more certain of success, Luo says. The spacecraft could launch in 15–20 years, he adds, around the same time as the Taiji group says that it could launch. Luo thinks that a simpler project is more realistic now, but says that TianQin could lay the groundwork for a Taiji-like project in the future.

Wu Ji, director-general of the Chinese Academy of Sciences' National Space Science Center, says that the TianQin and Taiji teams should merge. "If China decides to have a space gravitational mission, there should be an integrated one, with a new name probably. There is no way to support two missions at the same time."

Both Wu Yue-Liang and Luo are confident that their proposals will move forward to the concrete design phase in the next five years. Taiji currently receives money from the Chinese Academy of Sciences and TianQin from the city of Zhuhai — but both need much more cash. The LIGO discovery could increase their chances of success. The "government will know more the importance of fundamental research" in gravitational waves, says Wu Ji. "China should catch up in this area," he adds.

On 5 March, the Chinese central government released a draft list of 100 strategic projects that will be emphasized in the country's next five-year plan, which includes "a new generation of heavy launch vehicles, satellites, space platforms and new payload" and a "deep-space station".

Chinese researchers could also end up collaborating with Europe. As well as its main project, the Taiji group has outlined the possibility of a direct collaboration with eLISA: it would either contribute 1.5 billion yuan directly or develop its own scaled-down, 8-billion-yuan version of eLISA that would coordinate closely with the European effort, sharing data. Heinzel recommends that a united Chinese group work on one of these less ambitious options.

The direct contribution from China in particular could be a boon for eLISA. Originally, NASA collaborated with ESA on a planned space-based gravitational-wave observatory, named LISA. But the United States pulled out of LISA five years ago and ESA had to pare down the mission, resulting in the eLISA proposal. China's entry into the project could fill that hole, says Rainer Weiss, a physicist at the Massachusetts Institute of Technology in Cambridge, who is credited as the chief inventor of LIGO. This would perhaps allow Europe to pursue a design closer to that of LISA, which was better equipped than the eLISA proposal and would have had a longer mission lifetime.

A decision is needed soon if China is to achieve a launch date around 2030, cautions Heinzel. "Now is the time to do very serious technology development," he says. "It is time to start making decisions." ■

REPRODUCIBILITY

Statisticians issue warning on *P* values

Statement aims to halt missteps in the quest for certainty.

BY MONYA BAKER

Misuse of the *P* value — a common test for judging the strength of scientific evidence — is contributing to the number of research findings that cannot be reproduced, the American Statistical Association (ASA) warned on 8 March. The group has taken the unusual step of issuing principles to guide use of the *P* value, which it says cannot determine whether a hypothesis is true or whether results are important.

This is the first time that the 177-year-old ASA has made explicit recommendations on such a foundational matter, says executive director Ron Wasserstein. The society's members had become increasingly concerned that the *P* value was being misapplied, in ways that cast doubt on statistics generally, he adds.

In its statement, the ASA advises researchers to avoid drawing scientific conclusions or making policy decisions purely on the basis of *P* values (R. L. Wasserstein and N. A. Lazar *Am. Stat.* http://doi.org/bc4d; 2016). Researchers should describe not only the data analyses that produce statistically significant results, the society says, but all statistical tests and choices made in calculations. Otherwise, results may seem falsely robust.

Véronique Kiermer, executive editor of the Public Library of Science journals, says that the ASA's statement lends weight and visibility to longstanding concerns over undue reliance on the *P* value. "It is also very important in that it shows statisticians, as a profession, engaging with the problems in the literature outside of their field," she adds.

P values are commonly used to test (and dismiss) a 'null hypothesis', which generally states that there is no difference between two groups, or that there is no correlation between a pair of characteristics. The smaller the *P* value, the less likely an observed set of values would occur by chance — assuming that the null hypothesis is true. A *P* value of 0.05 or less is generally taken to mean that a finding is statistically significant and warrants publication. But that is not necessarily true, the ASA statement notes.

A *P* value of 0.05 does not mean that there is a 95% chance that a given hypothesis is correct. Instead, it signifies that if the null hypothesis is true, and all other assumptions made are valid, there is a 5% chance of obtaining a result at least as extreme as the one observed. And a *P* value cannot indicate the importance of a finding; for instance, a drug can have a statistically significant effect on patients' blood glucose levels without having a therapeutic effect.

Giovanni Parmigiani, a biostatistician at the Dana Farber Cancer Institute in Boston, Massachusetts, says that misunderstandings about what information a *P* value provides often crop up in textbooks and practice manuals. A course correction is long overdue, he adds. "Surely if this happened twenty years ago, biomedical research could be in a better place now."

FRUSTRATION ABOUNDS

Criticism of the *P* value is nothing new. In 2011, researchers trying to raise awareness about false positives gamed an analysis to reach a statistically significant finding: that listening to music by the Beatles makes undergraduates younger (J. P. Simmons *et al. Psychol. Sci.* **22,** 1359–1366; 2011). More controversially, in 2015, a set of documentary filmmakers published conclusions from a purposely shoddy clinical trial — supported by a robust *P* value — to show that eating chocolate helps people to lose weight. (The article has since been retracted.)

But Simine Vazire, a psychologist at the University of California, Davis, and editor of the journal *Social Psychological and Personality Science*, thinks that the ASA statement could help to convince authors to disclose all of the statistical analyses that they run. "To the extent that people might be sceptical, it helps to have statisticians saying, 'No, you can't interpret *P* values without this information," she says.

More drastic steps, such as a ban on publishing *P* values in articles instituted by at least one journal, could be counter-productive, says Andrew Vickers, a biostatistician at Memorial Sloan Kettering Cancer Center in New York City. He compares attempts to bar the use of *P* values to addressing the risk of automobile accidents by warning people not to drive — a message that many in the target audience would probably ignore. Instead, Vickers says that researchers should be instructed to "treat statistics as a science, and not a recipe".

But a better understanding of the *P* value will not take away the human impulse to use statistics to create an impossible level of certainty, warns Andrew Gelman, a statistician at Columbia University in New York City. "People want something that they can't really get," he says. "They want certainty." ■