THE ARISING

What the Clock Can Tell Us

BIOLOGICAL AGING CLOCKS



NOVEMBER 2026

What the Clock Can Tell Us

Biological age, changing numbers, and the life behind the test

The Arising · Biological-age clocks · November 2026 · John Griffin, MD

Your birthday does not change when a laboratory sends you a report. Your sense of what is possible might.

A biological-age result can arrive with an irresistible comparison: the years you have lived beside a number that seems to describe how old you really are. A younger result may feel like encouragement. An older one may feel like a warning. If a second test looks better after a new exercise plan, diet or supplement, the explanation can seem obvious: something worked.

But what, exactly, changed?

That is where the science becomes useful. A laboratory can measure something carefully without knowing whether a change means better health. Two different tests can disagree without either being a simple mistake. And a number can become more precise long before it becomes a reliable guide to what someone should do next.

Two studies help us explore those distinctions. One looked at how consistently 18 epigenetic clocks behaved across repeated measurements. The other revisited intervention studies using additional clocks. Neither was a trial proving that a treatment makes people live longer. Both ask an important question for anyone interested in healthy aging: when does a changing number deserve our trust? [1,2]

What an aging clock actually measures

Chronological age is the time since birth. A biological-age clock is a model that combines measurements to estimate something related to aging. The word “clock” makes that sound like reading a watch. In practice, it is closer to using a carefully constructed pattern-recognition tool.

The studies in this issue concern epigenetic clocks. These use patterns of DNA methylation, small chemical marks attached to DNA, rather than simply reading the sequence of its genetic letters. The model combines many measurements into a score. Some scores are expressed as years; others describe a rate or another aging-related pattern. [1,2]

That difference is not a detail to skip. A model built to estimate calendar age is doing a different job from one built around patterns associated with death risk or the pace of aging. Giving all of them the same label can make their results look more interchangeable than they are.

Imagine three maps of the same city. One shows streets, another public transport and another elevation. They describe the same place but answer different questions. Comparing aging clocks requires a similar first step: what was each one designed to tell us?

A score related to the risk of death in a group also does not tell an individual how many years they have left. Nor does lowering that score establish that a treatment will extend life. Those claims need their own evidence.

One sample twice, or one person twice?

Reliability sounds like a single quality. The study by Sehgal and colleagues shows why it helps to separate two kinds. [2]

Technical reliability asks whether the laboratory gets a similar result when it measures the same sample again. If two portions of one blood sample produce very different answers, interpretation becomes difficult before we even ask what is happening inside the person.

Biological reliability asks what happens when researchers collect another sample from the same person over a short interval. The laboratory procedure could be consistent while the person's measurements vary across collections.

Think of a bathroom scale. Repeatedly weighing the same fixed object tests the scale. Weighing yourself on different occasions involves the scale and changes in what is being weighed. The analogy is limited, but the distinction is useful: a reliable instrument does not make every change in its reading a lasting change in health.

Sehgal and colleagues used the TranslAGE research platform to compare 18 epigenetic clocks across existing datasets. Most performed well when the same samples were measured repeatedly under standard conditions. Some were more sensitive to technical details, including where a sample was placed on a testing slide or how DNA was extracted. Certain principal-component-based clocks were more robust across technical conditions. [2]

A principal component is a mathematical summary of patterns shared across many measurements. In these clocks, such summaries are used to make the score less vulnerable to noise in individual measurements. That can improve consistency; it does not, by itself, show that a score measures a treatment benefit. [1,2]

What changed when the person was sampled again?

Across short-interval collections from the same people, many clocks were less consistent. Some datasets included collections around meals, a stress task, altitude exposure or air pollution, as well as samples taken on different days. [2]

These were not large experiments proving that lunch or a stressful moment makes a person biologically older. The meal and stress analyses came from the same dataset of 34 young adults; the altitude analysis included 21 people, and the pollution analysis included 16. The event-related datasets lacked a comparable unexposed group sampled alongside them. The findings show variation across these collections, not a clean separation of what caused it. [2]

This is an important boundary for a headline. “A clock reading varied around a stressful event” is different from “stress aged these people by several years.” The first describes a measurement. The second adds a much larger claim about the body and the cause.

In the study, clocks with better biological reliability tended to give more consistent results when researchers examined links with thinking abilities and responses in intervention datasets. Technical reliability alone did not predict biological reliability. A laboratory can therefore perform a repeat measurement well while the score still presents challenges as a personal progress tracker. [2]

What this does not tell us about a test you can buy

The technical comparisons were conducted within laboratories and used particular research platforms, including older 450K and EPIC v1 measurement systems. They were not a comprehensive comparison of current consumer services, different laboratories or every way a company might process a sample. [2]

The study does not identify one universal commercial winner. It also does not show that all aging-clock research is unhelpful. It asks a more specific question: how much variation must researchers understand before interpreting a change as the effect of an intervention?

When different clocks tell different stories

The second paper, by Borrus and colleagues, examined what happens when findings reported with one clock are tested against others. [1]

The researchers revisited three intervention datasets. When they applied 11 additional clocks, those clocks did not corroborate the statistically significant decreases reported in the earlier single-clock analyses. “Statistically significant” means a result passed a chosen statistical threshold under the analysis's assumptions. It does not mean a person became meaningfully healthier, and failing that threshold does not prove that nothing changed. [1]

The result is a reason to ask better questions, not to declare every intervention a failure. Different clocks were built for different targets. An intervention might affect one aspect of biology more than another. Limited sample sizes and measurement variability can also make a real change harder to detect.

A noisy clock is not automatically more likely to produce a false positive, a result that appears to show an effect when there is none. Poor reliability can instead reduce sensitivity: the ability to pick up a real

change against the variation. The concern becomes different when someone examines many clocks and tells the story only through the one that looks most favorable. Choosing the result after seeing the answers can make the overall claim much less dependable. [1]

That is why the question “Which clock improved?” needs a companion: “Which clocks were chosen beforehand, and what happened to the others?”

Three ways the authors proposed strengthening confidence

The paper proposed looking for changes in reliable principal-component-based clocks, in clocks trained around mortality-related patterns or the pace of aging, and across multiple clocks moving in a consistent direction. Simulations and analyses of repeated measurements helped the researchers explore why reliability matters. [1]

These are proposed ways to interpret research. They are not a clinically validated checklist that turns a younger score into proof of rejuvenation. Agreement among several models would be more informative than a selectively chosen result, but it would still leave another question: did people's lives improve?

Better memory, greater strength, fewer illnesses and longer survival are outcomes in people. A clock change may eventually help researchers understand or predict some of those outcomes. It cannot simply stand in for them because it is expressed in years.

The people and interests behind the research

The two papers share investigators, so they should not be presented as independent replications by unrelated teams. Their questions complement each other, but the overlap matters when judging the breadth of support.

The authors also disclose consulting relationships with companies involved in biological-age testing. Authors of the reliability paper are co-inventors on a patent application concerning SystemsAge. These interests do not decide whether the findings are right or wrong. They are part of the context readers should have, particularly when a paper discusses which measurements are most useful. [1,2]

The strongest next step is research that tests the proposed improvements across additional teams, populations, laboratories and meaningful health outcomes. Better measurement is an opportunity worth pursuing, not a finished promise.

Give the score a useful job

Before buying a test, it helps to decide what question you want it to answer.

Curiosity is a legitimate reason to explore a new measurement. It is different from using the result to choose a medicine, judge whether a supplement works or decide that established care is no longer needed. The evidence required should grow with the importance of the decision.

Useful questions for a testing service or clinician include:

• **What is this score designed to estimate?** Calendar age, an aging-related risk pattern and a pace-of-aging measure are different targets.

• **How much do repeated results vary?** Ask about samples collected again from the same person, not only repeat testing of one sample.

• **What would count as a meaningful change?** A small numerical difference needs context about expected variation. Neither study supplies a universal cutoff for every test.

• **Was this test studied in people like me?** The population and sampling conditions affect how far a result can travel.

• **What would we do differently?** Ask what evidence supports that action and whether it improves something that matters beyond the score.

• **What is the full cost of the information?** Consider repeated testing, time and worry alongside the purchase price.

If no useful action follows, the result may still be interesting. It is less convincing as a reason to add expensive treatments or repeatedly change a health plan.

Three everyday situations

A person sees a younger score after starting strength training. They may also notice that carrying groceries feels easier. Those are two observations, not proof that the training caused the clock change. The improvement in daily ability has value in its own right. The clock result needs its own interpretation, including sampling conditions, expected variation and what the model measures.

A supplement advertisement highlights one favorable clock. The reader asks how many clocks were examined, whether the highlighted measure was chosen before the study, and whether the participants experienced a meaningful health benefit. A persuasive graph cannot answer those questions by itself. A study that reports every planned result is easier to judge than a story built around its best-looking number.

Someone receives an unexpectedly older result. It is understandable to feel unsettled. The result is not a diagnosis or a forecast of how many years remain. The next useful step is to understand the test and its limits, and discuss any genuine symptoms or health concerns through ordinary clinical care. Buying a second test immediately does not automatically resolve the first one's meaning.

These examples are illustrative, not reports of actual patients. They show how an aging score can prompt curiosity without being allowed to make every decision.

A future worth measuring

The ambition behind aging clocks is compelling. If a trustworthy measurement could help identify useful interventions sooner, it could make research more efficient and bring better care closer.

Getting there requires several things to be true at once. The measurement must be dependable. The change must mean what researchers think it means. And using it must help answer a question that matters for health or daily life.

The science is not diminished by asking those questions. It becomes more useful.

A good clock may help us understand the years. The life we hope to live remains the reason to measure them.

Sources and further reading

[1] Daniel S. Borrus, Raghav Sehgal, Jenel Fraij Armstrong and colleagues. When to Trust Epigenetic Clocks: Avoiding False Positives in Aging Interventions. Aging Cell. 2026;25(10):e70756. Primary publication record. Full text.

[2] Raghav Sehgal, Daniel S. Borrus, John Gonzalez and colleagues. Biological Versus Technical Reliability of Epigenetic Clocks and Implications for Disease Prognosis and Intervention Response. Aging Cell. 2026;25(8):e70635. Primary publication record. Full text.

Continue exploring

Read HRV & recovery, Wearables and Fitness, or explore the magazine collection.