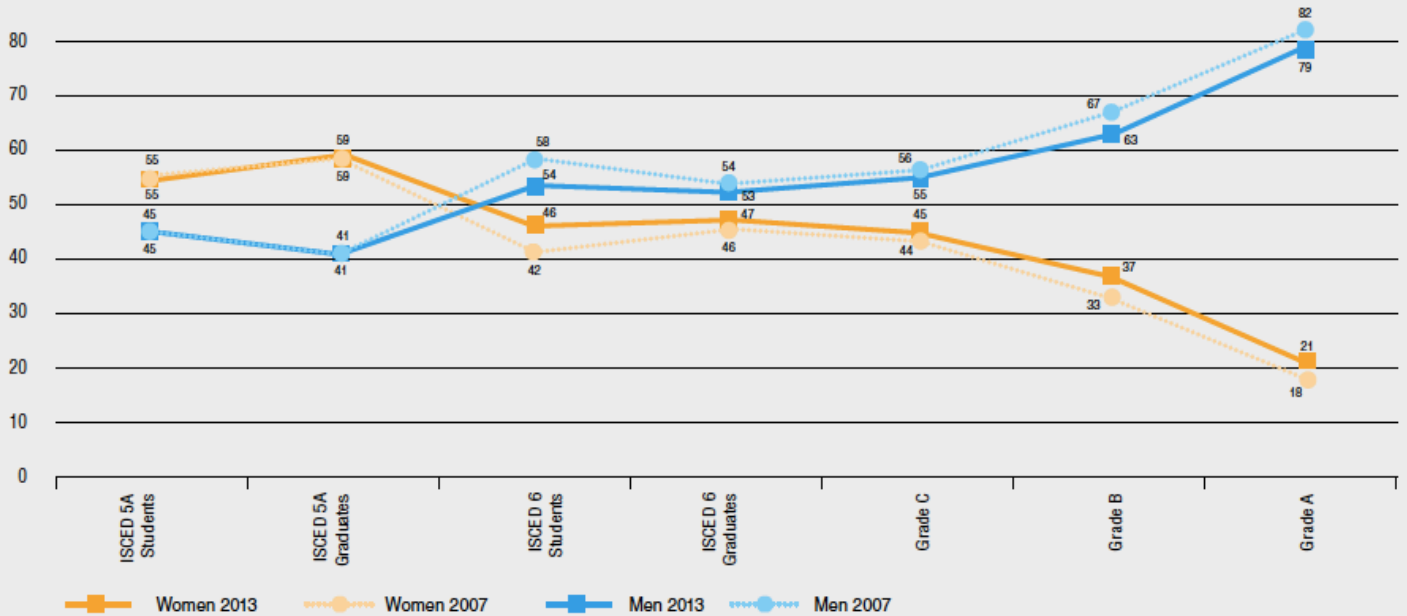
From *The independent*

Women who passed the Cambridge exams but never got the diploma prior to 1948; invited to receive it 50+ years later...

Box 1: Male and female academic career progression

The pattern whereby women in academic careers are significantly less likely to reach the top than men is referred to as the 'leaky pipeline or the 'vanish box'.[6]

Figure 1: The 'leaky pipeline' or 'vanish box' negatively affects women in academic careers across Europe [7] [8]



As Figure 1 shows, 59% of graduates but only 21% of (Grade A) professors[9] are female, whereas 41% of graduates and 79% of professors are male.

The graph also shows that progress between 2007 and 2013 has been minimal.[10] Overall growth in the proportion of women grade A academics is slow at about half a percentage point annually. The number of women employed as heads of universities or institutions entitled to deliver PhD degrees is 15% (EU-28 average, 2014) showing a similar growth rate to that of women grade A academics (*She Figures* 2015, p. 142, table 6.4.).[11]

Student organisations

What gender bias??

Some recent quotes from within the BSS

- 'If you had a good sense of what high quality means in international statistics, you would predict with probability close to 1, that all invited speakers would be male'
- 'The only problem I see is that we are stuck with incompetent women due to positive discrimination'

Harvard Project implicit: take the test

https://implicit.harvard.edu/implicit/education.html

- Compulsory test in several academic institutions
- The Implicit Association Test (IAT) measures attitudes and beliefs that people may be unwilling or unable to report. The IAT may be especially interesting if it shows that you have an implicit attitude that you did not know about.
- For example, you may believe that women and men should be equally associated with science, but your automatic associations could show that you (like many others) associate men with science more than you associate women with science.


PUSHING
THE FRONTIERS
OF INNOVATIVE
RESEARCH
ADVICE PAPER
NO.23 - JANUARY 2018
Implicit bias in academia:
A challenge to the meritocratic principle and to women's careers – And what to do about it
LEAGUE OF EUROPEAN RESEARCH UNIVERSITIES
• University of Amsterdam • Universitat de Barcelona • University of Cambridge • University of Copenhagen • Trinity College Dublin
• University of Edinburgh • University of Freiburg • Université de Genève • Universität Heidelberg • University of Helsinki
• Universiteit Leiden • KU Leuven • Imperial College London • University College London • Lund University • University of Milan
• Ludwig-Maximilians-Universität München • University of Oxford • Sorbonne University • Université Paris-Sud
• University of Strasbourg • Utrecht University • University of Zurich

Where are the problems for women in science - allegedly?

Lots of moments of evaluation by men (and women)
Under stated and implicit rules - often set by men

- Being hired/chosen for the job
- Evaluation of research output and impact
- Evaluation of grant proposals
- Opportunity for networking
- Getting promoted

Where can we find scientific evidence?

Some of the evidence presented in the literature is in nature

- experimental
- observational

Where can we find scientific evidence?

Some of the evidence presented in the literature is in nature

- experimental
- observational

btw: what exactly is the question?

The causal question and potential outcomes

Are we more or less likely to get

- the job
- the promotion
- the impact and citations
- the research grant
- ....

because we are male/female?

This is an active field of research. There are *many* publications presenting evidence.

We select a few here: from what sparked an early interest to more recent work.

Sex Roles, Vol. 41, Nos. 7/8, 1999

The Impact of Gender on the Review of the Curricula Vitae of Job Applicants and Tenure Candidates: A National Empirical Study

Rhea E. Steinpreis,[1] Katie A. Anders, and Dawn Ritzke
University of Wisconsin-Milwaukee

- 283 male and female academic psychologists (US)
- are sent one of 4 CVs, carrying the name of a:
  female/male for a job applicant/tenure candidate
- in total there were in fact just 2 cv's :
  from a single real-life scientist at two stages in her career
- between-groups design:
  every reviewer was sent just one of these cv's
- Names were changed to traditional male/female names

Reviewers were are asked: would you hire this candidate?

Result: both men and women are more likely to vote for male job applicant.

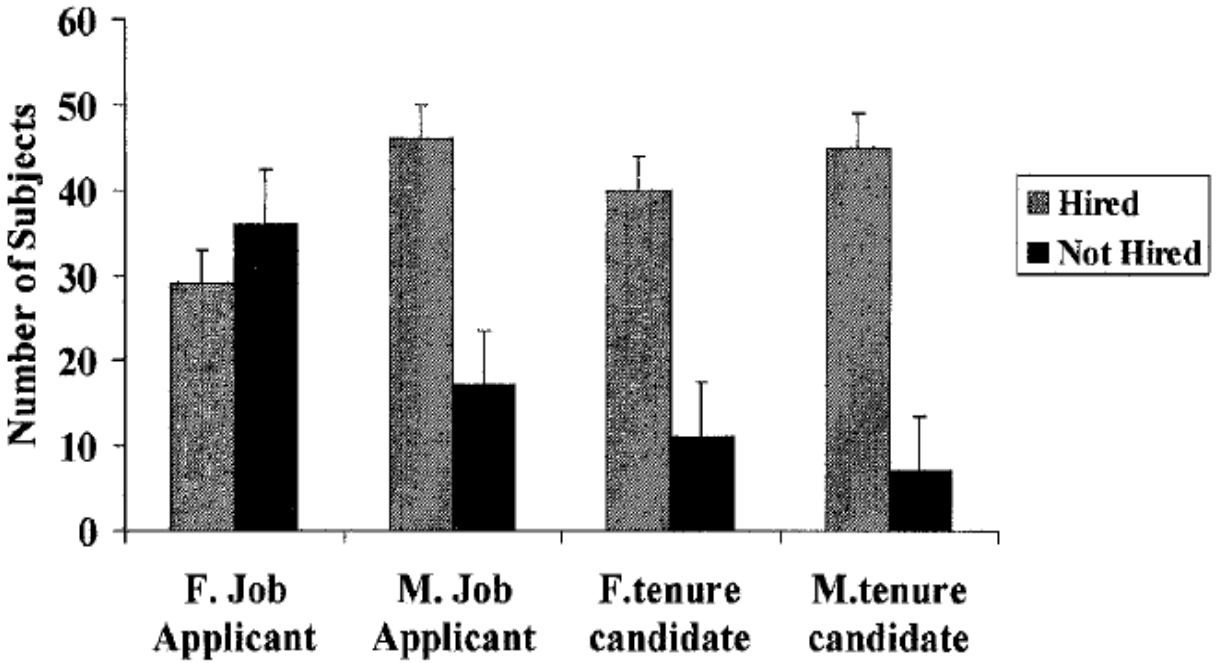


Fig. 1. Hireability of the job applicants and tenure candidates based on the quality of the curriculum vitae the participants were asked to evaluate.

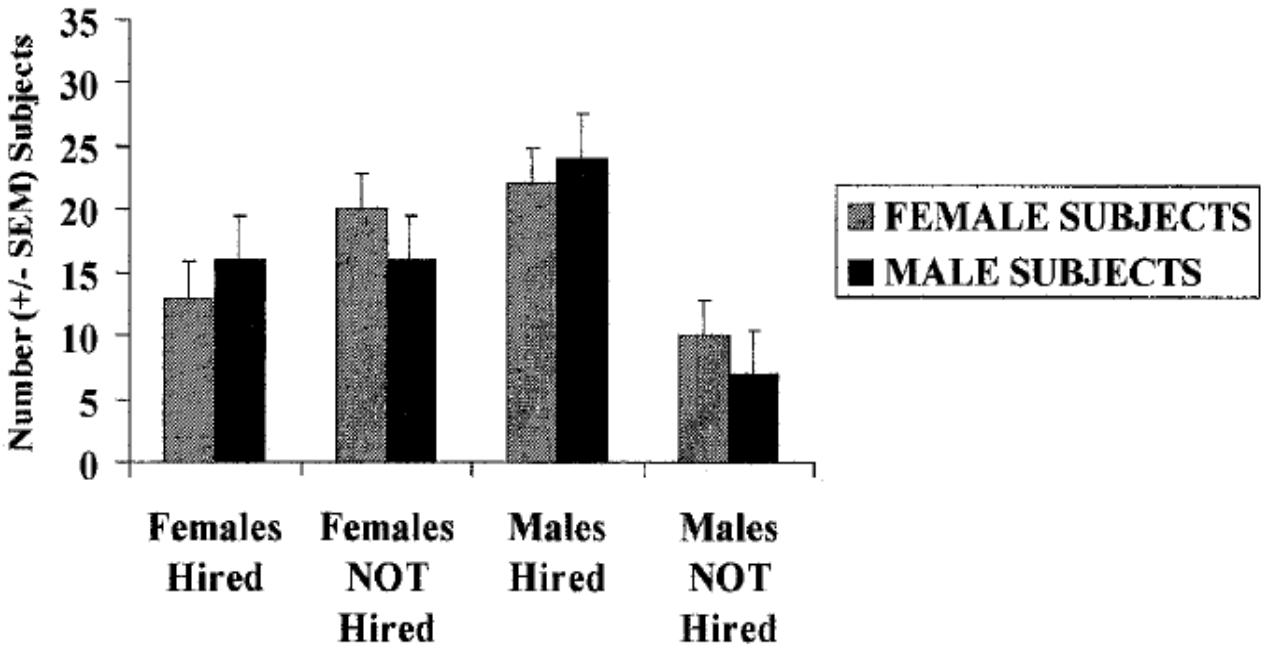


Fig. 5. Hireability of the job applicants as a function of the participant's gender.

Nepotism and sexism in peer review - Nature, May 1997

... in Sweden women hold 44% of biomedical PhDs, mere 25% of postdoctoral positions, 7% of professorial positions

- Our study strongly suggests that peer reviewers cannot judge scientific merit independent of gender
- The peer reviewers over estimated male achievements and/or underestimated female performance,
- ... as shown by multiple regression analyses of the relation between defined parameters of scientific productivity and competence score given by the reviewers of the Swedish research council

Swedish MRC review - 1995

Applicants for a postdoctoral position are rated by 5 reviewers giving a score form 0 to 4 on 3 domains:

- scientific competence
- relevance of research proposal
- quality of the proposed methodology

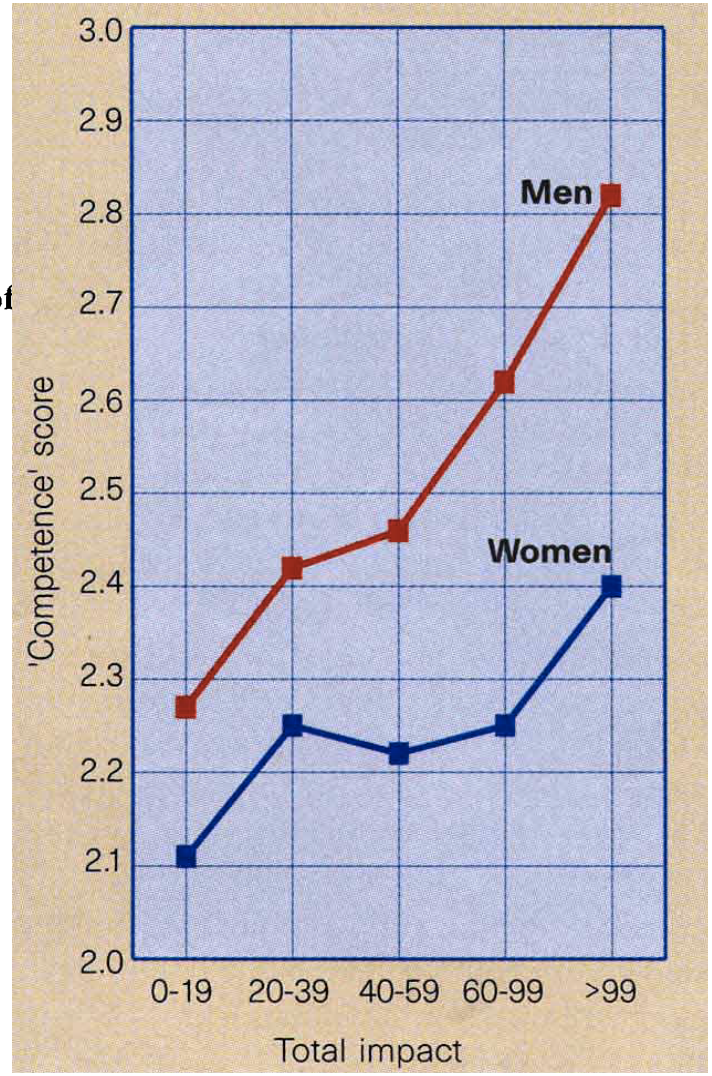
the 3 scores are multiplied for each reviewer (score from 0 to 64), then averaged over the 5 reviewers

Linear regression of the competence scores given by the MRC reviewers on:

- 'objective scores' of impact (based on publications)
- gender
- nationality
- and basic education, scientific field, university affiliation, evaluation committee assigned, postdoctoral experience abroad, letter of recommendation, affiliated with any of the members on the committee

Figure 1 **The mean competence score given to male (red squares) and female (blue squares) applicants by the MRC reviewers as a function of their scientific productivity, measured as total impact. One impact point equals one paper published in a journal with an impact factor of 1. (See text for further explanation.)**

NATURE | VOL 387 | 22 MAY 1997



one impact point equals one paper published in a journal with impact factor 1

Table 1 **Factors that significantly influenced peer reviewers' rating of scientific competence, according**

	Scientific productivity			*Additional points given by the reviewers for the following factors*		
Multiple regression model based on:	r^2	Intercept	Competence points per productivity unit	Male gender	Reviewer affiliation	Recommendation letter
Total impact	0.47	2.09	0.0033 *<0.00005**	0.21 *<0.00005*	0.22 *0.0008*	0.10 *0.04*
First-author impact	0.44	2.13	0.0094 *<0.0001*	0.24 *<0.00005*	0.20 *0.005*	NS
First-author citations	0.41	2.17	0.0054 *0.001*	0.23 *<0.00005*	0.23 *0.001*	NS

* Italicized numbers indicate *P*-values for the variable in question.
† Numbers in parentheses indicate 95% confidence interval.
NS, not statistically significant, *P*-value > 0.05.

Size of the influence of the non-scientific factors in productivity equivalents		
Male gender	Reviewer affiliation	Unit of measure
64 (35–93)†	67 (29–105)	Impact points
25 (14–36)	21 (6–36)	Impact points
42 (23–61)	42 (17–67)	Citations during 1994

Interpreted as

- a female needs to exceed scientific productivity by 64 impact points to make up for gender
  i.e. about 20 papers extra in impact factor 3 journals
- the mean total impact of this cohort of applicants was 40
- afilliation bonus was worth 67 impact points

Science communication 2013

The Matilda Effect in Science Communication: An Experiment on Gender Bias in Publication Quality Perceptions and Collaboration Interest

Science communication 2013

Abstract

An experiment with 243 young communication scholars tested hypotheses derived from role congruity theory regarding impacts of author gender and gender typing of research topics on perceived quality of scientific publications and collaboration interest. Participants rated conference abstracts ostensibly authored by females or males, with author associations rotated. The abstracts fell into research areas perceived as gender-typed or gender-neutral to ascertain impacts from gender typing of topics. Publications from male authors were associated with greater scientific quality, in particular if the topic was male-typed. Collaboration interest was highest for male authors working on male-typed topics. Respondent sex did not influence these patterns.

Assessment of potential bias in research grant peer review in Canada - CMAJ, 2015

ABSTRACT

BACKGROUND: Peer review is used to determine what research is funded and published, yet little is known about its effectiveness, and it is suspected that there may be biases. We investigated the variability of peer review and factors influencing ratings of grant applications.

METHODS: We evaluated all grant applications submitted to the Canadian Institutes of Health Research between 2012 and 2014. The contribution of application, principal applicant and reviewer characteristics to overall application score was assessed after adjusting for the applicant's scientific productivity.

RESULTS: Among 11 624 applications, 66.2% of principal applicants were male and 64.1% were in a basic science domain. We found a significant nonlinear association between scientific productivity and final application score that differed by applicant gender and scientific domain, with higher scores associated with past funding success and *h*-index and lower scores associated with female applicants and those in the applied sciences. Significantly lower application scores were also associated with applicants who were older, evaluated by female reviewers only (v. male reviewers only, –0.05 points, 95% confidence interval [CI] –0.08 to –0.02) or reviewers in scientific domains different from the applicant's (–0.07 points, 95% CI –0.11 to –0.03). Significantly higher application scores were also associated with reviewer agreement in application score (0.23 points, 95% CI 0.20 to 0.26), the existence of reviewer conflicts (0.09 points, 95% CI 0.07 to 0.11), larger budget requests (0.01 points per $100 000, 95% CI 0.007 to 0.02), and resubmissions (0.15 points, 95% CI 0.14 to 0.17). In addition, reviewers with high expertise were more likely than those with less expertise to provide higher scores to applicants with higher past success rates (0.18 points, 95% CI 0.08 to 0.28).

INTERPRETATION: There is evidence of bias in peer review of operating grants that is of sufficient magnitude to change application scores from fundable to nonfundable. This should be addressed by training and policy changes in research funding.

PNAS 2015: National hiring experiments reveal 2:1 faculty preference for women on STEM tenure track

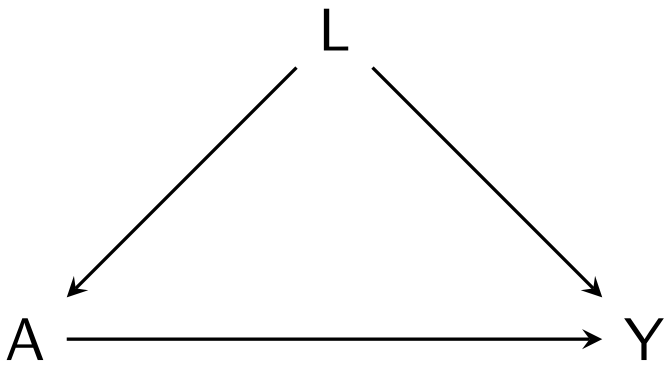
- 5 hiring experiments: faculty evaluated hypothetical female and male applicants, using systematically varied profiles disguising identical scholarship, for assistant professorships in biology, engineering, economics, and psychology.
- men and women faculty members from all four fields preferred female applicants 2:1 over identically qualified males with matching lifestyles (single, married, divorced), except for male economists, who showed no gender preference.
- Comparing different lifestyles revealed that women preferred divorced mothers to married fathers and that men preferred mothers who took parental leaves to mothers who did not.
- Findings, supported by real-world academic hiring data, suggest advantages for women launching academic science careers. (Suggestive framing of the CVs? Incentives?)

'The' causal effect of gender - more formally

What are we adjusting for,
and what is part of the gender package deal?

Observed data

- A for observed Actionable exposure level $->$ Gender?
- Y for the observed outcome
- L for observed Lifestyle factors or baseline covariates

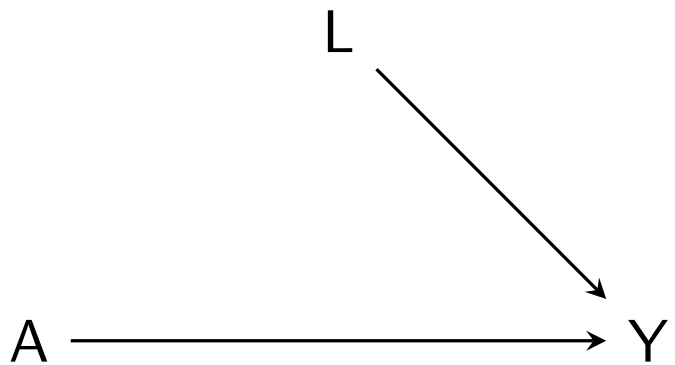


What can we learn? ITT

If A_1 **is randomised**

$\Rightarrow$ $L-$distribution common to all $A-$levels

$\Rightarrow$ learn causal effect of gender perception on evaluation



Comparing (mean) outcomes between **randomized** arms

$$E(Y|A=1) - E(Y|A=0)$$

measures a comparative causal effect...

What can we learn? ITT

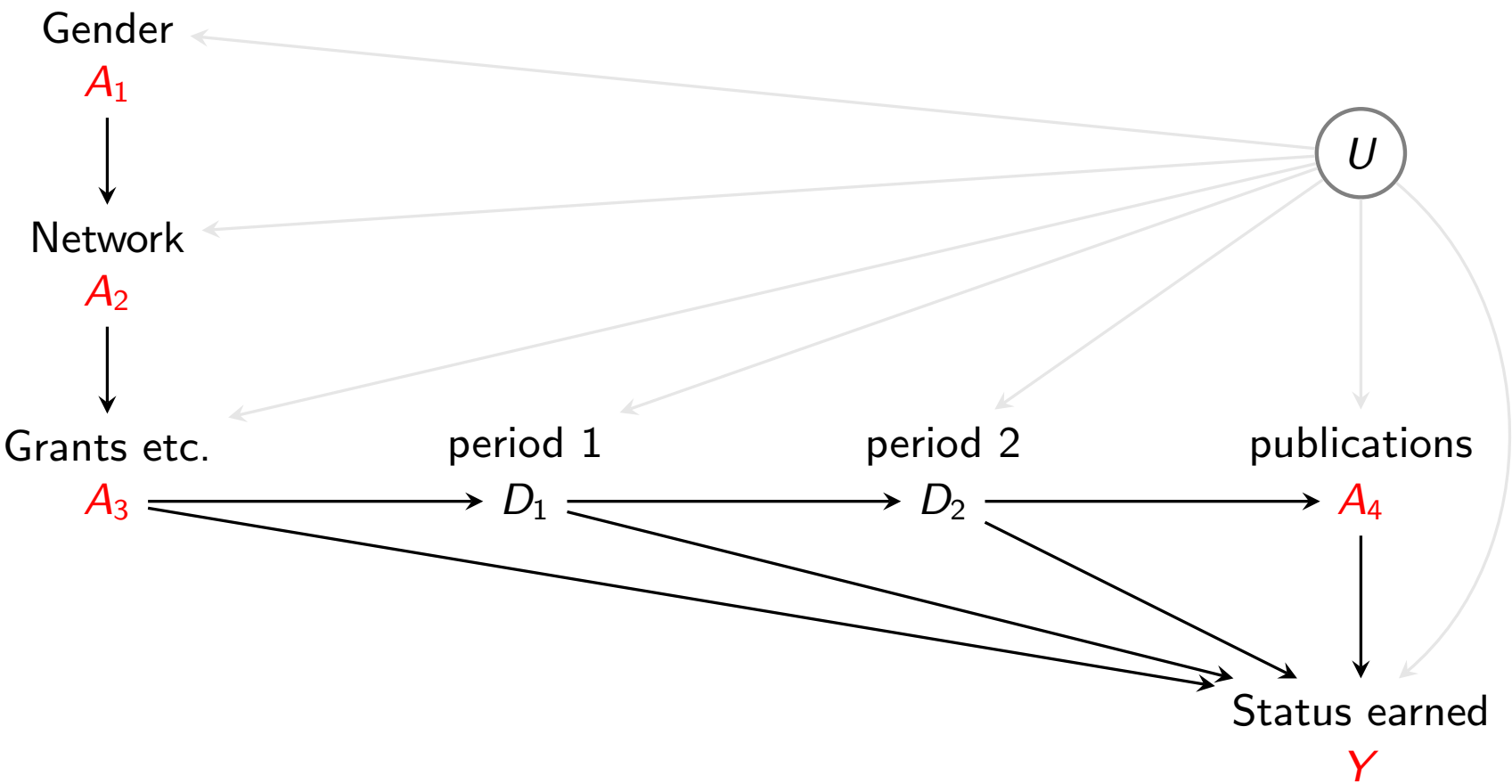
A_1 Randomised $\Rightarrow$
causal effect of gender perception on evaluation
compares (mean) outcomes between randomized arms

ACE: Average Causal Effect also called
ATE: Average Treatment Effect

$$A \text{ randomized} \Rightarrow E(Y|A=1) - E(Y|A=0)$$

averaging over a population of male and female assessors, specific country, type of applicants
What is the gender perception effect on top of a 'female cv' ?

Observed exposure measures



Direct effect of gender and mediation analysis

- Consider outcome - gender (or race or...) association
- as being mediated by a number of factors
  - that are 'post baseline'
  - and you might wish to exclude the effect of
- to assess a causal effect of gender

VanderWeele TJ, Robinson WR., Epidemiology, 2014

Is there gender bias in science?

Take home message

- Please measure your own implicit bias at the Harvard website
- Look for evidence in your own environment
  - getting the data...
  - ask an expert in mediation analysis
- Have both men and women as invited speaker, grant reviewer...
- Speak up when you see it
- Make sure to have mentors for the young faculty

Avoid suggestive images - as in September 2018 cover

unknown_1_8.jpeg 650×846 pixels 19/10/18 05:42

Thank you!

Some more references:

- VanderWeele TJ, Robinson WR. On causal interpretation of race in regressions adjusting for confounding and mediating variables. Epidemiology (Cambridge, Mass.). 2014 Jul;25(4):473.
- Jackson JW. Explaining intersectionality through description, counterfactual thinking, and mediation analysis. Social psychiatry and psychiatric epidemiology. 2017 Jul 1;52(7):785-93.
- Schisterman EF, Swanson CW, Lu YL, Mumford SL. The changing face of epidemiology: gender disparities in citations?. Epidemiology (Cambridge, Mass.). 2017 Mar;28(2):159.