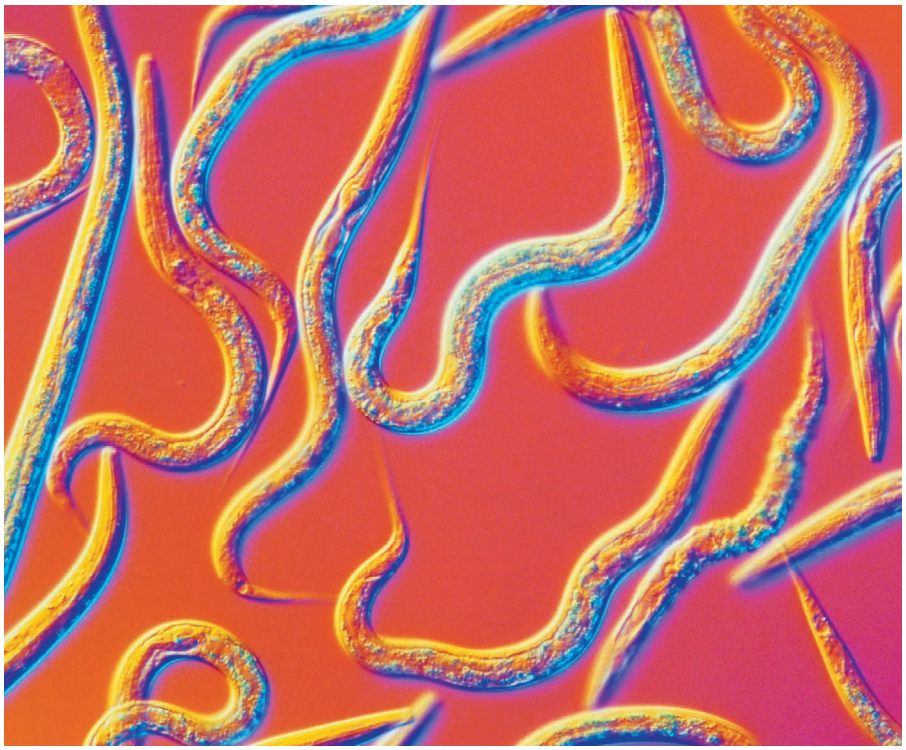
COMMENT

EVOLUTION The astonishing range of Charles Darwin's domestic discoveries **p.389**

CONSERVATION How power politics hijack US parks for private gain **p.390**

SPACE Documentary celebrates 40 years of NASA's Voyager probes **p.392**

ETHOLOGY Patrick Bateson, pioneer in animal-behaviour science, remembered **p.394**



SINCLAIR STAMMERS/SPL

The free-living roundworm *Caenorhabditis elegans* is about 1 millimetre long.

A long journey to reproducible results

Replicating our work took four years and 100,000 worms but brought surprising discoveries, explain **Gordon J. Lithgow**, **Monica Driscoll** and **Patrick Phillips**.

About 15 years ago, one of us (G.J.L.) got an uncomfortable phone call from a colleague and collaborator. After nearly a year of frustrating experiments, this colleague was about to publish a paper[1] chronicling his team's inability to reproduce the results of our high-profile paper[2] in a mainstream journal. Our study was the first to show clearly that a drug-like molecule could extend an animal's lifespan. We had found over and over again that the treatment lengthened the life of a roundworm by as much as 67%. Numerous phone calls and e-mails failed to identify why this apparently simple experiment produced different results between the labs. Then another lab failed to replicate our study. Despite more experiments and additional publications, we couldn't work out why the labs were getting different lifespan results. To this day, we still don't know.

A few years later, the same scenario played out with different compounds in other labs[3,4]. Around the same time, there was a roiling debate about whether resveratrol — a compound found in red wine — could extend lifespan in lab animals.

The possibility of drugs that stall ageing launched companies and a scientific subfield, but work in the field brought the realization that robust longevity outcomes could be challenging to replicate. Ageing research has long battled to distance itself from pseudoscientific claims. Irreproducible results from respected labs raised the spectre of yet more false promises. This had a chilling effect: some researchers (including G.J.L.) paused work on pharmacological compounds for years.

Nonetheless, scores of publications continued to appear with claims about compounds that slow ageing. There was little effort at replication. In 2013, the three of us were charged with that unglamorous task.

We have certainly not resolved discrepancies in the literature. But, by tracking the individual lifespans of more than 100,000 worms, we have found how crucial it is to understand sources of variability between labs and experiments. We even see hints of new biology that may explain discrepancies.

BROADER PROBLEM

Improved reproducibility often comes from pinning down methods. Scientists studying autophagy — the process by which cells remove degraded components — have coordinated efforts to craft and update extensive guidelines on, for instance, how to quantify that a component has been engulfed or how to verify that a gene is involved in the process[5]. In another, now-famous example, two cancer labs spent more than a year trying to understand inconsistencies[6]. It took scientists working side by side on the same tumour biopsy to reveal that small differences in how they isolated cells — vigorous stirring versus prolonged gentle rocking — produced different results.

Subtle tinkering has long been important in getting biology experiments to work. Before researchers purchased kits of reagents for common experiments, it wasn't unheard of for a team to cart distilled water from one institution when it moved to another. Lab members would spend months tweaking conditions until experiments with the new institution's water worked as well as before.

Sources of variation include the quality and purity of reagents, daily ▶

fluctuations in microenvironment and the idiosyncratic techniques of investigators[7]. With so many ways of getting it wrong, perhaps we should be surprised at how often experimental findings are reproducible.

TO THE WORMS

One organization focusing on inconsistencies in ageing studies was the National Institute on Aging (NIA) at the US National Institutes of Health. After all, the NIA was paying for the work and realized that the results needed to be iron-clad. The NIA also had a success story; its Intervention Testing Program had discovered that the drug rapamycin extended the lifespans of mice at three independent labs applying a common protocol[8]. The NIA initiated the *Caenorhabditis* Intervention Testing Program with similar goals, and tasked us — three researchers with different areas of expertise who had not worked together before — to systematically test ageing interventions in the nematode *Caenorhabditis elegans*.

In the lab, these worms live on agar plates and are fed a diet of live bacteria. Test compounds are mixed into the agar or feedstock, and longevity is assessed by mobility — if a worm moves when prodded with a metal wire, it's alive! This technique has been used by worm researchers for more than 25 years. Our first task, to develop a protocol, seemed straightforward.

But subtle disparities were endless. In one particularly painful teleconference, we spent an hour debating the proper procedure for picking up worms and placing them on new agar plates. Some batches of worms lived a full day longer with gentler technicians. Because a worm's lifespan is only about 20 days, this is a big deal. Hundreds of e-mails and many teleconferences later, we converged on a technique but still had a stupendous three-day difference in lifespan between labs. The problem, it turned out, was notation — one lab determined age on the basis of when an egg hatched, others on when it was laid.

We decided to buy shared batches of reagents from the start. Coordination was a nightmare; we arranged with suppliers to give us the same lot numbers and elected to change lots at the same time. We grew worms and their food from a common stock and had strict rules for handling. We established protocols that included precise positions of flasks in autoclave runs. We purchased worm incubators at the same time, from the same vendor.

We also needed to cope with a large amount of data going from each lab to a single database. We wrote an iPad app so that measurements were entered directly into the system and not jotted on paper to be entered later. The app prompted us to include full descriptors for each plate of worms, and ensured that data and metadata for each experiment were proofread (the strain names MY16 and my16 are not the same). This simple technology removed small recording errors that could disproportionately affect statistical analyses.

Once this system was in place, variability between labs decreased. After more than a year of pilot experiments and discussion of methods in excruciating detail, we almost completely eliminated systematic differences in worm survival across our labs[9] (see 'Worm wonders').

Our data revealed another, startling source of variation. Even in a single lab performing apparently identical experiments, we could not eliminate run-to-run differences. After dozens of experiments, we realized that some cohorts of worms could partition into one of two modes of ageing: short-lived or long-lived. The reason so far is not clear, but the discovery shows that, if we want to understand a chemical compound's effects, we need to repeat more experiments than we realized. We also need to consider both the number of worms studied and the number of independent experiments conducted.

It was against this backdrop that we began testing compounds. To probe the influence of genetic background, we used a wide range of *Caenorhabditis* strains and species. *C. elegans* and *C. briggsae* may look alike under the microscope, but genetically the two species are as different as a whale and a bat[10].

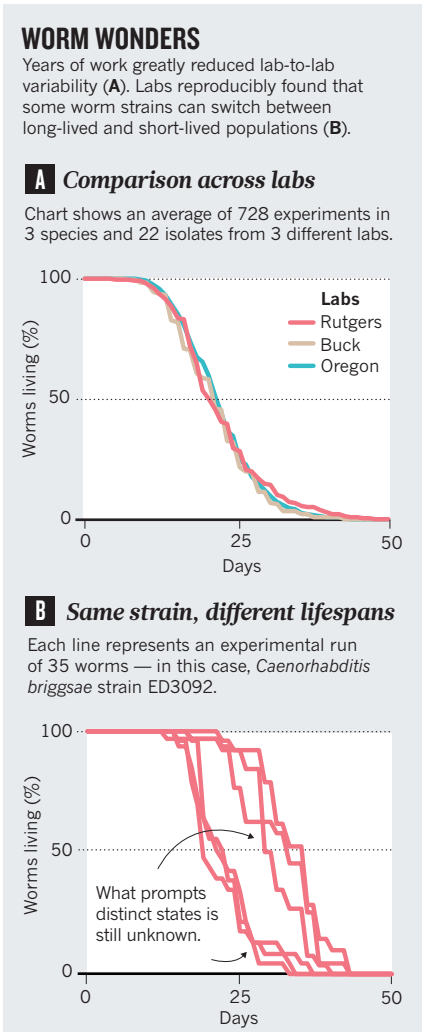
Our results are both encouraging and sobering. We anticipate that compounds that extend lifespan across such genetic diversity may be good candidates for further study. (Microscopic worms might thus eventually be used to find life-extending compounds in humans.) We have found one compound that lengthens lifespan across all strains and species. Most do so in only two or three strains, and often show detrimental effects in others. This raises a series of intriguing mechanistic questions, and might help to explain the genetic influences in an individual's response to an ageing intervention.

Other discrepancies are probably based on small methodological differences in survival assays. For example, small fluctuations in temperature in incubators and lab benches are likely to be significant, as are differences in food quality.

It is a rare project that specifies methods with such precision. In fact, several mouse researchers have argued that standardization is counterproductive — better to focus on results that persist across a wide range of conditions than to chase fragile findings that occur only within narrow parameters.

We argue that another way forward is to chase down the variability and try to understand it within a common environment. We are now working together to search for molecular differences that distinguish short-lived and long-lived batches of worms within the same strain, a phenomenon we never could have uncovered had we not eliminated nearly all other sources of variability.

We have learnt that to understand how life works, describing how the research was done is as important as describing what was observed. ■



SOURCE: REF. 9

Gordon J. Lithgow *is a geneticist at the Buck Institute for Research on Aging in Novato, California, USA.* **Monica Driscoll** *is a developmental biologist at Rutgers University in Piscataway, New Jersey, USA.* **Patrick Phillips** *is an evolutionary geneticist at the University of Oregon in Eugene, USA.*
e-mail: glithgow@buckinstitute.org

1. Keaney, M. & Gems, D. *Free Radic. Biol. Med.* **34,** 277–282 (2003).
2. Melov, S. *et al. Science* **289,** 1567–1569 (2000).
3. Petrascheck, M., Ye, X. & Buck, L. B. *Nature* **450,** 553–556 (2007).
4. Zarse, K. & Ristow, M. *PLoS ONE* **3,** e4062 (2008).
5. Klionsky, D. J. *Mol. Biol. Cell* **27,** 733–738 (2016).
6. Hines, W. C., Su, Y., Kuhn, I., Polyak, K. & Bissell, M. J. *Cell Rep.* **6,** 779–781 (2014).
7. Barrows, N. J., Le Sommer, C., Garcia-Blanco, M. A. & Pearson, J. L. *J. Biomol. Screen.* **15,** 735–747 (2010).
8. Harrison, D. E. *et al. Nature* **460,** 392–395 (2009).
9. Lucanic, M. *et al. Nature Commun.* **8,** 14256 (2017).
10. Cutter, A. D., Jovelin, R. & Dey, A. *Mol. Ecol.* **22,** 2074–2095 (2013).