THE ARISING

OCTOBER 31, 2026 · JOHN GRIFFIN, MD

The difference between a promising idea and something that works



Original conceptual illustration · Room to breathe

A promising idea becomes more useful when we ask exactly what happened in the test. These ten articles follow clinical benefits, disappointing comparisons, diagnostic tools and new biological discoveries. The aim is to keep the human stakes visible while giving uncertainty its proper place.

Ten full written articles · Alive. Awake. Aware.

In this edition

01 Fewer terrifying COPD flare-ups: two trials put an inflammatory target to a meaningful test

02 The liver test that could spare a biopsy—but only for the question it answers well

03 After lung-cancer chemoradiation, a second drug raised the harm without proving the extra benefit

04 A 17-point absolute difference in leukemia relapse deserves attention—and close transplant scrutiny

05 The AI inbox paradox: writing a reply faster does not necessarily give clinicians their evenings back

06 The evidence review can fail before the statistics begin

07 A psychedelic trial's most important result may be the outcome it missed

08 Sophisticated exercise personalization changed the laboratory profile—not the functional result

09 The sleep study that makes thirty seconds look like an arbitrary boundary

10 A virus found in the human gut turns out to infect a bacterium—not necessarily the person

About this copy

Recovered and prepared October 31, 2026, from the dated source issue or existing reader selection. The issue date is preserved; this PDF does not establish a publication date. Source qualifications are retained. This packaging pass is not a fresh independent audit of every study.

Use NEWSLETTER.html for copying into Substack and upload the image files separately. The longer original written-article reference, where supplied, is separate from this reader selection.

01 / Fewer terrifying COPD flare-ups: two trials put an inflammatory target to a meaningful test

Source: New England Journal of Medicine, **first published online September 8, 2026; October 29 journal issue**; OBERON and TITANIA phase 3 randomized, double-blind trials; primary article · DOI 10.1056/NEJMoa2606998. **Evidence:** high-quality paired randomized evidence. **Clinical state:** specialist appraisal now; investigational drug and indication require contemporaneous regulatory review.

An exacerbation of chronic obstructive pulmonary disease is not an abstract disease marker. It may mean another course of treatment, an ambulance ride, or the loss of confidence to leave home. Both OBERON and TITANIA tested whether inhibiting interleukin-33 could prevent these episodes in people who continued standard inhaled maintenance treatment despite previous exacerbations. Participants received 300 mg subcutaneous tozorakimab or placebo every four weeks for 52 weeks.

In former smokers, annualized moderate or severe exacerbations fell from **1.90 to 1.34** in OBERON (rate ratio 0.71, 95% CI 0.57–0.88) and from **2.07 to 1.37** in TITANIA (rate ratio 0.66, 95% CI 0.55–0.80). Similar findings appeared in the full current- and former-smoker populations. The overall-analysis groups reported about 0.59 fewer exacerbations per patient-year in each trial, an understandable gain even though rates are population averages rather than promised individual events. Reported overall adverse-event percentages were 70.4% versus 77.2% and 80.1% versus 79.8%, respectively.

What this changes: There is convincing randomized evidence for fewer exacerbations in the enrolled populations, with replication across two trials. Long-term survival, comparative effectiveness against other biologics, real-world access, and cost-effectiveness remain separate questions. Do not extrapolate efficacy to all COPD phenotypes or assume it is already routinely available.

Source edition: October 31, 2026 · AR-20261031-READER-01 · existing reader feature

02 / The liver test that could spare a biopsy—but only for the question it answers well

Source: The Lancet, issue dated October 31, 2026; LITMUS prospective multicenter diagnostic-accuracy study, DOI 10.1016/S0140-6736(26)01145-1, PubMed. **Evidence:** substantial prospective diagnostic comparison, not an outcome trial. **Clinical state:** promising for diagnostic-pathway design; assay selection/indication require validation.

Metabolic liver disease often progresses silently until its consequences are serious. Now that treatments are advancing, the immediate obstacle is identifying the right people without sending everyone for an invasive biopsy. The LITMUS investigators enrolled **1,590 adults with biopsy-confirmed disease** across 13 countries, centrally reviewed the histology, and compared a dozen individual biomarkers with 18 panels.

The most discriminating assay depended on what clinicians were trying to diagnose. For **advanced fibrosis**, SomaSignal reached an area under the receiver-operating curve (**AUC 0.92**, 95% CI 0.90–0.94); FibroMeter combined with vibration-controlled elastography reached **0.87**. For **at-risk steatohepatitis**—inflammation plus fibrosis stage 2 or higher—SomaSignal reached **0.83** (0.80–0.85). For cirrhosis, FibroMeter-VCTE reached **0.94**. Those are promising ranking measures, not the chance that a particular positive test is correct; predictive values depend on prevalence, selected thresholds and the next decision.

What this changes: We may be approaching more effective tiered pathways: cheap initial tests, focused multiparameter tests, biopsy when uncertainty still matters. But the biopsy-enriched cohort and comparison conditions can differ from an ordinary community screening population, and no test's AUC by itself proves fewer cirrhosis complications or better survival.

Source edition: October 31, 2026 · AR-20261031-READER-02 · existing reader feature

03 / After lung-cancer chemoradiation, a second drug raised the harm without proving the extra benefit

Source: JAMA Oncology, **published online October 29**, phase 3 R-ALPS double-blind randomized trial, 553 patients at 42 centers in China. Primary study · DOI 10.1001/jamaoncol.2026.3918. **Evidence:** high-quality randomized study for the studied consolidation strategies. **Clinical state:** important oncology appraisal, not a direct comparison with all global standards of care.

More targeted treatment is tempting when a cancer remains likely to recur. R-ALPS tested the PD-L1 antibody benmelstobart after chemoradiation for unresectable stage III non-small-cell lung cancer, with or without the antiangiogenic drug anlotinib. The first stage compared both regimens against placebo; the additional stage strengthened the direct comparison between active regimens. This structure matters when interpreting the headline numbers.

In the first randomized stage, median independently assessed progression-free survival was **15.1 months** for the combination, **9.7** for benmelstobart alone and **4.2** for placebo; both active regimens beat placebo. But when both randomized stages were pooled for the direct combination-versus-monotherapy question, median PFS was **17.4 versus 11.2 months**, hazard ratio **0.82** (95% CI 0.63–1.08), **p=0.12**. The extra drug had **not established superiority**. Grade 3–5 treatment-related adverse events were **48.8%** with the combination, **29.4%** with benmelstobart alone and **19.7%** with placebo; grade 5 treatment-related events were reported in **7.2%, 1.4% and 1.5%**, respectively.

What this changes: A numerical median difference can coexist with uncertainty substantial enough that the randomized superiority claim fails. The excess serious and fatal toxicity is a patient-important counterweight. Existing treatment standards, cancer molecular subgroup, geography and subsequent therapies must be considered before changing care.

Source edition: October 31, 2026 · AR-20261031-READER-03 · existing reader feature

04 / A 17-point absolute difference in leukemia relapse deserves attention—and close transplant scrutiny

Source: The Lancet Haematology, online/indexed in the October 29–31 retrieval window, November issue designation; eight-center Chinese phase 3 randomized trial in KMT2A-rearranged leukemia. PubMed · DOI 10.1016/S2352-3026(26)00250-400250-4). **Evidence:** randomized phase 3, relatively small molecularly selected population. **Clinical state:** transplant-specialist evaluation.

For a patient with a high-risk leukemia, transplant conditioning is an effort to destroy disease before donated immune cells establish themselves. In this study, **130 patients** were randomized to conventional busulfan-cyclophosphamide conditioning or the same regimen preceded by decitabine. Median follow-up was **34.7 months**. At two years, the cumulative incidence of relapse was **12%** with decitabine conditioning versus **29%** in the comparator group: an **absolute difference of 17 percentage points**, with hazard ratio **0.41** (95% CI 0.20–0.87; p=0.017).

There was no signal in the reported abstract that those gains required a major increase in some severe adverse-event categories: grade 3–5 infections were **42% versus 38%**, acute graft-versus-host disease **35% versus 34%**, and reported deaths attributed to adverse events **6% versus 8%**. Yet the confidence interval around the relapse effect is wide, and two-year survival, nonrelapse mortality, transplant toxicity in other settings and longer follow-up remain important to examine in the full publication.

What this changes: The trial provides a potentially meaningful modification to a specific transplant regimen in a molecularly defined leukemia, not a general longevity or anticancer treatment. Its rationale should be assessed alongside real survival and quality-of-life outcomes.

Source edition: October 31, 2026 · AR-20261031-READER-04 · existing reader feature

05 / The AI inbox paradox: writing a reply faster does not necessarily give clinicians their evenings back

Source: JAMA Network Open, **published online October 30, 2026**, observational EHR implementation study, original paper · DOI 10.1001/jamanetworkopen.2026.41364. **Evidence:** large real-world observational comparison; association, not randomized efficacy. **Clinical state:** operations measurement, not autonomous communication.

It sounds obvious that if a language model can prepare a good portal reply, the clinician should spend less time responding. The actual experience of a large health system was more complicated. Investigators analyzed **983,154 patient-portal message responses**, involving **1,542 clinicians and 272,091 patients**, to examine an EHR-integrated AI draft tool. Adoption was associated with an **8.6% increase in turnaround time**. For messages on which clinicians actually **used** the tool, turnaround time was associated with a **29.5% reduction**. The headline gains did **not** translate to an overall reduction in after-hours responding.

Those results are not contradictory. Clinicians may select AI for some messages, may use extra time reviewing others, and may respond to more messages because the work becomes easier to initiate. One cannot infer the causal effect of using the tool from comparing selected messages with other messages in an observational deployment.

What this changes: A productivity demonstration needs to track more than draft-generation speed. It needs elapsed patient wait, time spent editing, message count, response quality, missed clinical issues, after-hours work, and whether the task has moved to another staff member.

Source edition: October 31, 2026 · AR-20261031-READER-05 · existing reader feature

06 / The evidence review can fail before the statistics begin

Source: JAMA Network Open, **online October 29, 2026**; systematic audit of Cochrane diagnostic meta-analyses, article and PubMed, DOI 10.1001/jamanetworkopen.2026.40850. **Evidence:** methodology audit, not a new clinical intervention.

Our whole research system rests on a simple promise: when we report the effect of a treatment or accuracy of a test, the underlying numbers still mean what the original investigators measured. A new systematic audit reveals how easily that promise can break. Researchers reviewed **46 systematic reviews**, **879 diagnostic comparisons**, and the primary data from **6,173 studies**. Data extraction could not be reproduced for **338 studies (5.5%)**; errors affected **137 quantitative diagnostic meta-analyses** and **34 of 46 reviews (73.9%)**.

Not every error changed the bottom line. Among the 137 reproduced meta-analyses, substantial changes occurred in about **0.7%** for sensitivity, **2.9%** for AUC and **3.6%** for diagnostic odds ratios according to the paper's thresholds. But some did change conclusions, and the principal error types—wrong numbers, incorrect arithmetic or derived values, mismatched groups, and misread reference standards—are exactly the sort of problems that a fluent summary can conceal.

What this changes: A review's evidence tier is not a guarantee that the data were faithfully lifted from a trial. The underlying source tables, participant counts and comparator direction have to survive an independent check.

Source edition: October 31, 2026 · AR-20261031-READER-06 · existing reader feature

07 / A psychedelic trial's most important result may be the outcome it missed

Source: Journal of Affective Disorders, published/indexed **October 30, 2026**; randomized double-blind pilot in treatment-resistant depression, original record, DOI 10.1016/j.jad.2026.122676. **Evidence:** small randomized pilot. **Clinical state:** investigational, supervised clinical research only.

Anhedonia—the loss of the ability to enjoy ordinary life—is often one of the hardest aspects of major depression to treat. Investigators randomly assigned **20 patients** to psilocybin **25 mg or 1 mg**, with psychological support. The prespecified primary endpoint was improvement on a measure of anhedonia after **one week**. The between-group difference was **11.2 points**, but its 95% confidence interval ran from **−1.0 to 23.3**, and **p=0.073**. In other words, the study **did not establish the primary superiority claim**.

Secondary results were more encouraging. At week eight, the anhedonia difference on the same scale was **14.7 points** (95% CI 2.5–26.8); overall depression severity also favored the higher dose at weeks one and eight. Those findings merit replication, but the small pilot, multiple endpoints, psychologically active treatment and potential functional unblinding limit what can be concluded about durability and generalizability.

What this changes: The result belongs in a serious mental-health edition because it identifies both a plausible direction for research and a failed primary test. Neither a promising secondary measure nor an early clinical signal makes unsupervised psychedelic use appropriate.

Source edition: October 31, 2026 · AR-20261031-READER-07 · existing reader feature

08 / Sophisticated exercise personalization changed the laboratory profile—not the functional result

Source: Scandinavian Journal of Medicine & Science in Sports, newly available/indexed October 29–31; November issue listing, 12-week randomized study, record · DOI 10.1111/sms.70392. **Evidence:** small randomized intervention, functional outcomes measured. **Clinical state:** standard safe training principles still matter; new profiling rule unproven.

It is tempting to think the more precisely a training program diagnoses a person's weakness, the better the resulting exercise prescription must be. This study tested the assumption directly. Among adults aged **65 or older**, investigators measured leg-press force and velocity profiles and randomized **57 starters with a velocity deficit** to training loads either matched or deliberately mismatched to that deficit over **12 weeks**.

The intervention did what the underlying mechanics predicted: force capacity improved more in the mismatched-load group (**p=0.040**), while maximal movement velocity improved more in the matched-load group (**p=0.009**). But changes in **overall muscle power and functional performance did not significantly differ** between groups (reported p≥0.065). The experiment was too small to conclude the strategies are equivalent, but it does show that moving the mechanism one wishes to move does not necessarily create a measurable advantage in daily function.

What this changes: The clinically important targets remain walking, balance, rising from a chair, climbing stairs, and maintaining independence. Individualization is valuable when evidence connects the extra assessment and effort to better outcomes; it should not be assumed beneficial from a more intricate dashboard.

Source edition: October 31, 2026 · AR-20261031-READER-08 · existing reader feature

09 / The sleep study that makes thirty seconds look like an arbitrary boundary

Source: npj Digital Medicine, **October 30, 2026**, research/validation paper, PubMed · DOI 10.1038/s41746-026-03406-7. **Evidence:** multicenter model development/validation, no prospective clinical treatment-outcome evidence. **Clinical state:** worthwhile blinded technical evaluation.

Traditional polysomnography compresses the complexity of sleep into standard **30-second scored epochs**. That convention is practical, not a claim that the brain changes states only twice a minute. The AnySleep team developed a channel-flexible deep learning system using **more than 20,000 overnight recordings**, representing **over 200,000 hours** across **28 datasets**. It can score sleep from varying EEG and eye-movement channel arrangements, and at finer temporal resolutions.

The authors report performance at least comparable with established approaches for conventional sleep staging; sub-30-second inference identified brief wake intrusions resembling arousals and improved prediction of selected sleep-related conditions, including apnea, narcolepsy type 1 and insomnia, compared with 30-second staging. A major strength is deliberate testing across heterogeneous recordings and channel setups. However, algorithm-level agreement and predictive modeling do not establish fewer missed diagnoses, better therapy choices, lower costs or improved sleep in real patients.

What this changes: There may be clinically important information between the usual categories, especially when connecting sleep fragmentation to daytime function. But home wearable readings and formal EEG/EOG-based models are not interchangeable, and actual deployment will depend on artifacts, populations, reference scoring and external testing.

Source edition: October 31, 2026 · AR-20261031-READER-09 · existing reader feature

10 / A virus found in the human gut turns out to infect a bacterium—not necessarily the person

Source: Nature Microbiology, **published online October 29, 2026**; WashU-led experimental microbiology research, full primary article · DOI 10.1038/s41564-026-02523-z. **Evidence:** culture plus germ-free mouse and molecular localization, not a human clinical outcome. **Clinical state:** fundamental discovery; no new treatment or infection diagnosis.

A virus has been detected in human and animal stool for decades. It was therefore tempting to describe it as a virus infecting those hosts. The problem was that nobody had demonstrated the infection of human cells or reliably grown the agent in a culture system. Researchers working with picobirnaviruses—first reported in 1988—finally followed a different possibility: the virus might be traveling through the human gut because it infects a resident bacterium.

They propagated two human stool-associated viral samples in **germ-free mice after microbiota transplantation**, and grew one in an **anaerobic bacterial community**. Antibiotics suppressed its growth, RNA fluorescence imaging localized viral RNA inside bacterial cells, and cell-enrichment sequencing identified a candidate bacterial host, Candidatus Anaerotignum faecalis. This establishes at least one member of the picobirnavirus family as an **RNA bacteriophage**, expanding the recognized major clades of RNA phages from two to three. It does **not** prove every member of the family targets bacteria rather than animals.

What this changes: Where a sequence is found is not necessarily what it infects. Disease associations in stool metagenomics need ecological and causal interpretation, not just a name supplied by a genome database.

Source edition: October 31, 2026 · AR-20261031-READER-10 · existing reader feature

---

Alive. Awake. Aware.

GO DEEPER

Keep following the question.

The daily issue is a beginning. Choose a reading companion, explore a connection, or return to the wider Arising universe.

Reading the evidence →

Sleep & rhythm →

Browse all public booklets →

Explore the collections →

Follow connections in the Atlas →

Visit Arising Health →

Meet the book →

Meet The Long Game →

The public Atlas connects questions across topics. The book and game pages introduce projects in development.

Alive. Awake. Aware.