



STANDARD MODEL

Questions raised about the use of 'ALS mice' are prompting a broad reappraisal of the way that drugs are tested in animal models of neurodegenerative disease. **Jim Schnabel** reports.

Several years ago, clinical neurologist Michael Benatar set out to find a drug he could test on some of his patients with amyotrophic lateral sclerosis (ALS). This condition is apt to strike otherwise healthy adults, slowly destroying the neurons that control their muscles. Its roster of famous sufferers includes the physicist Stephen Hawking and the late US baseball star Lou Gehrig.

Except in rare cases, such as Hawking's, ALS progresses inexorably, causing respiratory failure within a few years of diagnosis. Decades of study have revealed a few mutant genes that could cause familial forms of the disease, but no one knows what causes the vast majority of cases — and despite trial after trial of prospective therapies, no therapy has ever been shown to have a major impact on the disease.

One of the first steps Benatar, at Emory University School of Medicine in Atlanta, Georgia, took was to review published data on more than 150 drug tests that other researchers had conducted in the standard mouse model of ALS. Engineered to carry multiple copies of the mutated superoxide dismutase 1 (*SOD1*) gene that causes some inherited ALS cases,

the mouse reliably develops and succumbs to a neuron-killing disease that closely resembles the human condition.

But as Benatar reviewed these mouse studies, he was dismayed to find that the data were of little use to him. Most of the published experiments, including some in top-rank journals, had been done "with small sample sizes, with no randomization of treatment and control groups, and without blinded evaluations of outcomes", he says.

Benatar also found that in the spread of reported results for some drugs, there was statistical evidence that only positive results had been published. Informal conversations with other researchers convinced him that some had tried and failed to confirm reported positive results, but had never published those non-confirmations.

All in all, a body of data that should have yielded useful information was, to Benatar, "questionable at best". He wrote a paper making what he could of the data, and also making clear how flawed he thought they were.

When Benatar's paper¹ was published in *Neurobiology of Disease* in early 2007, it seemed

to have little impact. But in the year and a half since, other investigators have come to similar, and indeed stronger, conclusions. "There is a dawning realization that we may not have designed our mouse drug trials rigorously enough," says Melanie Leitner, chief scientist at Prize4Life, a non-profit organization based in Cambridge, Massachusetts, that promotes ALS research.

That realization is spreading: some researchers are coming to believe that tests in mouse models of other neurodegenerative conditions such as Alzheimer's and Huntington's may have been performed with less than optimal rigour. The problem could in principle apply "to any mouse model study, for any disease", says Karen Duff of Columbia University in New York, who developed a popular Alzheimer's mouse model. In May, a dozen preclinical researchers and mouse model experts thrashed out the issues in a web discussion on the 'Alzforum' website, an online venue where researchers routinely gather to debate neurodegenerative disease issues.

"There has to be sort of a course correction in the field," says Lorenzo Refolo, who oversees grants for preclinical work on neurodegenerative diseases at the National Institute of Neurological Disorders and Stroke in Bethesda, Maryland, "otherwise these practices are just going to continue."

The results of drug tests in mice have never translated perfectly to tests in humans. But in recent years, and especially for neurodegenerative diseases, mouse model results have seemed

ALS/DTI

nearly useless. In the past year, for example, three major Alzheimer's drug candidates, Alzhemed (3-amino-1-propanesulphonic acid), Flurizan (tarenflurbil) and bapineuzumab, all of which had seemed powerfully effective in mouse models, have performed weakly or not at all in clinical trials involving thousands of human Alzheimer's patients.

In the case of ALS, close to a dozen different drugs have been reported to prolong lifespan in the SOD1 mouse, yet have subsequently failed to show benefit in ALS patients. In the most recent and spectacular of these failures, the antibiotic minocycline, which had seemed modestly effective in four separate ALS mouse studies since 2002, was found last year to have worsened symptoms in a clinical trial of more than 400 patients².

The minocycline clinical trial, and the other ill-fated ALS-drug trials that preceded it, would never have happened had the prior mouse studies been done properly, says Sean Scott, president of the ALS Therapy Development Institute (ALS TDI), a non-profit biotech company based in Cambridge, Massachusetts. The ALS TDI was set up in 1999 to swiftly screen approved drug compounds for any that could slow the disease in ALS mice. In 2001, it found that ritonavir, an antiretroviral drug used against HIV, seemed to extend lifespan in the mice and the ALS TDI initiated a small safety trial in ALS patients. "But instead of just moving onto the next drug," says Scott, "we kept screening ritonavir in the mice, in the hope that we could learn how it worked and maybe improve on it."

The drugs don't work

Scott and his colleagues were shocked to find that when they scaled up the ritonavir tests by adding more mice, the drug's effect on lifespan didn't become statistically clearer — it disappeared altogether. For the patients in the safety trial, the drug also showed no benefit, and at the highest dose, says Scott, "it was even a little bit detrimental".

Chastened by the experience and with the funding to find out what had gone wrong, the ALS TDI spent the next few years refining its mouse-trial methods and trying to characterize likely sources of error. Its scientists also applied their improved methodological rigour in further tests of prospective ALS drugs, including drugs such as minocycline that had seemed effective in other labs' mouse studies. "But really in the end," says Scott, "we were heartbroken, because even using dramatically more animals than any of those other labs, using very sophisticated pharmacological formulations, and paying attention to levels of the drugs achieved in the mouse nervous systems, we just

could not get any of those drugs to work."

At the time, in late 2006, the ALS TDI was funded in part by a US\$200,000 grant from the Muscular Dystrophy Association (MDA), which has supported ALS research since the 1950s. Sharon Hesterlee, director of research development at the MDA, oversaw the grant and remembers Scott and his colleagues at meetings "taking some flak from members of the academic community, who in effect said 'if you can't reproduce any of our results you're just not doing it right'".

But Hesterlee had already concluded that the ALS TDI "were actually so much more rigorous in their approach, that if there

mouse was not excluded appropriately from a study. Such an exclusion should have happened if, for example, an animal died young of non-ALS causes — Scott came to believe that mouse colonies in some academic labs were "filthy" with infections — or failed to express enough copies of the mutant *SOD1* gene. Further noise could creep in if the treatment and

Design, power, and interpretation of studies in the standard murine model of ALS

SEAN SCOTT¹, JANICE E. KRANZ², JEFF COLE¹, JOHN M. LINCECUM¹, KENNETH THOMPSON¹, NANCY KELLY¹, ALAN BOSTROM², JILL THEODOSS¹, BASHAR M. AL-NAKHALA¹, FERNANDO G. VIEIRA¹, JEYANTHI RAMASUBBU¹ & JAMES A. HEYWOOD¹

¹ALS Therapy Development Institute, Cambridge, Massachusetts, and ²Department of Epidemiology and Biostatistics, University of California, San Francisco, USA

Abstract
Identification of *SOD1*^{G93A} as the mutated protein in a significant subset of familial amyotrophic lateral sclerosis (FALS) cases is considered the standard model for FALS and ALS therapeutic studies. To date, there have been no published studies describing therapeutic agents that extend the lifespan of this mouse. However, no therapeutic effect was observed in a clinical trial of 5429 *SOD1*^{G93A} mice from our efficacy studies to quantify the



Sean Scott identified sources of noise in mouse model studies for amyotrophic lateral sclerosis.

was anything to find they would have seen it". The MDA soon organized a much larger grant to the ALS TDI, one that would provide it with about \$6 million a year to continue its work. At the same time, Hesterlee also pressured the organization to publish its analysis of the problems with *SOD1* mouse tests. "That way our own grant reviewers could start applying those lessons to other mouse studies," she says.

The analysis that the ALS TDI eventually put together overlapped with Benatar's, but also went further: Scott and his colleagues concluded that the previous positive drug trials in ALS mice were likely to have been so plagued by non-drug-related variations in mouse lifespan that this 'noise' was really all they had measured. In the context of small sample sizes and a bias against negative results, they noted, a high degree of noise could easily have led to the appearance of positive results even when no drug effect had existed.

In the ALS TDI's own work, the greatest source of noise had been observed when a

control groups were not evenly matched by gender and litter-membership, because mouse lifespan naturally varied according to these factors. Above all, Scott and his colleagues came to recommend starting each study with at least 24 mice in each treatment or control group — roughly double the norm — to ensure that any real drug effect would statistically rise above the noise.

Scott submitted the ALS TDI findings to the *Proceedings of the National Academy of Sciences* in late 2006 where, he says, "one reviewer flamed us" over the organization's failure to reproduce the positive mouse studies. Scott says he just "wanted to get it out there" and the paper was published in the journal *Amyotrophic Lateral Sclerosis* in January of this year³.

The immediate response was "muted", Leitner observes. "I do think it's a difficult thing for an academic researcher, working as hard and as fast as you can, to have some group tell you that you've made mistakes in your experimental design — especially when that group isn't necessarily operating under the same constraints as you are."

As a biotech company, the ALS TDI had never before published its research in a peer-reviewed journal. Although seven PhD scientists oversaw its laboratory work, its founder, James Heywood, was not a scientist but an engineer who had started the organization after his brother received an ALS diagnosis. Sean Scott had joined the organization as a volunteer after his mother developed the disease. He held a BA in rhetoric from the University of California, Berkeley, and had no formal training as a scientist. And he was now first author on a paper which concluded, in essence, that some high-powered academic researchers had been chasing their tails for years.

Robert Friedlander, at Harvard Medical School's Brigham and Women's Hospital, and the lead author on the first positive study of minocycline in SOD1 mice⁴ defends his work, saying that three other labs independently found similar results. "The fact that ALS TDI did not reproduce these results raises questions as to their methodologies," he says. As for the failed clinical trial of minocycline, Friedlander suggests that the drug may have been given to patients at too high a dose — and a lower dose might well have been effective. "In my mind, that was a flawed study," he says.

Neurologist Jeff Rothstein, who runs a large ALS research lab at Johns Hopkins University School of Medicine in Baltimore, Maryland, says of ALS TDI, "they've done some nice statistics". But the company's failure to reproduce his lab's positive study in 2002 of the anti-inflammatory drug Celebrex (celecoxib) in SOD1 mice⁵ might have been due to differences in study design, he says. Rothstein says that his lab confirmed Celebrex's biological effect at reducing neuroinflammation, whereas ALS TDI didn't look for it. "Were they at variance with us because they never got biological efficacy? Hard to know," says Rothstein. Celebrex later failed in a clinical trial in ALS patients.

The findings of Scott and his colleagues do seem to be resonating now among preclinical researchers. Still, Leitner thinks that too many scientists — including those working on conditions other than ALS — remain unaware that their mouse-model studies may be flawed.

Non-exacting standards

In the recent Alzforum event that Leitner organized, Scott and other researchers discussed the ALS TDI study and in general expressed concerns that these methodological issues went beyond ALS. "People will do an experiment once with ten animals and get a result, and if it's the right result it gets published in a high-profile journal," says Duff, who has followed the ALS TDI story with keen interest. "And there's no requirement that you show the effect again with a different set of mice, or in a larger group of mice, or in a different model."

"Many of the neurodegenerative models out there are on mixed and segregating genetic backgrounds," says Greg Cox, a forum participant and mouse geneticist who works for



"There is a dawning realization that we may not have designed our mouse drug trials rigorously enough."
— Melanie Leitner

the world's largest provider of research-grade mice, the Jackson Laboratory in Bar Harbor, Maine. "So in those cases even [inbred] littermates aren't genetically identical. And if a genetic background effect [for example, on mouse lifespan] is there, you could end up measuring those background effects more than the drug effect you want to focus on."

Mike Sasner, a neuroscientist also at the Jackson Laboratory, notes that spontaneous genetic changes often affect the disease-causing mutant gene directly. For SOD1 mice, and for Alzheimer's mice engineered to overexpress amyloid protein, changes are often seen in the number of copies of

the disease-driving 'transgene'. For Huntington's disease mice, spontaneous changes can alter the number of disease-causing repeat sequences within the transgene. "So if you're breeding these Huntington's mice over ten generations and you go from 100 repeats to 50 repeats, you're going to basically lose the phenotype," says Sasner.

He adds that the Jackson Lab now checks the genetic make-up of its transgenic mice to reduce the problem, but not every researcher is aware of such issues. "You might create a mouse in your lab and distribute it to ten different people," he says, "so there's ten different colonies all over the world, and they're all diverging from each other, genetically. So when I'm publishing my paper I'm talking about this mouse, and you think you have the same

mouse in your hands. But do you really?"

Sasner says that he and other scientists now hope to draft a formal document with guidance for preclinical researchers on these issues. Leitner applauds such efforts, but thinks the National Institutes of Health (NIH) needs to get involved too. "I believe that if the government doesn't encourage consideration of these issues and support them, it's going to be a very hard sell," she says. "Because basically the ALS TDI study suggests everyone needs to conduct much more thorough and expensive animal-model trials."

Duff agrees. "There just aren't the resources now to do really large, well-powered mouse studies. So I think the NIH should get ahead of the curve here, for example setting up a programme and sending out Requests for Applications to study this, to look more at the mouse models and, in the same way that Sean Scott's group did, to see why they're failing to translate."

Even Refolo thinks that the agency should take the lead in addressing these problems. "I think there just has to be a [new] policy," he says. But he emphasizes that the NIH grant process is regulated largely by the academic community itself. "The academics have to embrace the policy, and people who are reviewing grants and papers have to embrace it and also the editors of these journals have to embrace it." And so far Refolo sees no sign of a major change coming. "If [these issues] are being addressed in study section, where these grants are reviewed, it's at least below my radar."

Even if preclinical researchers ultimately do clear up the methodological flaws in mouse studies, they'll have other issues to deal with. For example, as Benatar noted in his 2007



Model behaviour: macaques have been engineered to carry the human Huntington's disease gene.

M. METZ

YANG, S.-H. ET AL. NATURE 453, 921-924 (2008)



Non-profit biotech ALS TDI challenged some of the methods commonly used in drug tests on mice.

paper, SOD1 mice are typically treated with drugs well before the onset of symptoms. Yet in nearly all human cases, Benatar says, “we have no capacity to initiate therapy presymptomatically, so it’s wishful thinking to suppose that success with presymptomatic treatment in a mouse is going to translate into efficacy in a human”.

The wrong model?

Perhaps the biggest issue facing the field is whether the mouse models faithfully reproduce the biology of the human disease. Alzheimer’s mouse models typically develop amyloid ‘plaques’ in their brains, but they do not develop an Alzheimer’s-like dementia and anti-amyloid strategies have repeatedly failed to slow the disease in clinical trials. Parkinson’s researchers have never had a good mouse model for the full disease process, and even the mouse model for Huntington’s disease — a relatively simple genetic disease — does not fully reproduce the clinical signs seen in humans with the disorder.

SOD1 mice have often been considered one of the most accurate animal models of any neurodegenerative disease. Although the SOD1 gene is mutated in only around 20% of human familial cases (representing 2–3% of all ALS) the disease that the animals get so closely resembles the common, sporadic form of human ALS

that the two maladies have been assumed to share a ‘final common pathway’ of neuronal destruction.

That assumption is now being questioned. In the past two years, researchers have found evidence suggesting that a DNA-binding protein, TDP-43, could be the trigger for sporadic ALS when it is malformed or improperly processed inside cells — and that SOD1-driven ALS might really be a distinct disorder. “The idea that the SOD1 model could be extended to sporadic ALS patients is not holding up,” says Virginia Lee, a neuropathologist at the University of Pennsylvania in Philadelphia, whose lab was the first to report the link between TDP-43 and sporadic ALS⁶. Her lab is one of several now racing to devise a mouse model of ALS carrying mutated TDP-43.



“I’m talking about this mouse, and you think you have the same mouse in your hands. But do you really?”

— Mike Sasner

The debate over SOD1’s relevance to the majority of human ALS cases feeds a broader worry, namely that it may be unrealistic to think of modelling the full complexity of ageing-related human brain disorders in mice whose disease course is usually accelerated by a single, crude genetic modification.

In theory, with their more human-like nervous systems, monkeys should make much better models of human neurodegenerative diseases. Aged vervets have already been used to test Alzheimer’s vaccines; and Anthony Chan and his colleagues at the Yerkes National

Primate Research Center in Atlanta, Georgia, have described the creation of macaques that carry the human Huntington’s disease gene, and suffer from a very similar disorder⁷. “I definitely think that the non-human primate models need to be brought to the forefront much more than they have been,” says John Morrison, an animal-model researcher at Mount Sinai School of Medicine, New York.

But aside from the moral issues that this would raise for some researchers, a switch to monkeys could be prohibitively expensive. Scott estimates that a six-month study of about 50 ALS mice would cost roughly \$100,000, whereas Stuart Zola, who heads the Yerkes Center, estimates that more than \$500,000 would be needed to study 50 ordinary macaques for two years — and possibly much more time and money would be required to study drugs for slow-burning neurodegenerative diseases whose effects may only become apparent with advanced age.

Scott, Duff, Rothstein and others suggest that mouse models should still be used, but that drug tests in them should target specific, disease-related molecular pathways — for example, TDP-43 accumulations, if they turn out to be relevant — instead of broader endpoints such as lifespan or behaviour, where mice and men are inherently mismatched. Scott says he now sees the SOD1 mouse as “perhaps a pathway model as opposed to a disease model, and if we can affect survival [with drugs], great”. The ALS TDI is currently finishing a study of gene expression patterns in various tissues from SOD1 mice and humans with ALS, to find what molecular pathways of disease they share — if any.

Mouse models could therefore end up being not only more difficult and expensive to use with acceptable rigour, but at the same time more narrowly predictive of the human condition. But whether preclinical researchers will accept such a radical change remains to be seen. “I think there’s a sense of desperation that we need a convenient model for bringing drugs to clinical trial,” says Benatar. “And I do sort of hear that concern.” But desperation, he adds, is an inadequate justification for the continued use of a poor model. “It’s a bit like the proverbial drunk who keeps looking for his lost keys under the lamp post, simply because the light’s better there.” ■

Jim Schnabel is a freelance writer based in Maryland.

1. Benatar, M. *Neurobiol. Dis.* **26**, 1–13 (2007).
2. Gordon, P. H. et al. *Lancet Neurol.* **6**, 1045–1053 (2007).
3. Scott, S. et al. *Amyotroph. Lateral Scler.* **9**, 4–15 (2008).
4. Zhu, S. et al. *Nature* **417**, 74–78 (2002).
5. Drachman, D. B. et al. *Ann. Neurol.* **52**, 771–778 (2002).
6. Neumann, M. et al. *Science* **314**, 130–133 (2006).
7. Yang, S.-H. et al. *Nature* **453**, 921–924 (2008).

ALS TDI

JACKSON LABORATORY