TEAM SCIENCE Focus on communication, not on new centres **go.nature.com/2cvbjpw**

FIELD WORK Polar researchers rescued from hungry bears **go.nature.com/2ctvbzk**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

ROY SCOTT/GETTY

REPRODUCIBILITY

Seek out stronger science

Want to learn how to design an experiment or analyse data? Training is there if you look.

BY MONYA BAKER

When graduate student Alyssa Ward took a science-policy internship, she expected to learn about policy — not to unearth gaps in her biomedical training.

She was compiling a bibliography about the reproducibility of experiments, and one of the papers, a meta-analysis, found that scientists routinely fail to explain how they choose the number of samples to use in a study. "My surprise was not about the omission — it was because I had no clue how, or when, to calculate sample size," Ward says. Nor had she ever been taught about major categories of experimental design, or the limitations of *P* values. (Although they can help to judge the strength of scientific evidence, *P* values do not — as many think — estimate the likelihood that a hypothesis is true.)

Ward's PhD programme required her to take courses in research ethics, but she already knew not to make up data. Instead, she wanted to know how to plan unbiased experiments and conduct rigorous analyses. "Mistakes are more important than misconduct," she says. "I wanted a course on mistakes." So she designed one.

After her internship, she convinced the head of her graduate programme at Johns Hopkins School of Medicine in Baltimore, Maryland, to let her pilot a 7-week course called 'Method, Logic and Experimental Design' for first-year graduate students. Next year, an expanded version of this course will roll out across multiple Hopkins programmes and be required for many trainees.

Scientific irreproducibility — the inability to repeat others' experiments and reach the same conclusion — is a growing concern. Much blame is placed on weak experimental and analytical practices that cause researchers to inadvertently favour exciting hypotheses. A *Nature* survey this year (*Nature* **533,** 452–454; 2016) found that 87% of more than 1,500 researchers named poor experimental design as a cause of irreproducibility; 89% blamed flaws in statistical analysis. Yet few early-career researchers receive formal instruction on these topics. Indeed, when Ward tried to recruit faculty members to lecture for her course, several declined because they had not received such training themselves.

INITIATIVE EXPECTED

Early-career scientists cannot expect to learn everything they need to know in their own laboratories, or even departments, says Alison Gammie, who administers research training programmes at the US National Institutes of Health (NIH) in Bethesda, Maryland. "Science is changing at an incredible rate, and the current principal investigators and investigators were trained in a different era," she says. Because many of the available training opportunities are new and not well known, scientists who want to improve their analytical skills must take the initiative to seek — or create — the resources they require (see 'The learning hunt').

Several researchers have supplied relevant advice in the form of books, articles and webcasts. David Glass, a director at the Novartis Institutes for BioMedical Research in Cambridge, Massachusetts, converted a short ▶

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

course he teaches into the handbook *Experimental Design for Biologists* in part because so little formal training is available. "It seemed odd that we weren't teaching grad students how to actually perform science," he says.

David Vaux, a cell biologist at the Walter and Eliza Hall Institute of Medical Research in Melbourne, Australia, has also penned articles in this vein. An experiment designed to probe alternative explanations can be more powerful than using statistics to ferret out differences between groups, Vaux notes. For example, researchers can engineer a mouse or cell so that a gene can be turned off selectively. However, that kind of manipulation is not an option for human studies. "In clinical sciences, you need statistics, but in basic biology, you should design an experiment where results give clear-cut answers."

When interpreting results, an analytical mindset is more important than any specific instructions for crunching numbers. Too many students focus on rote calculations without examining their data, he says. Researchers need to do enough experiments to know what kind of results to expect, and then treat data sceptically. Wide variations across identical experiments can point to faulty equipment, he says, whereas tiny variations may reveal bias.

If possible, Vaux says, the best option is to plot data points to see whether these alone point to conclusions. Such displays can be more convincing than, say, mean values with error bars because they make clear just how many individual measurements were made.

Before learning data analysis, researchers working with large data sets need to learn 'data wisdom', or the ability to assess how and whether the available data can answer a scientific question, says Bin Yu, a statistician at the University of California, Berkeley. She co-authored the article 'Ten simple rules for effective statistical practice', which has been viewed more than 100,000 times since it was published in June (R. E. Kass *et al. PLoS Comput. Biol.* **12,** e1004961; 2016). "If the data are unreliable, then no fancy method can save you," she says.

ADVICE AVAILABLE

Help with experimental design and analysis is more accessible than many researchers realize. The 60 or so research institutions that receive funding as hubs for the NIH Clinical and Translational Science Awards programme often also provide statistical advice. The Center for Open Science in Charlottesville, Virginia, offers free statistical consultations along with webinars and short tutorials on techniques such as pre-registration, a bias-thwarting strategy that requires researches to formally document analysis plans before collecting results. And the CHDI Foundation, a non-profit drug-development organization targeting Huntington's disease, has assembled a panel of experts on protocol and statistics that anyone in that research community can engage.

But many scientists do not know how to work with statisticians effectively, says Andrew Vickers, a biostatistician at Memorial Sloan Kettering Cancer Center in New York City. For example, it is best to approach a statistician about analysis early, ideally before collecting data.

At Vanderbilt University in Nashville, Tennessee, a course for second-year graduate students teaches them how to work with statisticians. Students bring data from their lab to the university's 'walk-in statistician clinic', and learn how to balance statistical exploration and rigour. But most importantly, they learn that the stats clinic exists — course designer and pharmacologist Joey Barnett hopes that this message will reach their principal investigators, too. "If they are good students they will act as a vector and deliver this to their mentor." Barnett has also observed that trainees learn to seek resources beyond their labs after taking the course.

GETTING TO CLASS

Most research universities offer courses in applied statistics and, increasingly, in data analysis. These can be useful, but trainees may find themselves lumped in with economics or business students with very different core interests. "Scientists learn skills better when they are taught in a domain-specific way than when you shuttle them off to math and computer science departments," says Ethan White, an ecologist at the University of Florida, Gainesville, who has designed courses in quantitative methods.

Such discipline-focused courses are not always available, but faculty members are generally sympathetic when students ask for help, says Randall Reed, a molecular biologist who acted as the faculty sponsor for the course Ward designed. "She was more successful than I would have been in engaging faculty and getting them to participate."

Ward spoke to more than a dozen faculty members in an attempt to find guest lecturers for her course. But recruiting professors was easier than pitching content at the appropriate level, she says. "We're looking for the magic that is general enough to help everyone but specific enough so that students know how to apply it." She found a solution in copious extra-credit assignments that, for example, required graduate students to work out how to validate reagents in their own labs. Plans for subsequent years include a series of online lectures supplemented with in-person discussions and projects tailored to specific disciplines.

"Help with experimental design and analysis is more accessible than many researchers realize."

Resources need not come from a trainee's own institution. Next year, a trio of experimental psychologists will launch a week-long residential course on methods for reproducible science for early-career researchers at Cumberland Lodge in Windsor, UK, taught by global experts in psychology and statistics. And Rafael Irizarry at Harvard University in Cambridge, Massachusetts, offers a massive open online course on data analysis in reproducible genomics; some 30,000 people a year sign up for the most basic module.

Scientific societies are also starting to step up to the plate, says Gammie. The US Society for Neuroscience, for example, has created a series of webinars about experimental design, analysis and reporting for its members. The American Physiological Society is creating similar resources. Other focused organizations, such as the American Society for Cell Biology and the Federation of American Societies for Experimental Biology, have sponsored relevant workshops and set guidelines.

Peter Grabitz, a medical student at the Charité university hospital in Berlin, co-founded the Berlin Open Science Group to learn how to handle data more robustly and to share them more broadly. The group of researchers and open-science advocates began as an informal gathering coordinated by the online platform Meetup, but has now swelled to include 450 members keen to swap technical tips on data-management tools. If researchers make the effort to seek advice, Grabitz says, they can find people willing to provide it. ■

Monya Baker *writes and edits for* Nature *from San Francisco, California.*

ANALYSIS AIDS

The learning hunt

Training resources for statistics and experimental design are available, if not immediately apparent. For more, see go.nature.com/2cwbl8e.

- **Seek discipline-specific training.** Before enrolling in classes aimed at maths, business or computer science, enquire how students from scientific disciplines are accommodated and what courses are tailored for your discipline.
- **Ask around.** Your institution's department head or postdoctoral office may know of resources that your supervisor has never heard of. So might the education committee in your local or national scientific society.
- **Search.** Simply putting words such as 'training' or 'beginner' with 'experimental design' or 'data analysis' and the name of your discipline or institution into a search engine can reveal just-offered courses or instructive publications.
- **Organize.** If the training you need is not available online or nearby, set up an interest group or even a course. M.B.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.