

► organization in New York City.

Money is limited, as is the pool of people available for clinical trials, which become larger and more complex as a vaccine or antibody treatment progresses towards the market. “We will need prioritization,” Warren says. “That view needs to be driven by science and financial realities, and the decision process needs to be clear and transparent.”

Another issue facing researchers is how to improve the likelihood that people at risk of HIV infection will take preventive treatments.

Success is not guaranteed. Truvada, a daily pill for preventing HIV infection, has not reduced the number of new HIV cases globally since regulators approved it six years ago. In eastern and southern Africa, for instance, young women rarely take the drug, even though they account for 26% of the region’s new infections. Tian Johnson, founder of the African Alliance for HIV Prevention in Johannesburg, South Africa, says that researchers did not adequately consider how poverty, pregnancy, discrimination and abuse might affect whether young

women at risk are likely to seek out Truvada. “If you disregard the complexity of a woman’s daily life and reality, you put at risk the millions of dollars you invest in developing a product,” Johnson says.

Despite the challenges ahead, the fact that these discussions are happening is an important step forward, says Feinberg. “You can’t keep your head in the sand,” he says. “You need to work ahead and think of ways that we as a research-development community can solve these problems — and they are solvable.” ■

PUBLISHING

Duplicated images could soon be identified by an automated test

Team says technique finds reused images even if they have been rotated and resized.

BY DECLAN BUTLER

Researchers have developed an automated technique that they say can quickly detect duplicate images among hundreds of thousands of papers. If it proves successful, the software could make it easy for editors to screen images before publication — something that currently requires great effort and is done by only a few publications.

Daniel Acuna, a machine-learning researcher at Syracuse University in New York, and his two colleagues described their algorithm on 22 February (D. E. Acuna *et al.* Preprint at bioRxiv <http://dx.doi.org/10.1101/269415>; 2018).

Acuna says he isn’t making the full algorithm public, because that could trigger false allegations. Instead, his team plans to license it to journals and research-integrity offices. Lauran Qualkenbush, director of the Office for Research Integrity at Northwestern University in Chicago, Illinois, and vice-president of the US Association of Research Integrity Officers, says she has discussed the approach with Acuna. “It would be extremely helpful for a research-integrity office,” she says.

In early 2015, Acuna’s team used the algorithm to extract more than 2.6 million images from the 760,000 articles then in the open-access subset of the PubMed database of biomedical literature. These included micrographs of cells and tissues, and gel blots. The algorithm then zoomed in on the most feature-rich areas — where colour and greyscales vary most — to extract a characteristic digital ‘fingerprint’ of each image.

The researchers only compared images

across papers from the same first and corresponding authors, to avoid the computational load of comparing every image against every other one. But the system could pick up potential duplicates even if they had been rotated, resized or had their contrast or colours changed. The trio then manually examined a sample of around 3,750 of the flagged images to judge whether the dupli-

“It would be extremely helpful for a research-integrity office.”

cates were suspicious or potentially fraudulent. On the basis of their results, they predict that 1.5% of the papers in the database would contain suspicious images, and that 0.6% of the papers would contain fraudulent images.

The researchers haven’t been able to benchmark the accuracy of their algorithm, says Hany Farid, a computer scientist at Dartmouth College in Hanover, New Hampshire — because there isn’t a database of known duplicate or non-duplicate scientific images against which they could test the tool.

At present, many journals check some images, but relatively few have automated processes. For instance, *Nature* runs random spot checks on images in submitted manuscripts. (*Nature*’s news team is editorially independent of its journal team.)

To detect image reuse across the literature, publishers would need to create a shared database of all published images against which articles submitted for publication could be compared, says IJsbrand Jan Aalbersberg, head of research integrity at the Dutch publishing giant Elsevier.

There are currently no plans for a

publisher-wide system for image checking, but that is partly because the technologies are not yet mature, says Ed Pentz, executive director of Crossref, a non-profit collaboration of 10,000 publishers. Crossref runs a service that enables publishers to routinely screen submitted manuscripts for plagiarism.

Elsevier says it would support such an initiative for images. Two years ago, the company set up a 3-year, €1-million (US\$1.2-million) partnership with Humboldt University in Berlin to study article mining and to identify research misconduct. On 25 January, the project announced that it intends to create a database of images from retracted publications. Such a data set would provide a bank of test images for researchers developing automated screening of images in publications. ■

CORRECTION

In saying that everyday atomic hearts have equal protons and neutrons, the News story ‘Physicists plan first antimatter road trip’ (*Nature* **554**, 412–413; 2018) didn’t take account of the fact that some elements, such as hydrogen and lithium, have uneven numbers of protons in their most abundant form.

The News Feature ‘The entangled web’ (*Nature* **554**, 289–292; 2018) misstated the leadership of the Dutch demonstration quantum network. The project is co-led by Ronald Hanson and Stephanie Wehner of Delft University of Technology in the Netherlands and Erwin van Zwet at the Dutch research organization TNO in The Hague.