



# THE FRAUGHT QUEST TO ACCOUNT FOR SEX IN STUDIES

Funders and publishers are increasingly asking researchers to account for the role of sex in experiments – a requirement that’s contentious and hard to get right. **By Emily Willingham**

In 2016, pharmacologist Susan Howlett wrote up a study on how hormone levels during pregnancy affect heart function and sent it off to a journal.

When the reviewers’ comments came back, two of the three had asked an unexpected question: where were the tissues from male mice?

Because they were studying high hormone levels related to pregnancy, Howlett, at Dalhousie University in Halifax, Canada, and her team had used only female animals. “I was really surprised that they wanted us to repeat everything in males,” she said.

Nonetheless, they obliged, and their findings were published in 2017. As expected, they found no effect of the hormone progesterone on heart function in males; in females, it influenced the activity of cardiac cells<sup>1</sup>.

Howlett had mixed feelings about the request to add males. “It was a big ask and it was a lot more research.” But in general, she adds, it’s really important to factor sex into studies. “I’m a big proponent of doing experiments in both males and females.”

Many of science’s gatekeepers – granting agencies and academic journals – feel the same way. Over the past decade or so, a growing list of funders and publishers, including

the US National Institutes of Health (NIH) and the European Union, have been asking researchers to include two sexes in their work with cells and animal models.

Two major catalysts motivated these policies. One was a growing recognition that sex-based differences, often related to hormone profiles or genes on sex chromosomes, can influence responses to drugs and other treatments. The other was the realization that including two sexes can increase the rigour of scientific inquiry, enhance reproducibility and open up questions for scientific pursuit.

When studies do include two sexes, the results can be important for health. For example, sex is known to affect people’s responses to common drugs, including some antibiotics<sup>2</sup>, and the risk of cardiovascular disease seems to rise at a lower blood pressure in women than in men<sup>3</sup>.

COVID-19 offers another ready example of why sex should be considered. More men die from the disease<sup>4</sup>, whereas women seem more susceptible to the lingering constellation of symptoms known as long COVID<sup>5</sup>.

The big advantage of looking at more than one sex, says Sabine Oertelt-Prigione, a physician who specializes in gender medicine at Radboud University Medical Center in

Nijmegen, the Netherlands, is that “you might find potential pathways or solutions or new questions that you wouldn’t find otherwise”.

But hoped-for improvements in reproducibility and rigour have been slow to materialize. The policies have generated considerable confusion and controversy over when and how to work the different sexes into study designs, and some researchers argue that ‘sex’ as currently defined is too binary and blunt.

“The number of scientists who accept the importance of studying sex is growing,” said Janine Clayton, director of the NIH Office of Research on Women’s Health (ORWH) in Bethesda, Maryland, in comments e-mailed to *Nature*. “However, there is room for improvement.”

## Diminished representation

As more and more women entered the research arena in the mid- to late twentieth century, some of them began to notice that many clinical studies neglected to include two sexes.

The dearth of female participants resulted in part from a reaction to a tragedy: the use of a sedative called thalidomide during pregnancy had been found to cause congenital anomalies. One upshot was that in 1977 the US Food and Drug Administration (FDA) recommended that almost all women who could become pregnant be excluded from early-phase clinical trials – those that test the safety and efficacy of therapies in healthy volunteers. A policy meant to protect women ended up leaving a vacuum of information on how drugs affect them.

It began to dawn on researchers and funders that excluding a large proportion of the population from these studies or blending the sexes for analyses would have clinical consequences. In response, in 1990, the NIH established the ORWH, and three years later began requiring that women be included in clinical research.

In basic science, however, sex was sidelined until much more recently. A dozen years ago, funders and publishers began to address the imbalance. In 2010, the Canadian Institutes of Health Research implemented a requirement to incorporate sex and gender analyses; in 2013, the EU introduced similar guidelines, which it beefed up into a mandate in 2020. In 2016, the same year that Howlett’s team was asked to add a second sex to their work, the NIH enacted a policy calling for the inclusion of two sexes in studies involving cells, tissues and animals, in part as a way to find signals of sex effects well before any clinical studies were done.

The publishing community is pushing for similar clarity. In 2016, it published the Sex and Gender Equity in Research (SAGER) guidelines, which set out how to report sex-based differences in published research. Individual publishers, including Springer Nature (which publishes *Nature*), have their own policies encouraging researchers to report results by

# Feature

sex, defined as a cluster of biological traits, and sometimes also gender, which is socially defined.

Even getting to that point wasn't easy. Clayton has spearheaded the efforts at the ORWH to account for 'sex as a biological variable' (SABV) since 2012. "I watched her go through it, every year, she and others," says Londa Schiebinger, a specialist in the history of science at Stanford University in California, who has been closely involved in the work. "Just to get sex as a biological variable through the institutes, she had to go to each of these [NIH] institutes and argue her case."

The expectation of the NIH's SABV policy, according to Clayton, is that researchers look for influences of sex or of sex differences – or provide a clear justification for studying a single sex. "Looking for influences of sex or sex differences," wrote Clayton, "is an opportunity, not an obstacle."

But even as the policy was launched, some researchers felt it was the latter.

## The complexity of sex

Accounting for sex in animal and cells studies is not as simple as it might sound.

Delineating sexes on the basis of broad indicators, such as anatomy, elides the deeper complexity of hormones, the key actors in many identified or potential differences between males and females. People who didn't train as endocrinologists "might not know these things", says Jessica Tollkuhn, a molecular biologist at Cold Spring Harbor Laboratory in New York.

Defining sex as a crude binary, predicated on the chromosomes present, or on specific anatomy, could be too limiting. Some species, such as the nematode worm *Caenorhabditis elegans*, have one sex that makes only sperm cells and one that makes both sperm and egg cells. And in a vast assortment of species, sex is determined environmentally rather than chromosomally. And still other species can change sex during their lifetime. Placing cells, tissues or even whole organisms into a pair of categories takes on layers of difficulty in these contexts.

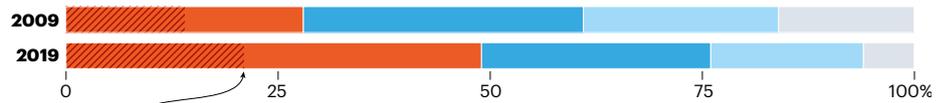
Critics have also argued that there is a logistical problem with the policy: including two sexes will require more animals.

"There's this assumption that if you're doing mouse research and you want to consider both sexes, you'll have to double the numbers," says Irene Miguel-Aliaga, a geneticist at Imperial College London who helped to shape a mandate to use both sexes launched by the UK Medical Research Council earlier this year. Doubling might be needed if sex differences drive a study's hypothesis, but for exploratory purposes, "you just have to have enough animals to tell whether whatever you're finding is relevant to both sexes", she says.

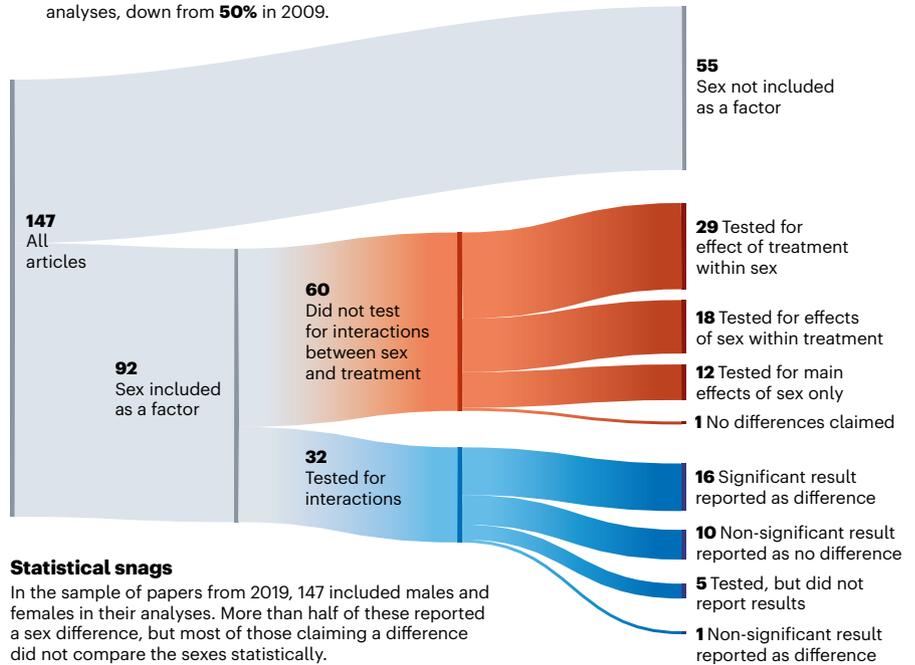
On average, sample sizes might need to

## SEX STUDIES SCRUTINIZED

A sample of 720 papers published in 34 biology journals in 2009 and 2019 suggests that the proportion including males and females has increased substantially. However, of those published in 2019 that did include two sexes, fewer than half conducted analyses that would reveal any effects of sex.



In 2019, **42%** of studies that included males and females conducted sex-based analyses, down from **50%** in 2009.



### Statistical snags

In the sample of papers from 2019, 147 included males and females in their analyses. More than half of these reported a sex difference, but most of those claiming a difference did not compare the sexes statistically.

increase by as much as one-third to meet this bar<sup>6</sup>. The problem with that, says Evan Rosen, chief of the Division of Endocrinology, Diabetes and Metabolism at Beth Israel Deaconess Medical Center in Boston, Massachusetts, is that "mouse work is expensive, and one frustrating aspect of this new stance is that the NIH often demands that we do studies in female mice but balks at providing sufficient funds".

Earlier this year, he and his team published<sup>7</sup> an expansive human and mouse atlas of a type of fat tissue called white adipose tissue, and they ran into an interesting problem: most mouse studies in the field are done in males, which tend to have a lot more fat than do females. By contrast, most human samples are taken during weight-loss surgery, which has an overwhelmingly female patient population. As they began work on their atlas, they realized that their mice and human populations were skewed in opposite ways, and had to ensure that they included tissues from female mice and male humans. In the end, says Rosen, "we did see big differences between lean people and obese people and lean mice and obese mice, but sex fizzled out as a comparator".

Miguel-Aliaga says that even such "negative" findings of no differences are informative.

"It's still good to know that whatever you're studying doesn't show sexual dimorphism or that the treatment it might lead to could apply to both sexes," she points out. Doing these studies is "a win-win".

## Rough road

These policies were meant to compel change, but many scientists struggle to comply with them routinely or to incorporate sexes properly into studies. In the e-mail sent to *Nature*, Clayton notes that, by 2015, 22 years after the NIH established its clinical-trials requirement, fewer than one-third of evaluated NIH-funded randomized controlled trials were including two sexes in their studies or providing an explanation for not doing so. A 2018 review found that the needle had largely remained in place for the previous 14 years<sup>8</sup>.

When women are included in trials, it is often in proportions that do not tally with the real-life prevalence of certain diseases in that group. A 2020 review published by Clayton and her colleagues found that of the 11 disease categories the authors analysed from 2014 to 2018, women were under-represented in 7, including liver and kidney diseases<sup>9</sup>.

Compliance with the newer policy in

animal and cell studies is even patchier. Nicole Weitowich, a SABV researcher in the Department of Medical Social Sciences at Northwestern University's Feinberg School of Medicine in Chicago, Illinois, co-authored a report<sup>10</sup> looking at how sex inclusion in animal studies changed between 2009 and 2019. In 9 research areas across 34 journals, she and her colleagues found that the proportions of studies including two sexes had risen. But in eight of those fields, analysis of data by sex had not increased, and authors rarely explained the omission (see 'Sex studies scrutinized').

Weitowich singles out neuroscience. Studies in this field showed a big increase in including two sexes, yet fewer than half bothered to specify numbers for each sex. That's a reproducibility issue. Sex inclusion is "great", she says, but "if we're not doing sex-based analyses, we're essentially leaving half the data on the table".

A follow-up study by a different group took a closer look at how the same batch of studies had handled the data<sup>11</sup>. Only a minority of them reported data by sex, and in those that did, the sex-based analyses were inappropriate – in 70% of cases not even comparing treatment effects between the sexes – or the results were misinterpreted.

One common error was inferring a sex-based difference if an outcome was significant within one sex but not within the other, even though the two sexes hadn't been compared directly. Values for one group might have a wider range around the average than do values for the other group, for example, just because of individual differences. Testing the groups for significance separately would not show whether they were different; they must be compared with one another using a statistical test.

But the report also described the opposite bias: the risk of erasing genuine sex effects. This risk arises when authors pool sexes for analyses without considering sex as a factor, which they sometimes did even when preliminary calculations indicated sex differences.

COVID-19 again provides a recent example of how the wrong analyses can muddy insight. A 2020 report<sup>12</sup> found differences in the levels of immune and inflammation-related molecules between men and women with COVID-19. But a follow-up analysis<sup>13</sup> by Sarah Richardson, a science historian and director of the Harvard GenderSci Lab in Cambridge, Massachusetts, and her colleagues pointed to errors in the analyses. For three of the results, Richardson and her colleagues wrote, the differences were within the same sex, not between the sexes. For example, in women, the levels of one signalling molecule differed significantly at baseline between those whose condition worsened and those whose condition remained stable, but this pattern did not hold for men.

The original authors had concluded that the

result represented a "between-sex" difference, even though the two sexes hadn't been compared directly. Richardson and her colleagues, by contrast, did a direct comparison and found that the two differences were not significantly distinct, suggesting that sex had no role. They concluded that social factors, such as gender and ethnicity, rather than sex could underlie some of the differences originally attributed to sex.

Some researchers agree that such social factors should be accounted for in clinical trials. But these variables are harder to measure and incorporate. The process of getting sex included as an NIH policy priority "would have been a lot more difficult for gender, even if ultimately it's very difficult to separate sex and gender as health determinants", says Madeleine Pape, a sociologist at the University of Lausanne in Switzerland.

Schiebinger, whose group has spent several years developing questionnaires that address gender for use in clinical trials, hopes that the NIH will include gender as a sociocultural variable one of these days. But it is "waiting for better measures", she says.



## PEOPLE ARE NOT PARTICULARLY GOOD AT CHANGE."

The SAGER guidelines and publishers' own policies on sex and gender are meant to encourage authors to include and report on both sexes. But journals' adherence to the policies is sporadic. An informal review in 2021 suggested that some journal editors continued to resist adoption of SABV policies, asserting that they were not applicable to their fields<sup>14</sup>.

Complaints about lagging adherence and slow uptake are not unexpected, says Eliza Bliss-Moreau, a psychologist at the California National Primate Research Center at the University of California, Davis. "People are not particularly good at change," she notes. She also says that the length of the NIH funding cycles have built in a lag for policies to catch up. "There have been many things put into policy that people have griped about, and 10 or 15 years later, they are just how things are done."

### Partial progress

Despite the bumpy ride, the federal guidelines that were put into place in the early 1990s have led to some important medical discoveries, perhaps a signal that key revelations could emerge from basic research in a few years.

For instance, there are sex-based differences

in the heart's electrical response to several classes of drug, including antidepressants and antibiotics. As a result, sex-based dose adjustments are now recommended for some drugs<sup>2</sup>.

Steroid hormones such as oestrogens and androgens are thought to be primary actors in many of these differences between men and women. For example, women metabolize propranolol, a blood-pressure drug from a class known as beta blockers, more slowly than men do<sup>15</sup>. Researchers think that sex-related steroid hormones acting on the liver can exert these effects. Other factors could include body size and composition, such as the fat:muscle ratio, which tends to be higher in women.

The cut-offs for risk might also differ between men and women. A 2021 analysis of cardiovascular risk related to systolic blood pressure shows what happens if data for two sexes are pooled rather than analysed appropriately<sup>3</sup>. The authors found that when data were pooled, the range for increased risk was a systolic pressure of 120–129 millimetres of mercury (mmHg). But the sex-specific analyses showed that for women, the risk actually begins to climb when systolic blood pressure tops 110 mmHg. If other studies solidify these findings, the result would be a sea change in risk calculation for cardiovascular disease.

That study, as it happens, "was very much inspired and motivated by an NIH request for applications" about sex differences in health outcomes, says Susan Cheng, a cardiologist at Cedars-Sinai Medical Center in Los Angeles, California, and senior author on the report. Without that call for studies specifically designed to look for sex differences, she says, "we had a lot of ideas, but not a thematic focus". Their findings that men and women differ in risk cut-offs "was actually a real 'eureka moment'", Cheng says. "I was like, 'how did we not see this before?'" She attributes the results to the NIH's challenge. "They made it all happen."

Emily Willingham is a science journalist based in California and Texas.

1. Feridooni, H. A. et al. *Am. J. Physiol. Heart Circ. Physiol.* **312**, H46–H59 (2017).
2. Obias-Manno, D. et al. *J. Womens Health* **16**, 807–817 (2007).
3. Ji, H. et al. *Circulation* **143**, 761–763 (2021).
4. Doerre, A. & Doblhammer, G. *PLoS ONE* **17**, e0268119 (2022).
5. Sylvester, S. V. et al. *Curr. Med. Res. Opin.* **38**, 1391–1399 (2022).
6. Buch, T. et al. *J. Mol. Med.* **97**, 871–877 (2019).
7. Emont, M. P. et al. *Nature* **603**, 926–933 (2022).
8. Geller, S. E. et al. *Acad. Med.* **93**, 630–635 (2018).
9. Feldman, S. et al. *JAMA Netw. Open* **2**, e196700 (2019).
10. Weitowich, N. C., Beery, A. & Woodruff, T. *eLife* **9**, e56344 (2020).
11. Garcia-Sifuentes, Y. & Maney, D. L. *eLife* **10**, e70817 (2021).
12. Takahashi, T. et al. *Nature* **588**, 315–320 (2020).
13. Shattuck-Heidorn, H. et al. *Nature* **597**, E7–E9 (2021).
14. Peters, S. A. E. et al. *BMJ Glob. Health* **6**, e007853 (2021).
15. Tamargo, J. et al. *Eur. Heart J. Cardiovasc. Pharmacother.* **3**, 163–182 (2017).