The Illusion of Optimization

— Dr. Zain Hakeem

Abstract

THE DESIRE TO OPTIMIZE HEALTH is rational. Modern medicine has improved diagnosis, extended lifespan, and reduced suffering. If treating disease works, it is reasonable to assume that intervening earlier, measuring more precisely, and adjusting more variables should work even better.

We all know that standard medical practice feels incomplete. Visits are short. Explanations are thin. Recommendations are tied to numbers that are rarely unpacked.

It is completely natural, then, to want to take the things that standard medicine already does, including blood work, scans, and screening tests, and simply do *more* of it. The impulse is so natural that I cannot truly blame anyone for not understanding its flaws intuitively.

Understanding the counter-intuitive truth that doing more and better tests must inevitably fail as a form of optimization, requires a significantly deeper level of study. That level of understanding unfortunately goes far beyond most doctors, let alone the general public.

After all, standard medicine drifted toward metric-chasing long before influencer culture accelerated it. The medical profession has become beholden to the interests of insurance companies, exacerbating the challenge of caring for patients through lack of time, lack of explanation of reasoning, and arrogance in their emphasis on faithful adherence to national guidelines. Insurance companies reinforce this behavior by treating improvement in metrics as "quality", regardless of the real science, or real patient outcomes.

Surface problems are addressed, and deeper concerns dismissed. The bare minimum of function is considered the gold standard, and higher health aspirations are brushed aside.

Into that gap has grown a rapidly expanding industry of "health optimization", emerging at the intersection of digital influence, consumer biotechnology, and preventive medicine. This movement promotes the aggressive use of laboratory testing, "advanced" imaging, biometric tracking, and personalized protocols, with the stated aim of extending lifespan and enhancing long-term health.

While these concepts are promoted as proactive, cutting-edge, and scientifically superior, their underlying logic relies on a style of reasoning called "quantified quackery": any surrogate biomarker with an observational association and a plausible mechanism is treated as causation. "Improving" such markers is treated as definitive proof of improvement in outcomes, without any direct evidence that such is the case. The label "advanced" replaces any need for such outdated things as scientific evidence of clinical benefit.

The extensive dissection of biochemical mechanisms gives a veneer of scientific expertise, and even as they announce "correlation isn't causation", influencers nonetheless continue to promote correlative links and associations between biomarkers and risks.

In doing this, influencers merely extended the flawed playbook of modern medical practice, removing the few remaining guardrails smuggled into guidelines by thoughtful scientists, despite the meddling of big pharma.

In their correct perception of big pharma's influence, and of medicine's slow adoption of new information, influencers thought they could step in and make medical recommendations, without needing any of the experience that grants a license to practice, and without assuming any accountability for the consequences of their recommendations.

Influencers implicitly promote their information as "more expert" than your physician, when in fact they are demonstrating nothing more than a Dunning-Kruger-style confidence, arising from the amateur's failure to understand the true complexity of biology and treatment, including the extensive history of medical failures resulting from exactly their brand of "reasoning".

This document will expose exactly that failure by critically examining the scientific foundations of the optimization paradigm. Drawing on established principles from clinical epidemiology, including the limits of surrogate endpoints, measurement variability, regression to the mean, Bayesian reasoning, reference range construction, and overdiagnosis, this paper will dismantle the underlying model that turns biological storytelling into overconfident medical advice, whether licensed or unlicensed.

Historical examples from medicine demonstrate that modifying biological markers does not reliably translate into reductions in morbidity or mortality, and that increased testing in low-risk populations predictably generates false positives, overdiagnosis, and net harm.

This document will explain why the methods influencers promote can only ever provide *the illusion of optimization*, and why chasing these fairy-lights is a distraction from the true opportunity that lies in another direction.

By sharply distinguishing between correlation and causation, measurement and meaning, detection and benefit, this document seeks to clarify what constitutes reliable proof of improved health and longevity. Only by maintaining rigorous outcome-based criteria can medicine avoid repeating historical errors and prevent the commercialization of statistical noise as scientific progress.

The question that exposes the illusion is simple: what are your requirements for belief and disbelief? What are your standards of evidence that justify your decision to believe in a medical recommendation? For most people, without specific training, the answer ends up being "vibes". Unfortunately, that can lead us astray.

This document will deconstruct the ways in which "vibes"-based belief formation has led the health optimization movement astray. In my no-cost, biweekly livestreamed lecture, "An Evening with Dr.Z", I help you develop better standards of belief, and show what optimizations remain when those standards are enforced.

Marker / Outcome Decoupling

Improving Metrics Does Not Imply Improving Outcomes

THIS SECTION IS THE MOST IMPORTANT concept in this document. One of the key errors committed by health influencers (and frequently by standard medicine as well) is to confuse a *metric* with the disease process it represents. They assume that any intervention that modifies a metric will also improve the clinical outcome.

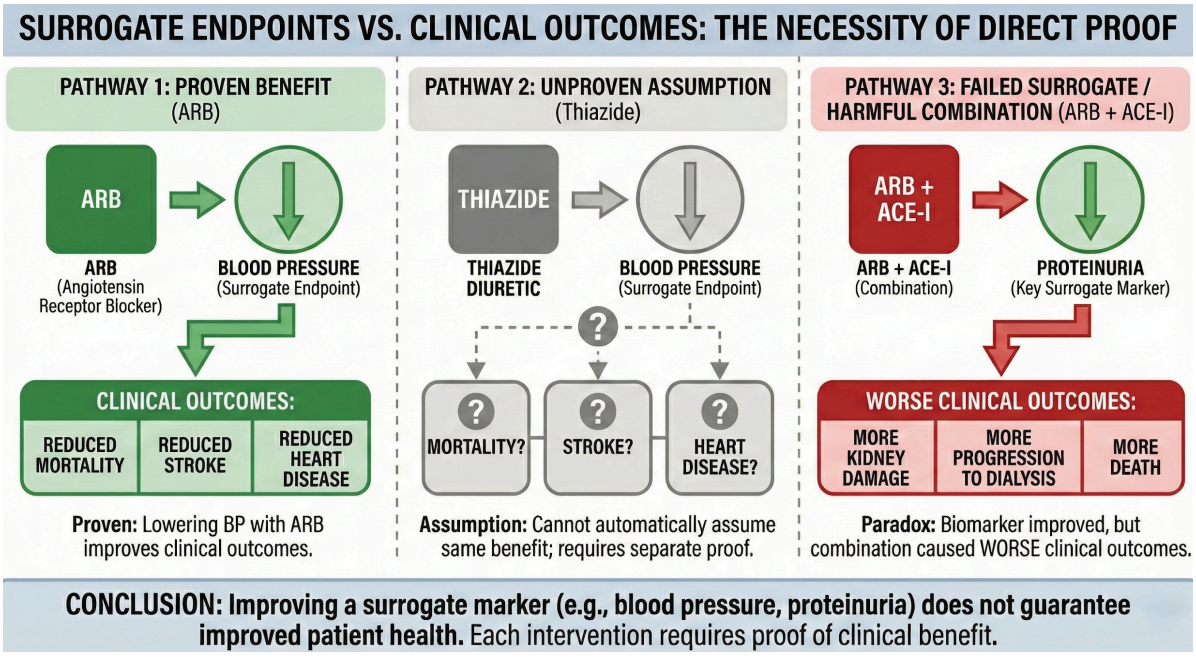
In fact, medical history is a graveyard of interventions that successfully modified a key factor believed to be important to a disease, but resulted in no actual benefit (or even caused harm) to the patient. Changing a metric does not imply, even slightly, a change in the outcome[1]. The improvement in outcome *must* be proven separately, in its own randomized controlled trial (RCT)[2], [3].

It is not enough to prove, for instance, that an angiotensin receptor blocker (ARB) medication lowers blood pressure. You must separately prove that by lowering blood pressure, the ARB thereby reduces mortality, stroke, and heart disease. Once you have proven that, you cannot automatically assume that a thiazide medication, which also lowers blood pressure, will provide the same benefits. You must prove that separately as well.

> **"Changing a metric does not imply, even slightly, a change in the outcome ... improvement in outcome *must* be proven separately, in its own randomized controlled trial (RCT)."**

Each intervention, and each combination of interventions, requires its own proof. For example, both ARBs and ACE inhibitors (ACE-I) lower blood pressure and reduce heart attacks, strokes, and kidney failure. Specifically, both reduce protein leakage into the urine (called "proteinuria"), which is a key marker of kidney failure. Logically, one might assume that taking both would be even better.

Indeed, taking both an ARB and an ACE-I reduced proteinuria more effectively than either one alone. Yet, it *also* caused more kidney damage, more progression to dialysis, and more death[4], [5]. The marker improved, which theoretically should have improved outcomes, but in reality, the combination caused worse outcomes while continuing to improve the biomarker.



How you change a metric is just as important as the change itself. Merely proving that an intervention affects a metric is insufficient without also proving a clinically relevant outcome.

To illustrate how often logical reasoning fails in medicine, I want to go through several examples that lie in that medical graveyard. Let's start with the COURAGE and ISCHEMIA trials. One might naturally assume that because blocked arteries cause heart attacks, opening partial blockages with stents, before a heart attack has occurred, would naturally reduce subsequent heart attacks and deaths[6].

It is a very clean, logical reasoning pattern: you see the narrowing on an angiogram, you know a heart attack is caused by a *total* blockage, and you conclude that fixing a partial blockage will prevent the catastrophe. Makes sense.

> "The logic was 'obvious,' but failed when tested."

But, when tested, putting stents into these blockages had no effect on reducing heart attacks or death[7]. The logic was "obvious," but failed when tested.

Another example is Hormone Replacement Therapy (HRT) for menopause. For years, it was considered "obvious" that giving estrogen to menopausal women would lower cardiovascular risk. Women had less cardiovascular disease than men, observational data indicated those taking HRT had fewer heart attacks than those who did not, and HRT was known to improve lipid profiles[8], [9]. It was a clean logical chain based on understood mechanisms and strong observational evidence[10]. It was so "obvious" that it was written into guidelines at the time[11], [12]. Unfortunately, when an RCT was finally done against a placebo, HRT was found to actually *increase* the cardiovascular risk, completely counter to the expectations of observational evidence and mechanistic reasoning[13].

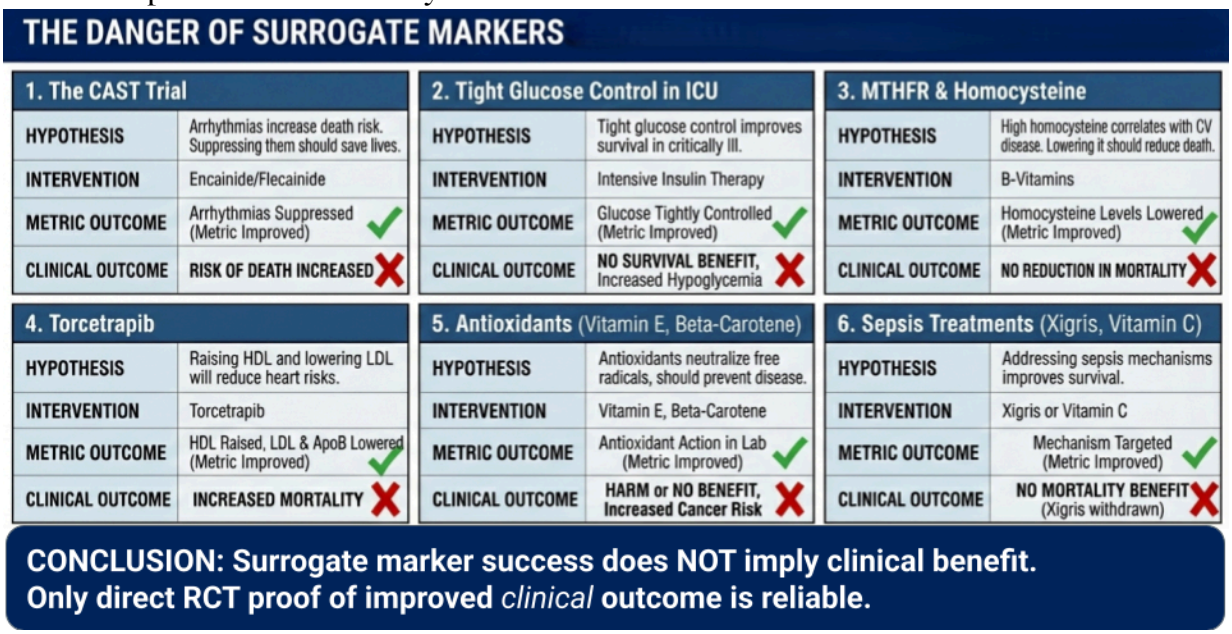
This pattern is far from an exception. One of the most dramatic examples is the CAST trial. It was observed that after a heart attack, people with more arrhythmias were at greater risk of death. It was presumed that suppressing those arrhythmias with medication would reduce that risk. We gave them drugs like Encainide and Flecainide, and while the arrhythmias were effectively suppressed, the risk of death actually increased. The metric was improved, but the outcome was worse[14], [15], [16].

We see this in the ICU, where tight glucose control was hypothesized to improve survival. Instead, it increased mortality[17].

And we see it again with MTHFR genes and homocysteine, which has a compelling biochemical story, and is very strongly correlated with cardiovascular disease and mortality. B-vitamins lower the levels effectively but do nothing to reduce mortality when actually tested[18], [19].

I could go on and on. Torcetrapib was effective at raising HDL and lowering LDL and ApoB, yet it increased mortality[20], [21]. Vitamin E and Beta-Carotene are antioxidants, and while they perform antioxidant actions in the lab, they result in harm or no benefit when given as supplements[22], [23], [24], [25], [26], [27], [28].

Standard medicine has made this mistake many times. Xigris was marketed with a sound biomechanical explanation for sepsis, but it was withdrawn because it was ineffective[29], [30], [31]. Vitamin C in sepsis has also been tried. Vitamin C levels drop whenever there's a severe infection. We know Vitamin C is important for infection fighting, and yet supplementing it in the ICU provides no mortality benefit[32].



Hopefully by now the pattern is clear - whether we're discussing NAD+, hyperbaric chambers, TRT, or statins, anyone can show changes in bloodwork, form mechanistic explanations, trace biochemical pathways, and cite observational data to create "advanced" protocols, but without direct RCT proof that an intervention changes a defined clinical

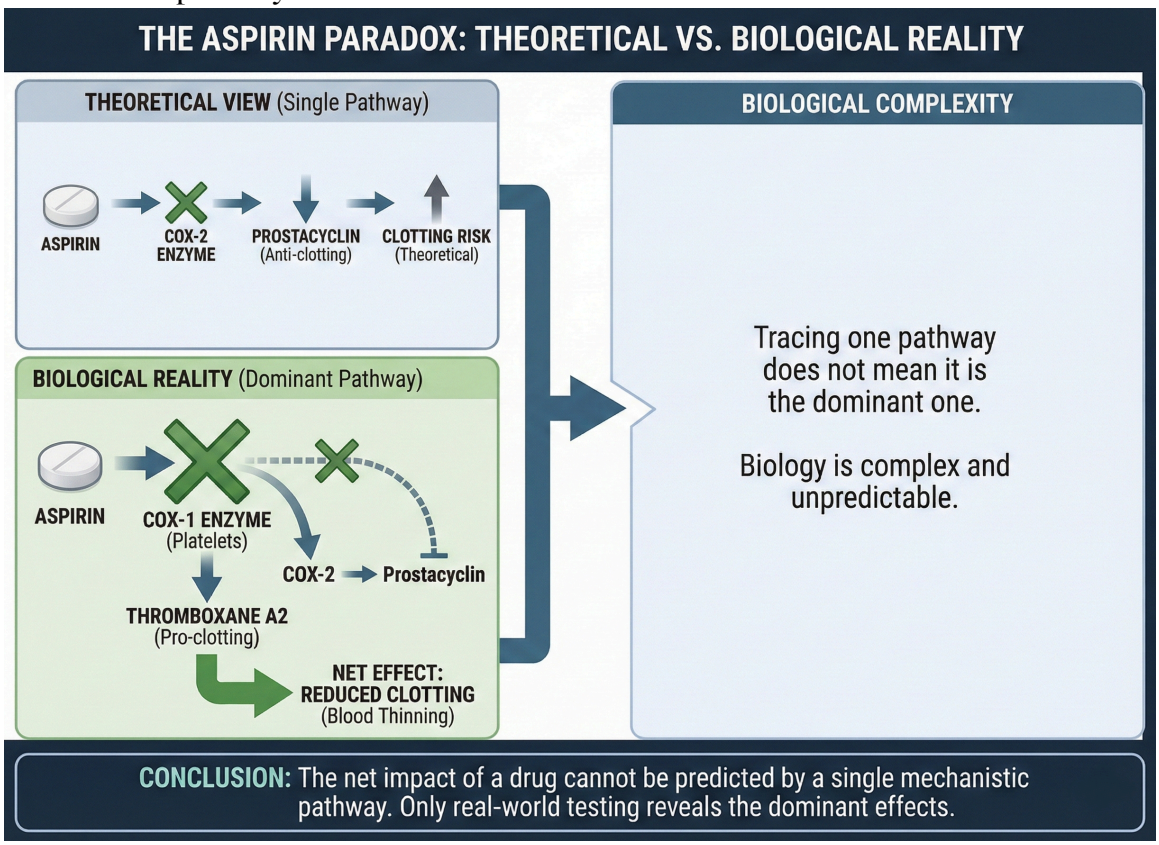
outcome (not a surrogate), you do not have reliable evidence that it is beneficial. In many cases, such interventions have turned out to be actively harmful.

This is the key error of "Medicine 3.0" and health influencer ideologies. They spend hours discussing biological processes and observational data to recommend protocols, giving the veneer of science … then they make claims of improved outcomes without actual proof, assuming that the "reasonable" chain of effects will predict the outcome. If you look at the history of medicine, you realize this "chain of reasoning implies outcomes" concept is *extremely unreliable*, undercutting the entire methodology of so-called "advanced" health recommendations.

Why Does This Keep Happening?

WHY DO WE KEEP FALLING into this trap? One reason is that mechanistic reasoning is seductive. The human brain likes storytelling and causal inferences. In our everyday lives, correlation and causation are often linked, but in observational statistics, two things can have a tight correlation without any causal connection.

Biology is complex and unpredictable. Tracing one pathway does not mean it is the dominant one. For example, Aspirin inhibits COX-2, which can reduce prostacyclin, which in turn prevents clotting. An influencer could argue mechanistically that Aspirin *increases* the risk of clotting, only to find out that in reality it does the opposite, because an alternate, more dominant pathway is at work.



Torcetrapib lowered LDL and ApoB, but increased death because of an unforeseen mechanism that raised blood pressure. The complexity of biological systems means you cannot predict the impact of a drug without testing it. (Some claim AI and better simulations will solve this, but for several reasons, I suspect that is not the case; at a minimum, it is certainly not true now.)

> "We frequently take the flawed shortcut of assuming that if the metric changed, the outcome must follow.

The other reason we fall into this trap is a desire for certainty in an uncertain domain, and a desire for urgent action when true medical knowledge takes time. Observational trials are cheap, correlation is easy to show, and metrics are fast to measure. If I want to show that a supplement changes a blood marker, I can do it in months. If I want to show it reduces death, I have to wait years, recruit thousands of people, and spend millions of dollars.

We frequently take the flawed shortcut of assuming that if the metric changed, the outcome must follow. Influencers double down on this mistake for views and clicks, expanding the error exponentially.

The Placebo Immunity Illusion

ONE MORE REASON WE CONTINUE to fall into this trap - we systematically underestimate the power of the placebo effect.

Most people, including many doctors, harbor a subconscious intuition that the placebo effect is weak, obvious, or easily identified. We tell ourselves, "I would know if a treatment was just a placebo," or "The improvement I felt was too significant to be psychological." We assume that if a benefit lasts for months, or if it involves a physical sensation like increased energy or reduced pain, it must be the result of a "real" physiological change.

Just as most people think they are above-average drivers, we naturally think that the placebo effect is something that "fools" other people, but not *me*. Of course I'm too smart to be "fooled". And yet, if there were a way for you to distinguish between placebo effects and biological effects, there would be no such thing as a placebo effect.

The reality is that placebo effects are often massive[33], [34], [35], [36], [37]. They are particularly dominant in the exact domains health influencers target: pain, focus, mood, sexual function, and athletic performance (even strength, hypertrophy and biomarkers)[37-a], [37-b], [37-c]. Furthermore, these effects can be incredibly persistent. When you combine the placebo effect with the "natural history" of a condition, the fact that many symptoms fluctuate and eventually improve on their own, and "regression to the mean", the statistical reality is that if you measure someone when they feel their worst, they are likely to feel better soon regardless of what you do. And in the process, you create a perfect storm of misattributed success.

I want to be clear about something. If you have personally experienced improvement after starting a supplement, a protocol, or a lifestyle change recommended by an influencer, I am not telling you that your experience isn't real. The placebo effect is very real. You felt better. That happened. But outside of a research study, all treatments are a blend of biological effects and placebo effects. As that blend approaches 0% biological and 100% placebo, it becomes important to carefully analyze the *risks*, because trading placebo benefits for biological risks is a dangerous game.

> "If cutting someone's knee open and "fixing" it can be a placebo effect, we must accept that a supplement protocol or an "optimized" diet is even more susceptible to the same problem."

The placebo effect can even be seen in surgery. For decades, arthroscopic surgery for a degenerative meniscal tear was considered a "gold standard" treatment. Surgeons were utterly convinced of its efficacy because they saw patients improve firsthand. The patients, having undergone a real procedure, were equally convinced. It seemed "obvious" that trimming a torn piece of cartilage would fix knee pain.

However, when researchers finally conducted sham-controlled trials, where patients were given a "placebo surgery" that involved an incision but no actual repair, the results were shocking. There was no difference in outcomes between those who had the "real" surgery and those who received the placebo, but the risks of surgery were the same for both[38]. If cutting someone's knee open and "fixing" it can be a placebo effect, we must accept that a supplement protocol or an "optimized" diet is even more susceptible to the same problem.

This is why anecdotes are the most dangerous form of evidence in medicine. Because humans are wired for storytelling, we see a personal success story and assume causation. We misattribute a natural recovery or a robust placebo response to the specific intervention an influencer is selling. These anecdotes become the foundation of testimonial marketing, creating a feedback loop of false confidence.

In truth, anecdotes are exactly where medicine most often gets fooled. Conviction, experience, and perceived patient improvement are not a substitute for a randomized controlled trial.

And that would be bad enough, but there are even more problems with the optimizers' approach.

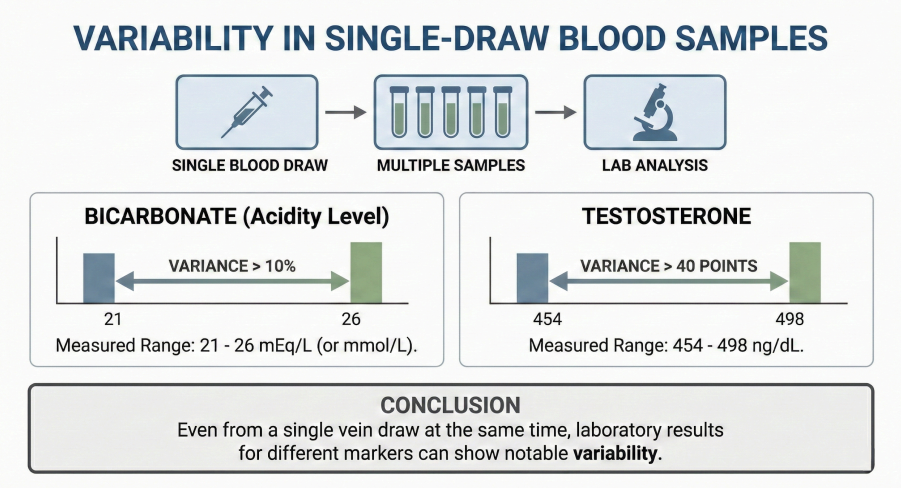
Measurement Variability

The problem with using a rubber ruler.

I HAVE PERSONALLY PERFORMED AN EXPERIMENT that I encourage anyone serious about biohacking to try for themselves under the care of a physician. I drew 15 samples of blood from a single vein at the same time and sent them to the lab. I ran an assortment of tests on that blood, including a standard chemistry panel, testosterone levels, and A1C.

> "bloodwork ... is a fuzzy instrument, not a precise one"

The level of variance was startling. Bicarbonate, a measurement of the acidity level in the blood, had a variance of more than 10% (21 to 26). Testosterone in my samples varied by more than 40 points (454 to 498). I want to emphasize that this blood was not drawn on different days or from different veins. It was the same human, at the same moment in time, the same blood, from the same vein, sent to the same lab, under identical conditions.



My little n-of-1 illustration is a known concept called analytical variability in the lab world. The reality is that bloodwork lab measurement is a fuzzy instrument, not a precise one, though influencers push a false sense of precision either through ignorance or intention[39].

Some assays are more precise than others. A1C, for instance, has one of the lowest assay variances, but even my A1C samples showed a 0.2% variance.

The idea that an influencer can claim to optimize your health by moving a marker like bicarbonate up by a few points is absurd, because the measurement itself isn't precise enough to support those kinds of claims. You may see someone online discuss increasing their testosterone by 40 points, using some protocol, yet I achieved that same increase merely by sending the same blood to the same lab twice.

The problem is compounded by biological variability. While measurement variability refers to the lab's imprecision, biological variability is the fact that our bodies truly

fluctuate over time. For instance, I drew a blood sample one hour after lifting weights and saw an increase in testosterone of more than 100 points (from 381 before to 536 after).

> "You may see someone online discuss increasing their testosterone by 40 points ... I achieved that same increase merely by sending the same blood to the same lab twice."

This is why the medical protocol for diagnosing low testosterone requires samples to be drawn in the morning, fasted, on at least two separate occasions[40], [41]. These standards evolved specifically to account for the fact that testosterone levels fluctuate with sleep-wake cycles, food, and activity, all layered on top of measurement imprecision.

Furthermore, the presence or absence of symptoms changes the interpretation of the lab value. If a healthy, athletic person who eats well has a low insulin level, we interpret that as high insulin sensitivity, which is a marker of health. If, however, insulin is low in a patient in the ICU with severe acidosis, it is indicative of pancreatic failure. The symptoms change the interpretation.

Yet, this reasoning is rarely extended to other hormones like testosterone. If we see low testosterone in a healthy, athletic young man who is effectively gaining muscle and maintaining a low body fat percentage, we should conclude that he has high testosterone sensitivity and that his body is functioning efficiently, just as we do with insulin. Instead, influencers pathologize this, claiming something needs to be "fixed" without any proof of outcomes or comparison to other hormone dynamics.

As we will discuss in a later section, pre-test probability is determined by symptoms, the test then adds information to increase or decrease the likelihood that an intervention is necessary[42]. For now, the takeaway is that biological and measurement variations have two problematic effects[43]. First, they undermine the influencer's tale of precision. Second, they create false narratives.

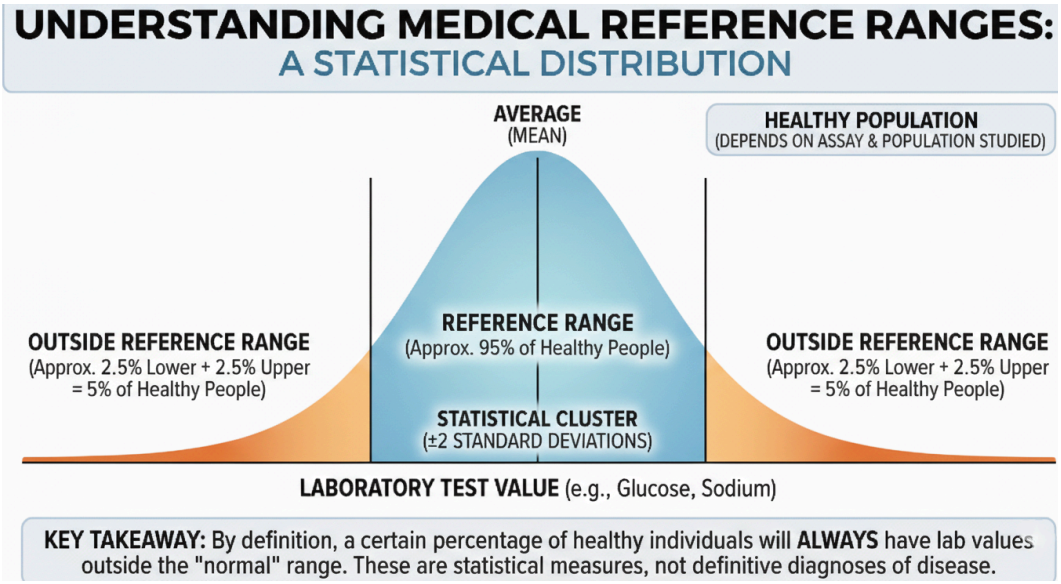
What do I mean by "false narratives"? If you measure enough things, you will eventually find an abnormality. You may then start a protocol or supplement, remeasure, and claim to see an improvement. Unfortunately, that "improvement" is often just lab variability or biological fluctuation, but the narrative is one of a successful intervention[44]. Without population level studies, these anecdotes can "feel" like evidence, when in fact they are just instances of biological and assay variance, and regression to the mean[45].

Before we proceed, I want to note that we should not take this to mean that bloodwork has no value. It has significant value when used in the way it was intended - to raise or lower suspicion for a disease based on a patient's symptoms and presentation, by looking for abnormal deviations from the normal range.

We should realize however, that hijacking the tool of bloodwork and using it differently than intended creates significant errors.

Reference Ranges and Thresholds.

Leaving aside the issue of variability and assuming for a moment that measurement is precise, we must consider the nature of the reference range. Reference ranges are simply measures of the distribution of results in a healthy population[46]. Thus, by definition, a certain percentage of healthy people will always have lab values outside the so-called "normal" range. These ranges are merely statistical clusters, usually two standard deviations from the average, depending on the particular assay being measured[47].



This leads to a crucial realization that influencers ignore: a value labeled "abnormal" because it sits at the extreme end of the bell curve may actually represent a functional adaptation that is physiologically superior for that specific individual's unique biological context. For that person, the outlier value is the functional optimum. An athlete with "low" hemoglobin may have athletic pseudoanemia, which is an important adaptation to their exercise needs. Attempting to "fix" their low hemoglobin would be harmful. On the other end of the spectrum, a patient with untreated sleep apnea may have high hemoglobin, which helps their body send oxygen to their organs even if they stop breathing temporarily. Once again, "optimizing" the lab result would actually harm the patient.

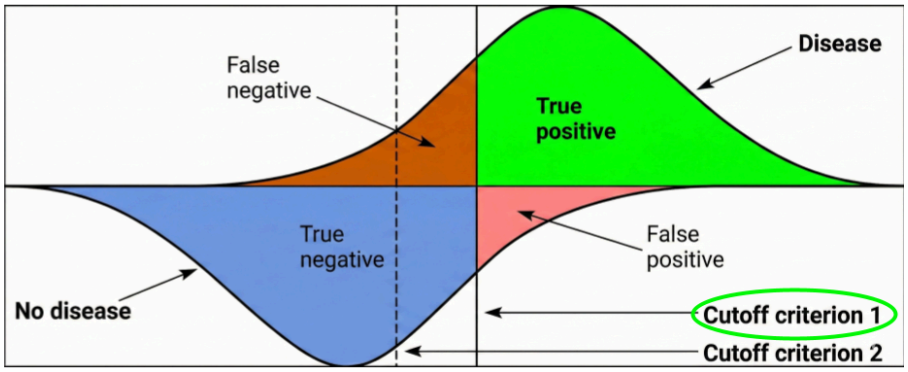
These reference ranges do not inherently predict outcomes. Furthermore, different populations have different ranges - pregnant women may have different distributions than non-pregnant women, or men may differ from women. But, we also cannot over-specify the population. If you only compare to men born in 1980, at low altitude, now living at high altitude, who lift weights and sprint, but don't do zone 2, who used to smoke, but now quit, work in an area with high cosmic radiation and exposure to benzene, who wear their protective equipment about 94% of the time, and who are married, but unhappily, and have a mistress, but no close friends, etc. etc. ... eventually, if you over-specify the population,

you end up just comparing the results to yourself at that moment, and everything is normal by definition.

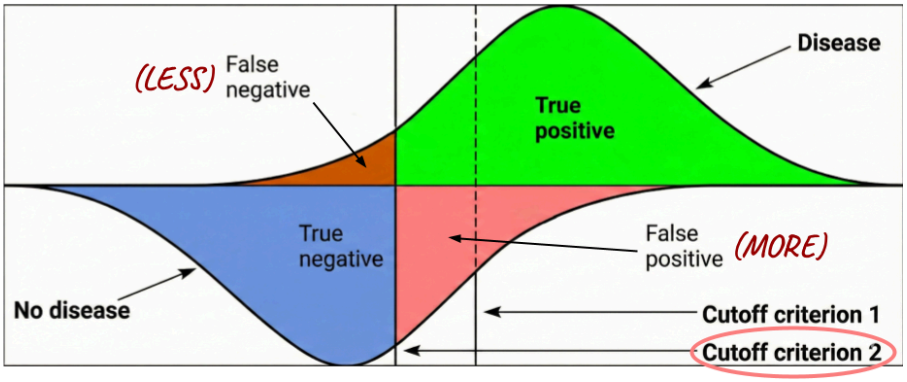
Since we can't do that, then comparison to a reference range is always a comparison to a group of people who are, in some ways, not similar to you[48], [49]. Most importantly, it is completely unproven that "improving" a level within the normal range necessarily creates better outcomes. As I have said before, that requires separate proof in its own trial.

In medicine, we use thresholds to detect abnormality. If you are outside the normal range and have symptoms consistent with that abnormality, we may suspect that correcting it will help. However, all such thresholds involve trade-offs. Whatever level we set as the threshold for "abnormal" creates a set of false positives and false negatives. Selecting a tighter threshold merely changes the ratio of these errors, it does not eliminate them. In clinical medicine, we choose these thresholds on a practical basis: if we want to minimize false negatives, we set a lower bar for abnormality, and vice versa.

1.



2.



The reason I explain this is because influencers perpetually invent so-called "optimal" ranges. Companies like Function Health have invented a whole series of optimal ranges that have no evidence of being superior[50] (or frankly any definition other than the blind assertion of being "optimal", unclear for what outcome, or in what way, or according to whom). They have no proof that moving your blood work toward these results leads to better outcomes. There is no causal evidence that achieving an "optimal" range is in any way beneficial.

In fact, there was a study of men with symptoms of low testosterone plus low-normal levels on blood testing. They were given either testosterone to bring them to high-normal,

or placebo. They couldn't tell the difference on any measure. There were no perceivable changes in symptoms at all[51], [52].

These so-called "optimal" ranges are merely products of existing bias. Having a higher testosterone level is often considered more optimal by influencers, yet we know from multiple studies in different animals that lower testosterone (via castration) actually leads to an increase in lifespan. This is also seen in human data regarding eunuchs and castrated individuals, who live significantly longer on average than similar non-castrated males[53], [54], [55].

The idea that higher testosterone levels lead to increased longevity is strongly counteracted by this data, and yet influencers label high levels as "optimal" even within the normal range. This ignores the sensitivity issues discussed earlier, the failed placebo trials, and the data on castration. It is purely an invention from influencer bias because higher testosterone "feels" better. It is based purely on vibes. The same is true for A1C and various other "optimal" ranges that have been invented with no proof (and in several cases, evidence of harm).

"These so-called 'optimal' ranges are merely products of existing bias."

Bayes' Theorem and False Positives.

THE NEXT ISSUE IS ONE THAT is statistically crucial and yet very poorly understood, even among physicians[56]. Multiple studies have shown that doctors frequently fail to grasp this concept, let alone influencers and the general public. Bayes' Theorem is one of the most central concepts in medicine, yet it can be counter-intuitive.

If I have a test that is extremely accurate, like a pregnancy test, it will have a very low rate of false positives and false negatives. However, accuracy alone does not tell us how likely a positive test is to be correct. The missing factor is the "pre-test probability" - how likely the result was to be positive *before* you even did the measurement.

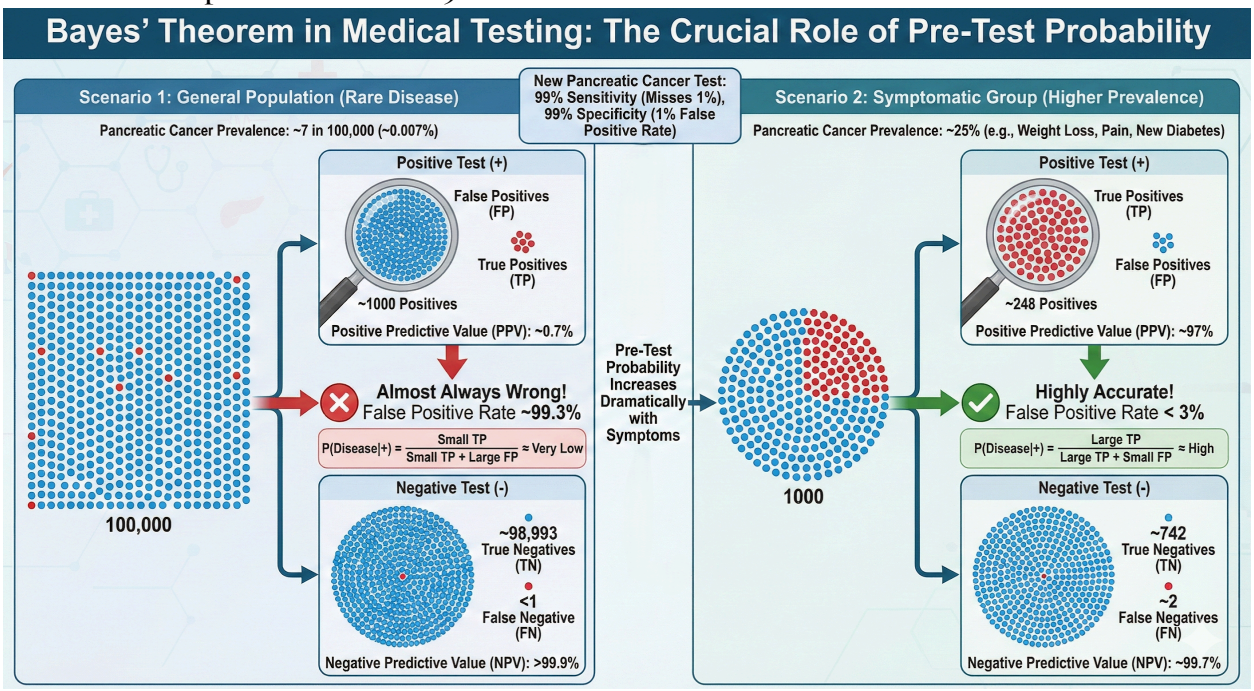
To illustrate, if I show you a positive pregnancy test, you might assume it is almost certainly correct. But if I then tell you the test was run on a man, you immediately know it is a false positive. Your estimate of the likelihood of a false positive goes from very low to 100%, because the pre-test probability of a man being pregnant is zero. No matter how accurate the test is, the context determines the interpretation.

This remains true for rare conditions in the general population. Suppose we invent a new test for pancreatic cancer that is 99% accurate: it has 99% sensitivity (it only misses 1% of cases) and 99% specificity (it only gives a false positive 1% of the time).

If you run this test on an average person in the general population, where the incidence of pancreatic cancer[57], [58] is roughly 7 in 100,000, Bayes' theorem calculates that a positive result would be a false positive 99.3% of the time. The test is "accurate," but because the

disease is so rare in that population, the test is almost always wrong when it claims to find something.

Contrast that with someone who has symptoms: weight loss, abdominal pain, and the new development of diabetes. In this symptomatic group, the pre-test probability is much higher, maybe as high as 25%. If you run the same 99% accurate test on this person, a positive result is a false positive less than 3% of the time.



I go through these examples because it is so poorly understood that tests do not function independently of their context. You cannot merely run a test and rely solely on the result. The symptoms of the patient change the interpretation of the test[59]. This fact, necessitated by the statistics of Bayes' Theorem, is one of the most crucial aspects of medical diagnostics.

Because this is ignored, influencers (and even doctors) continue to recommend testing healthy people who, by definition, have a very low pre-test probability. This guarantees a high rate of false positives. Testing is often framed as a way to bypass the "subjectivity" of symptoms, as if the lab value is more reliable than the patient's presentation. In reality, the context of the patient determines the value of the test.

I even created a teaching tool at rxbayes.com to help doctors with this reasoning, but as direct access to labs and scans increases, it is becoming progressively more important for the public to understand these complexities. If your pre-test probability of an abnormality is extremely low, an abnormal result is more likely a false positive than a true target for "optimization."

> "If your pre-test probability is extremely low, an abnormal result is more likely a false positive than a true target for 'optimization.' "

This problem compounds with every additional test. Reference ranges are typically set at the 95th percentile, meaning 5% of the healthy population will naturally test "abnormal" on any given marker[60], [61]. When you run large panels, finding a "problem" becomes a statistical certainty: with just 20 tests, the chance of receiving at least one false alarm is 64%. If you increase to 50 tests, that probability exceeds 92%. By increasing the number of tests, influencers guarantee the discovery of "abnormalities" that are medically irrelevant but commercially profitable.

After all, if you can market a supplement or intervention to "fix" that "abnormality" (which is likely irrelevant), you can see how lucrative it is to sell large blood panels. "Optimization" in this sense is nothing more than manufacturing a fictitious sense of disease.

Goodhart's Law and the Failure of Surrogate Targets

AT THIS POINT, THE PATTERN we've been seeing has a name. Goodhart's Law: *when a measure becomes a target, it ceases to be a good measure*[62].

Many biological measurements are genuinely useful indicators of health risk. Muscle strength, aerobic fitness, blood pressure, glucose control, and other markers often correlate strongly with long-term outcomes. These measurements became useful because improving them through real physiological changes, including exercise, improved metabolic health, smoking cessation, and blood pressure control, also tended to improve outcomes.

But the problem arises when we begin assuming that **anything that improves the measurement must also improve health**[63].

Once the number itself becomes the target, rather than the physiology behind it, the relationship breaks down.

> "Goodhart's Law: *when a measure becomes a target, it stops being a good measure.*"

We can see a clear application of this in the example of grip strength. In various studies, grip strength was used as a surrogate marker for overall body strength, and in those studies, it correlated strongly with longevity[64]. However, if you take that information to mean that you should specifically train your grip, and you engage only in grip-strengthening exercises, you have almost certainly not increased your longevity. You have merely broken the correlation between grip strength and total body strength that the study relied upon.

Manipulating the metric merely breaks the assumed connection, without actually changing the outcome.

We have already seen the same pattern repeatedly:

Arrhythmias could be suppressed, yet mortality increased. Blood sugar could be tightly controlled in ICU patients, yet mortality worsened. Homocysteine levels could be lowered, yet cardiovascular risk remained unchanged. HDL increased and LDL decreased, yet mortality rose in drug trials. The measurements moved. Outcomes did not.

> "Optimization culture assumes that because a marker correlates with disease risk, pushing the marker toward an "optimal" value must improve health."

Sometimes the disconnect is even more direct. Hemoglobin A1c is widely used as a marker of long-term glucose control. But A1c can fall without glucose physiology improving at all[65]. Blood donation will lower A1c by forcing the body to generate fresh red blood cells, without changing anything about blood sugar levels, or the damage caused by those levels[66].

The number improves. The underlying problem does not.

Optimization culture assumes that because a marker correlates with disease risk, pushing the marker toward an "optimal" value must improve health. But again and again, medicine has discovered that manipulating the marker is not the same as manipulating the disease. The *way* you change a metric is the thing that has to be tested against outcomes, otherwise you're confusing the map for the territory.

Goodhart's Law explains why surrogate endpoints and metrics do not necessarily translate to health. For all the reasons mentioned so far, you can see how blood work markers do not function the way influencers would have you believe they do.

But these issues are not limited to blood work. They apply equally to other diagnostic technologies.

MRI and Other "Early" Scans

Different Tool, Same Errors

Full-body MRI replicates all of these same problems[67]. Because of the dictates of Bayes' Theorem, a full-body MRI on a healthy person (low pretest probability) almost guarantees that most findings will be false positives, leading inevitably to overdiagnosis and overtreatment (the harms of which are real, but always handwaved away or minimized)[68].

Furthermore, population-level autopsy studies show that many people develop malignancies that could be detected on a slide or a scan but would have had no impact on their lives if left alone[69], [70], [71], [72], [73],[74]. We have certainly seen this with prostate cancer, to the point that screening recommendations have essentially been all but revoked (the guidelines say "individual decision", which is basically pushing liability onto the patient - AKA "no benefit" plus lobbying and special interests)[75], [76].

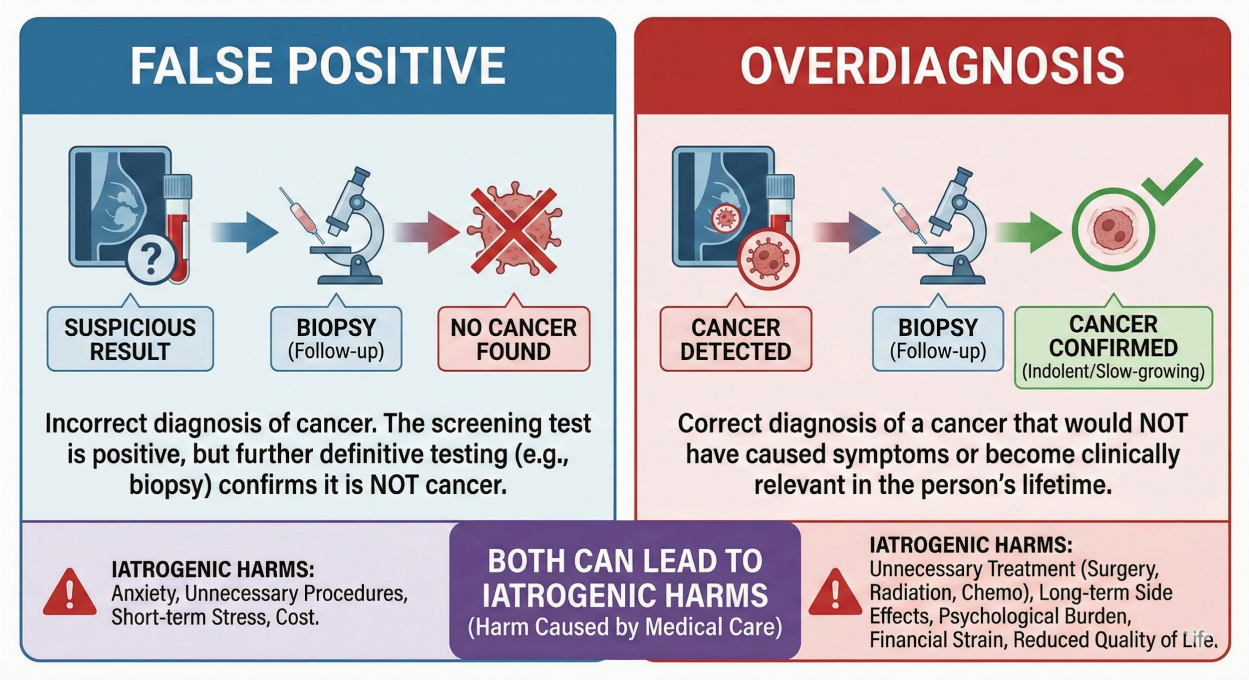
The issue occurs with mammography as well, though it is less-often discussed and more nuanced[77]. The USPSTF still sells the central claim as mortality reduction[78], [79]. That's the official story. But Cochrane's critique is blunt: the apparent benefit leans heavily on older trials with substantial methodological limitations[80]. When you restrict attention to better-quality evidence, the mortality signal is limited to disease-specific analysis, with no overall mortality benefit, and even the disease-specific claims are inflated by misclassification. The false-positive rate of mammography is significant, but even more concerning is the rate of overdiagnosis[81]. Overdiagnosis is a separate issue from false positives. False positive is incorrect suspicion or diagnosis of cancer. Overdiagnosis is (correct) diagnosis of a cancer that would not have become clinically relevant in the person's lifetime.

Cochrane estimates 10 women will be overdiagnosed (unnecessarily treated with surgery, radiation, endocrine therapy, chemo, etc.), for every 1 cancer death prevented per 2000 women invited for 10 years, with no improvement in overall longevity[82]. False positives are costly and anxiety-producing, but they usually (though not always) get figured out[83]. But overdiagnosis is a medical catastrophe[84], [85],[86], [87]. It converts healthy people into cancer patients on paper and exposes them to real harms, for zero benefit[88], [89].

"False positives are costly and anxiety-producing, but ...

overdiagnosis is a medical catastrophe.

It converts healthy people into cancer patients on paper and exposes them to real harms, for zero benefit."



And Bleyer and Welch describe mammography as the signature pattern of overdiagnosis: When mammogram screening programs are introduced, we see a significant increase in the detection of Stage 1 disease[90]. These women then undergo the full weight of modern medicine: chemotherapy, radiation, and surgery, and yet, on a ten-year timeframe, there is only a marginal reduction in late-stage disease, and no detectable reduction in overall mortality. Treating the stage 1 cancers creates stories of "cures," and yet on a population level there is only a slight dent in advanced disease, and no improvement in overall longevity.

It is possible that some (many?) Stage 1 cancers are halted by the immune system, never progressing or becoming a problem. Unfortunately, on an individual basis, we have no way to tell the difference. We cannot know in the moment whether a detected Stage 1 cancer is the type that will kill or the type that will not. Thus, we are forced to treat everyone. While individuals undergo toxic treatments, the population-level statistics show no benefit.

Effective screening requires a rare set of circumstances. The test must be accurate enough to overcome the Bayesian hurdle of testing a low-probability population. The disease must be treatable (otherwise why bother at all?). The treatment must be significantly more effective in the early stages than the late stages (otherwise, why not wait for symptoms?). And finally, the disease must not have a high rate of spontaneous regression or stabilization.

It is very rare for all these conditions to exist[91], [92], [93], [94], [95], [96]. They do exist for lung cancer in smokers, which is why low-dose CT scanning saves lives in that specific group[97], [98], [99]. However, the NordICC trial, the first randomized controlled trial involving colonoscopy failed to show a mortality benefit from an invitation to colonoscopy[100], [101].

Full-body MRI will exponentially expand these problems because most people do not understand these nuances. Screening is not a simple matter of "looking more carefully." More detection does not guarantee better outcomes. The "catch it early" mantra is deeply woven into our culture, yet it has been called into question for many of our most common cancers[102], [103].

There is a deeper and more concerning aspect of all of this. Just as we discussed with bloodwork optimization "regressing to the mean", an analogous process exists with malignancies. Because many cancers do not progress or may even spontaneously resolve, full-body MRIs generate countless stories of people who believe a scan "saved their life" after a tumor was detected and treated. In reality, many of those people were never at risk of dying from that tumor, and yet these anecdotes reinforce a false belief in early detection, while the actual harms of overdiagnosis and the associated interventions are ignored.

> "More detection does not guarantee better outcomes."

Conclusion

We started from the natural desire to pursue higher health aspirations. We recognized the failure of standard medicine to assist us in these goals, and explored the reasonable intuition that we could take medicine's tests and scans and just *do more*. And yet now, we have found that the very tools themselves are fundamentally flawed when applied to health optimization instead of disease detection.

Mechanistic reasoning cannot reliably predict outcomes based on changes in metrics. Interpreting markers outside the context of a patient's symptoms is statistically flawed. The inherent variability of lab measurements ensures that "improvements" are often just statistical noise. By running dozens or hundreds of tests on healthy people, influencers guarantee the discovery of "abnormalities" that are medically irrelevant but commercially profitable. And finally, more detection through advanced scanning often leads to overdiagnosis rather than longer life.

But before we go further, take a moment to notice how challenging and non-obvious the journey to this realization has been. Understanding the true complexity of these topics is really difficult.

You've needed to take in information about medical history, obscure statistics, unexpected outcomes, and subjectivity in analysis and diagnosis. Most people don't intuitively understand Bayes' Theorem, and almost no one has thought about (or heard of) overdiagnosis, analytical variability, or metric/outcome discordance. And yet, once you do understand these ideas, the entire underlying architecture of the "health optimization / Medicine 3.0" model collapses.

If you've gotten this far, you've already done more epistemic work than most people ever do on their own health decisions. If anyone, doctor or influencer, is to make medical recommendations, I believe that they should have an awareness of these same concepts, or else their authority is mere marketing, and they are perpetrating true harm, even if it can't be detected. I hope this document helps you spot these false authorities, whether they have an MD or a million followers.

I hope you are now able to spot when someone is violating your new understanding by equating metrics with outcomes, ignoring test interpretation fuzziness, and failing to account for the invisible (but real) harms of false positives and overdiagnosis.

None of this means that medicine doesn't work. Evidence-based interventions save lives every day. Blood pressure medications reduce strokes. Statins reduce second heart attacks in those who have had one. Scans and bloodwork help us make more accurate diagnoses in those with symptoms, which in turn allows more targeted treatments. Medicine and medical technologies are extraordinary when used in the way they were designed. The problem is not medicine. The problem is the hijacking of medical tools for purposes they were never designed to serve, and the sale of that hijacking as "advanced" care.

We need to stop giving charlatans a pass on bad science. But that still leaves you with a question that deserves a real answer: if this kind of optimization fails, then what actually *does* work?

The answer is simpler than the optimization industry wants you to believe. There are a small number of interventions with direct, high-quality evidence of extending human lifespan. Not surrogate markers. Not mechanistic reasoning. Actual evidence of actual people living longer.

Understanding why these specific interventions meet my standards of evidence, when so many others do not, is the critical next step. This is the entire focus of my no-cost private briefing.

Reserve your seat here: https://riverrockmedical.com/briefing

References

1. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med.* 1989;8(4):431-440. URL: https://pubmed.ncbi.nlm.nih.gov/2727467/

2. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med.* 1996;125(7):605-613. URL:https://www.acpjournals.org/doi/10.7326/0003-4819-125-7-199610010-00011

3. Manyara AM, Davies P, et al. Definitions, acceptability, limitations, and guidance in the use and reporting of surrogate end points in trials: a scoping review. *J Clin Epidemiol.* 2023. URL:https://pubmed.ncbi.nlm.nih.gov/37380118/

4. Yusuf, S., et al. (2008). Telmisartan, Ramipril, or Both in Patients at High Risk for Vascular Events. *New England Journal of Medicine*, 358(15), 1547–1559. URL:https://doi.org/10.1056/NEJMoa0801317

5. Fried, L. F., et al. (2013). Combined Angiotensin Inhibition for the Treatment of Diabetic Nephropathy. New England Journal of Medicine, 369(20), 1892–1903. URL:https://doi.org/10.1056/NEJMoa1303154

6. Boden WE, O'Rourke RA, Teo KK, et al; COURAGE Trial Research Group. Optimal medical therapy with or without PCI for stable coronary disease. *N Engl J Med.* 2007;356(15):1503-1516. URL: https://www.nejm.org/doi/full/10.1056/NEJMoa070829

7. Maron DJ, Hochman JS, Reynolds HR, et al; ISCHEMIA Research Group. Initial invasive or conservative strategy for stable coronary disease. *N Engl J Med.* 2020;382(15):1395-1407. URL: https://www.nejm.org/doi/full/10.1056/NEJMoa1915922

8. Stampfer, M. J., et al. (1991). Postmenopausal Estrogen Therapy and Cardiovascular Disease: Ten-year Follow-up from the Nurses' Health Study. *New England Journal of Medicine*, 325(11), 756–762. URL:https://doi.org/10.1056/NEJM199109123251102

9. The Writing Group for the PEPI Trial. (1995). Effects of Estrogen or Estrogen/Progestin Regimens on Heart Disease Risk Factors in Postmenopausal Women: The Postmenopausal Estrogen/Progestin Interventions (PEPI) Trial. *JAMA*, 273(3), 199–208. URL:https://pubmed.ncbi.nlm.nih.gov/7807658/

10. Walsh, B. W., et al. (1991). Effects of Postmenopausal Estrogen Replacement on the Concentrations and Metabolism of Plasma Lipoproteins. *New England Journal of Medicine*, 325(17), 1196–1204. URL:https://pubmed.ncbi.nlm.nih.gov/1922206/

11. American College of Physicians. (1992). Guidelines for Counseling Postmenopausal Women about Preventive Hormone Therapy. *Annals of Internal Medicine*, 117(12), 1038–1041. URL:https://pubmed.ncbi.nlm.nih.gov/1443972/

12. Godsland, I. F. (2001). Effects of postmenopausal hormone replacement therapy on lipid, lipoprotein, and apolipoprotein (a) concentrations: a quantitative review. *Metabolism*, 50(7), 801–813. URL:https://pubmed.ncbi.nlm.nih.gov/11334901/

13. Rossouw JE, Anderson GL, Prentice RL, et al; Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results. *JAMA.* 2002;288(3):321-333. URL: https://jamanetwork.com/journals/jama/fullarticle/195120

14. CAST Investigators. Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. *N Engl J Med.* 1989;321(6):406-412. URL: https://pubmed.ncbi.nlm.nih.gov/2473403/

15. Manson JAE, Hsia J, Johnson KC, et al. Estrogen plus progestin and the risk of coronary heart disease. *N Engl J Med.* 2003;349(6):523-534. URL: https://www.nejm.org/doi/full/10.1056/NEJMoa030808

16. Echt DS, Liebson PR, Mitchell LB, et al; CAST Investigators. Mortality and morbidity in patients receiving encainide, flecainide, or placebo: the Cardiac Arrhythmia Suppression Trial. *N Engl J Med.* 1991;324(12):781-788. URL: https://www.nejm.org/doi/full/10.1056/NEJM199103213241201

17. Finfer S, Chittock DR, Su SY, et al; NICE-SUGAR Study Investigators. Intensive versus conventional glucose control in critically ill patients. *N Engl J Med.* 2009;360(13):1283-1297. URL: https://www.nejm.org/doi/full/10.1056/NEJMoa0810625

18. Bønaa KH, Njølstad I, Ueland PM, et al. Homocysteine lowering and cardiovascular events after acute myocardial infarction. *N Engl J Med.* 2006;354(15):1578-1588. URL: https://www.nejm.org/doi/full/10.1056/NEJMoa055227

19. Ebbing M, Bønaa KH, Nygård O, et al. Mortality and cardiovascular events in patients treated with homocysteine-lowering B vitamins after coronary angiography: a randomized controlled trial. *JAMA.* 2008;300(7):795-804. URL: https://jamanetwork.com/journals/jama/fullarticle/182412

20. Barter PJ, Caulfield M, Eriksson M, et al. Effects of torcetrapib in patients at high risk for coronary events. *N Engl J Med.* 2007;357(21):2109-2122. URL: https://www.nejm.org/doi/full/10.1056/NEJMoa0706628

21. American College of Cardiology. ILLUMINATE (Investigation of Lipid Level Management to Understand its Impact in Atherosclerotic Events) — Presented at AHA 2007. URL: https://www.acc.org/Latest-in-Cardiology/Clinical-Trials/2010/02/23/19/09/ILLUMINATE--Presented-at-AHA-2007

22. The Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *N Engl J Med.* 1994;330(15):1029-1035. URL: https://www.nejm.org/doi/full/10.1056/NEJM199404143301501

23. Omenn GS, Goodman GE, Thornquist MD, et al. Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease. *N Engl J Med.* 1996;334(18):1150-1155. URL: https://www.nejm.org/doi/full/10.1056/NEJM199605023341802

24. Yusuf S, Dagenais G, Pogue J, Bosch J, Sleight P; HOPE Study Investigators. Vitamin E supplementation and cardiovascular events in high-risk patients. *N Engl J Med.* 2000;342(3):154-160. URL: https://www.nejm.org/doi/full/10.1056/NEJM200001203420302

25. Albanes D, Heinonen OP, Taylor PR, et al. Alpha-tocopherol and beta-carotene supplements and lung cancer incidence in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study: effects of baseline characteristics and study compliance. *J Natl Cancer Inst.* 1996;88(21):1560-1570. URL: https://pubmed.ncbi.nlm.nih.gov/8901854/

26. Yusuf S, Dagenais G, Pogue J, Bosch J, Sleight P; HOPE Study Investigators. Vitamin E supplementation and cardiovascular events in high-risk patients. *N Engl J Med.* 2000;342(3):154-160. URL: https://www.nejm.org/doi/full/10.1056/NEJM200001203420302

27. Lonn E, Bosch J, Yusuf S, et al; HOPE and HOPE-TOO Trial Investigators. Effects of long-term vitamin E supplementation on cardiovascular events and cancer: a randomized controlled trial. *JAMA.* 2005;293(11):1338-1347. URL: https://jamanetwork.com/journals/jama/fullarticle/200541

28. American College of Cardiology. The Heart Outcomes Prevention Evaluation Study — HOPE Vitamin E and HOPE-TOO. 2005 update page. URL: https://www.acc.org/Latest-in-Cardiology/Clinical-Trials/2010/02/23/19/07/HOPE--Vitamin-E-and-HOPE-TOO

29. U.S. Food and Drug Administration (FDA). FDA Drug Safety Communication: Voluntary market withdrawal of Xigris (drotrecogin alfa [activated]) due to lack of efficacy. 2011. URL: https://www.fda.gov/drugs/drug-safety-and-availability/fda-drug-safety-communication-voluntary-market-withdrawal-xigris-drotrecogin-alfa-activated-due

30. Ranieri VM, Thompson BT, Barie PS, et al; PROWESS-SHOCK Study Group. Drotrecogin alfa (activated) in adults with septic shock. *N Engl J Med.* 2012;366(22):2055-2064. URL: https://www.nejm.org/doi/full/10.1056/NEJMoa1202290

31. Lai PS, Thompson BT. Why activated protein C was not successful in severe sepsis and septic shock: are we still tilting at windmills? Curr Infect Dis Rep. 2013 Oct;15(5):407-12. doi: 10.1007/s11908-013-0358-9. PMID: 23925482; PMCID: PMC3829688. https://pmc.ncbi.nlm.nih.gov/articles/PMC3829688/

32. Lamontagne F, Masse MH, Menard J, et al; LOVIT Investigators and the Canadian Critical Care Trials Group. Intravenous vitamin C in adults with sepsis. *N Engl J Med.* 2022;386(25):2387-2398. URL: https://www.nejm.org/doi/full/10.1056/NEJMoa2200644

33. Colloca L. The placebo effect in pain therapies. *Annu Rev Pharmacol Toxicol*. 2019;59:191-211. doi:10.1146/annurev-pharmtox-010818-021542. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC6402571/

34. van Lennep JPA, Trossèl F, Otten RHJM, van Middendorp H, Evers AWM, Szadek KM. Placebo effects in low back pain: A systematic review and meta-analysis of the literature. *Eur J Pain*. 2021;25(9):1876-1897. doi:10.1002/ejp.1811. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC8518410/

35. Walsh BT, Seidman SN, Sysko R, Gould M. Placebo response in studies of major depression: variable, substantial, and growing. *JAMA*. 2002;287(14):1840-1847. doi:10.1001/jama.287.14.1840. Available from: https://pubmed.ncbi.nlm.nih.gov/11939870/

36. Weinberger JM, Houman J, Caron AT, Patel DN, Baskin AS, Ackerman AL, et al. Female sexual dysfunction and the placebo effect: A meta-analysis. *Obstet Gynecol*. 2018;132(2):453-458. doi:10.1097/AOG.0000000000002733. Available from: https://pubmed.ncbi.nlm.nih.gov/29995725/

37. Chhabra B, Szabo A. Placebo and nocebo effects on sports and exercise performance: A systematic literature review update. *Nutrients*. 2024;16(13):1975. doi:10.3390/nu16131975. Available from: https://pubmed.ncbi.nlm.nih.gov/38999724/

37-a. Lindberg, K., Bjørnsen, T., Vårvik, F.T. et al. **The effects of being told you are in the intervention group on training results: a pilot study.** Sci Rep 13, 1972 (2023). Available from: https://doi.org/10.1038/s41598-023-29141-7.

37-b. Ariel, Gideon B. and W. A. Saville. **Anabolic steroids: the physiological effects of placebos.** Medicine and science in sports 4 (1972): 124-126. Available from: https://www.arielnet.com/articles/show/adi-pub-01242/anabolic-steroids-the-physiological-effects-of-placebos

37-c. Crum AJ, Corbin WR, Brownell KD, Salovey P. **Mind over milkshakes: mindsets, not just nutrients, determine ghrelin response.** Health Psychol. 2011 Jul;30(4):424-9; discussion 430-1. doi: 10.1037/a0023467. PMID: 21574706. Available from: https://pubmed.ncbi.nlm.nih.gov/21574706/

38. Sihvonen R, Paavola M, Malmivaara A, et al. **Arthroscopic partial meniscectomy versus sham surgery for a degenerative meniscal tear.** *N Engl J Med.* 2013;369(26):2515-2524. URL: https://www.nejm.org/doi/full/10.1056/NEJMoa1305189

39. Fraser CG. **Reference change values.** *Clin Chem Lab Med.* 2011;50(5):807-812. doi:10.1515/CCLM.2011.733. URL: https://pubmed.ncbi.nlm.nih.gov/21958344/

40. Bhasin S, Brito JP, Cunningham GR, Hayes FJ, Hodis HN, Matsumoto AM, et al. **Testosterone therapy in men with hypogonadism: An Endocrine Society clinical practice guideline.** *J Clin Endocrinol Metab.* 2018;103(5):1715-1744. Available from: https://pubmed.ncbi.nlm.nih.gov/29562364/

41. Mulhall JP, Trost LW, Brannigan RE, Kurtz EG, Redmon JB, Chiles KA, et al. **Evaluation and management of testosterone deficiency: AUA guideline.** *J Urol.* 2018;200(2):423-432. Available from: https://pubmed.ncbi.nlm.nih.gov/29601923/

42. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. **Helping doctors and patients make sense of health statistics.** *Psychol Sci Public Interest.* 2007;8(2):53-96. URL: https://www.stat.berkeley.edu/~aldous/157/Papers/health_stats.pdf

43. AcuteCareTesting. **Biological variation — what's it all about?** (includes RCV concept + calculation framing). URL: https://acutecaretesting.org/en/articles/biological-variation--whats-it-all-about

44. Badrick T. **Biological variation: Understanding why it is so important?** *Clin Biochem.* 2021. URL: https://www.sciencedirect.com/science/article/pii/S2352551720301621

45. Barnett AG, van der Pols JC, Dobson AJ. **Regression to the mean: what it is and how to deal with it.** *Int J Epidemiol.* 2005;34(1):215-220. URL: https://pubmed.ncbi.nlm.nih.gov/15333621/

46. **CALIPER (Canadian Laboratory Initiative on Pediatric Reference Intervals).** Healthcare Professionals page (definition: 2.5th–97.5th percentiles). URL: https://caliperproject.ca/caliper/healthcare-professionals/

47. Ozarda Y. **Reference intervals: current status, recent developments and future considerations.** *Biochem Med (Zagreb).* 2016;26(1):5-16. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC4783089/

48. Røys EÅ, Viste K, Jones GD, et al. **Estimating Reference Change Values Using Routine Patient Data: A Novel Pathology Database Approach.** *Clin Chem.* 2025;71(2):307-316. URL: https://academic.oup.com/clinchem/article/71/2/307/7874401

49. Lacher DA, Hughes JP, Carroll MD. **Biological variation of laboratory analytes based on the 1999–2002 National Health and Nutrition Examination Survey.** *Natl Health Stat Report.* 2010;(21). URL: https://stacks.cdc.gov/view/cdc/5756

50. Function Health. **The Function Index.** FunctionHealth.com. Available from: https://www.functionhealth.com/campaign/the-function-index

51. Mok SF, Fennell C, Savkovic S, Turner L, Jayadev V, Conway A, Handelsman DJ. **Testosterone for androgen deficiency-like symptoms in men without pathologic hypogonadism: A randomized, placebo-controlled cross-over with masked choice extension clinical trial.** *J Gerontol A Biol Sci Med Sci.* 2020;75(9):1723-1731. doi:10.1093/gerona/glz195. Available from: https://pubmed.ncbi.nlm.nih.gov/31425577/

52. Snyder PJ, Bhasin S, Cunningham GR, Matsumoto AM, Stephens-Shields AJ, Cauley JA, et al. **Lessons from the Testosterone Trials.** *Endocr Rev.* 2018;39(3):369-386. doi:10.1210/er.2017-00234. Available from: https://pubmed.ncbi.nlm.nih.gov/29522088/

53. Garratt M, et al. **Sterilization and contraception increase lifespan across vertebrates.** *Nature.* 2026;649(8099):1264-1272. Epub 2025 Dec 10. doi:10.1038/s41586-025-09836-9. URL: https://pubmed.ncbi.nlm.nih.gov/41372417/

54. Hamilton JB, Mestler GE. **Mortality and survival: comparison of eunuchs with intact men and women in a mentally retarded population.** *J Gerontol.* 1969;24(4):395-411. doi:10.1093/geronj/24.4.395. URL: https://pubmed.ncbi.nlm.nih.gov/5362349/

55. Min KJ, Lee CK, Park HN. The lifespan of Korean eunuchs. *Curr Biol*. 2012;22(18):R792-R793. doi:10.1016/j.cub.2012.06.036. URL: https://pubmed.ncbi.nlm.nih.gov/23017989/

56. Austin LC. Physician and nonphysician estimates of positive predictive value in diagnostic vs mass screening mammography: an examination of Bayesian reasoning. *Med Decis Making*. 2019. URL: https://journals.sagepub.com/doi/abs/10.1177/0272989X18823757

57. Rawla P, Sunkara T, Gaduputi V. Epidemiology of Pancreatic Cancer: Global Trends, Etiology and Risk Factors. *World J Oncol*. 2019;10(1):10-27. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC6396775/

58. National Cancer Institute. SEER Cancer Stat Facts: Pancreatic Cancer. URL: https://seer.cancer.gov/statfacts/html/pancreas.html

59. Navarrete G, Correia R, Sirota M, Juanchich M, Huepe D. Doctor, what does my positive test mean? From Bayesian textbook tasks to personalized risk communication. *Front Psychol*. 2015;6:1327. doi:10.3389/fpsyg.2015.01327. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC4585185/

60. NCCLS (now CLSI). How to Define and Determine Reference Intervals in the Clinical Laboratory; Approved Guideline—Second Edition (C28-A2). 2000. URL: https://mdcpp.com/doc/standard/NCCLSC28-A2-2000.pdf

61. CLSI. Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; *Approved Guideline (C28-A3)*. 2008. (Official document page / citation portal). URL: https://www.researchgate.net/publication/281239796_Defining_Establishing_and_Verifying_Reference_Intervals_in_the_Clinical_Laboratory_Approved_Guidelines_CLSI_document_C28-A3_Vol_28_No_3

62. Goodhart C. Problems of monetary management: the UK experience. In: *Papers in Monetary Economics*. Reserve Bank of Australia; 1975. URL: https://en.wikipedia.org/wiki/Goodhart%27s_law

63. Mattson C, Bushardt RL, Artino AR Jr. "When a Measure Becomes a Target, It Ceases to be a Good Measure." *J Grad Med Educ*. 2021;13(1):2-5. doi:10.4300/JGME-D-20-01492.1. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC7901608/

64. Leong DP, Teo KK, Rangarajan S, et al. Prognostic value of grip strength: findings from the Prospective Urban Rural Epidemiology (PURE) study. *Lancet*. 2015;386(9990):266-273. URL: https://pubmed.ncbi.nlm.nih.gov/25982160/

65. U.S. Food and Drug Administration. 510(k) Review Memorandum / Summary — FreeStyle Libre 2 Flash Glucose Monitoring System (K193371). 2020. (Notes significant ascorbic acid interference). URL: https://www.accessdata.fda.gov/cdrh_docs/reviews/K193371.pdf

66. Dijkstra A, et al. Whole blood donation affects the interpretation of HbA1c. *PLOS ONE*. 2017;12(1):e0170807. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC5261611/

67. Richter A, et al. The effects of incidental findings from whole-body MRI on the frequency of biopsies and detected malignancies or benign conditions in a general population cohort study. *Eur J Radiol*. 2020. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC7524843/

68. Kwee RM, Kwee TC. Whole-body MRI for preventive health screening: a systematic review and meta-analysis. *J Magn Reson Imaging*. 2019;50(5):1489-1503. URL: https://pubmed.ncbi.nlm.nih.gov/30932247/

69. Bell KJL, Del Mar C, Wright G, Dickinson J, Glasziou P. Prevalence of incidental prostate cancer: a systematic review of autopsy studies. *Int J Cancer*. 2015;137(7):1749-1757. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC4682465/

70. Thomas ET, Del Mar C, Glasziou P, Wright G, Barratt A, Bell KJL. Prevalence of incidental breast cancer and precursor lesions: a systematic review and meta-analysis. *BMJ Open*. 2017;7(9):e016879. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC5712106/

71. American Thyroid Association. Prevalence of thyroid cancer found in autopsy studies has remained stable over time. *Clinical Thyroidology for Patients*. 2016;9(12):10. URL: https://www.thyroid.org/patient-thyroid-information/ct-for-patients/december-2016/vol-9-issue-12-p-10/

72. Poelhekken K, et al. The natural history of ductal carcinoma in situ (DCIS) in modelling studies (non-progressive DCIS range). Breast. 2023. Available from: https://www.sciencedirect.com/science/article/pii/S0960977623005350

73. American Thyroid Association. Thyroid cancer over-diagnosis is associated with increased incidence but not increased mortality. 2021. Available from: https://www.thyroid.org/patient-thyroid-information/ct-for-patients/december-2021/vol-14-issue-12-p-7-8/

74. Uozaki H, Kikuchi Y, Watanabe M, et al. Trends in the Hidden Burden of Cancer in an Autopsy-Based Study Over 66 Years in Japan. *JAMA Netw Open*. 2026;9(2):e2557812. doi:10.1001/jamanetworkopen.2025.57812. URL: https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2844664

75. Jacklin C. More men die with prostate cancer than because of it. *Prostate Int*. 2021. Available from: https://www.sciencedirect.com/science/article/pii/S2468294220300605

76. Sandhu GS, Andriole GL. Overdiagnosis of prostate cancer. JNCI Monogr. 2012. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC3540879/

77. Kowalski AE. Mammograms and Mortality: How Has the Evidence Evolved? J Econ Perspect. 2021 Spring;35(2):119-140. doi: 10.1257/jep.35.2.119. PMID: 34421215; PMCID: PMC8371936. https://pmc.ncbi.nlm.nih.gov/articles/PMC8371936/

78. U.S. Preventive Services Task Force. Breast Cancer: Screening. USPSTF Recommendation Statement (Final, 2024). URL: https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/breast-cancer-screening

79. Nicholson WK, et al. Screening for Breast Cancer: US Preventive Services Task Force Recommendation Statement. *JAMA*. 2024. URL: https://jamanetwork.com/journals/jama/fullarticle/2818283

80. Gøtzsche PC, Jørgensen KJ. Screening for breast cancer with mammography. *Cochrane Database Syst Rev*. 2013;2013(6):CD001877. doi:10.1002/14651858.CD001877.pub5. URL: https://pubmed.ncbi.nlm.nih.gov/23737396/

Gøtzsche PC, Jørgensen KJ. Screening for breast cancer with mammography (Cochrane Library record). 2013. URL: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD001877.pub5/information

81. Flemban AF, et al. Overdiagnosis due to screening mammography for breast cancer: systematic review & meta-analysis. 2023. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC10051653/

82. Gøtzsche PC, Jørgensen KJ. Screening for breast cancer with mammography. *Cochrane Database Syst Rev*. 2013;(6):CD001877. doi:10.1002/14651858.CD001877.pub5. Available from: https://pubmed.ncbi.nlm.nih.gov/23737396/

83. Bond M, Pavey T, Welch K, et al. Psychological consequences of false-positive screening mammograms: systematic review. J Med Screen. 2013. Available from: https://pubmed.ncbi.nlm.nih.gov/23540978/

84. Carter JL, Coletti RJ, Harris RP. Quantifying and monitoring overdiagnosis in cancer screening: a systematic review of methods. *BMJ*. 2015;350:g7773. URL: https://www.bmj.com/content/350/bmj.g7773

85. Bae JM. Overdiagnosis: epidemiologic concepts and estimation. Epidemiol Health. 2015;37:e2015004. Available from: https://www.e-epih.org/journal/view.php?number=770

86. Srivastava S, Hanash S, Pan S, et al. Cancer overdiagnosis: a biological challenge and clinical dilemma. *Nat Rev Clin Oncol*. 2019;16(6):349-358. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC8819710/

87. Dunn BK, Woloshin S, Xie H, et al. Cancer overdiagnosis: a challenge in the era of screening. *J Natl Cancer Cent*. 2022;2:235-242. URL: https://pubmed.ncbi.nlm.nih.gov/36568283/

88. O'Brien MER, Borthwick A, Rigg A, et al. Mortality within 30 days of chemotherapy: a clinical governance benchmarking issue for oncology patients. *Br J Cancer*. 2006;95(12):1632-1636. doi:10.1038/sj.bjc.6603498. URL: https://pubmed.ncbi.nlm.nih.gov/17160081/

89. El-Tamer MB, Ward BM, Schifftner T, Neumayer L, Khuri S, Henderson W. Morbidity and mortality following breast cancer surgery in women: national benchmarks for standards of care. *Ann Surg*. 2007;245(5):665-671. URL: https://pubmed.ncbi.nlm.nih.gov/17457156/

90. Bleyer A, Welch HG. Effect of three decades of screening mammography on breast-cancer incidence. *N Engl J Med*. 2012;367(21):1998-2005. doi:10.1056/NEJMoa1206809. URL: https://pubmed.ncbi.nlm.nih.gov/23171096/

91. Hernán MA, Robins JM. Per-Protocol Analyses of Pragmatic Trials. *N Engl J Med*. 2017;377(14):1391-1398. doi:10.1056/NEJMsm1605385. URL: https://pubmed.ncbi.nlm.nih.gov/28976864/

92. International Council for Harmonisation (ICH). *ICH E9(R1): Statistical Principles for Clinical Trials—Addendum on Estimands and Sensitivity Analysis in Clinical Trials*. 2019 (Step 4). URL: https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf

93. U.S. Food and Drug Administration. E9(R1) Statistical Principles for Clinical Trials: Addendum—Estimands and Sensitivity Analysis. 2021. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e9r1-statistical-principles-clinical-trials-addendum-estimands-and-sensitivity-analysis-clinical

94. Schulz KF, Altman DG, Moher D; for the CONSORT Group. CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomised Trials. *BMJ*. 2010;340:c332. doi:10.1136/bmj.c332. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC2860339/

95. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 Explanation and Elaboration: Updated Guidelines for Reporting Parallel Group Randomised Trials. *BMJ*. 2010;340:c869. doi:10.1136/bmj.c869. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC2844943/

96. Hopewell S, et al. **CONSORT 2025 statement: updated guideline for reporting randomised trials.** Lancet. 2025. URL: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736%2825%2900672-5/fulltext

97. McCoy CE. **Understanding the Intention-to-treat Principle in Randomized Controlled Trials.** *West J Emerg Med*. 2017;18(6):1075-1078. doi:10.5811/westjem.2017.8.35985. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC5654877/

98. Santos-Gallego CG, et al. **Per-Protocol Versus Intention-to-Treat in Clinical Trials.** *J Am Heart Assoc*. 2022;11:e025561. doi:10.1161/JAHA.122.025561. URL: https://www.ahajournals.org/doi/10.1161/JAHA.122.025561

99. Bretthauer M, et al. **Effect of Colonoscopy Screening on Risks of Colorectal Cancer and Related Death.** *N Engl J Med*. 2022;387:1547-1556. doi:10.1056/NEJMoa2208375. URL: https://pubmed.ncbi.nlm.nih.gov/36214590/

100. **Trends in the hidden burden of cancer in an autopsy-based cohort (latent colorectal cancer prevalence example).** 2026. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC12878435/

101. ClinicalTrials.gov. **The Northern-European Initiative on Colorectal Cancer (NordICC) (NCT00883792).** URL: https://www.clinicaltrials.gov/study/NCT00883792

102. Mukherjee S. **"The Catch in Catching Cancer Early."** *The New Yorker*. June 16, 2025.
URL: https://www.newyorker.com/magazine/2025/06/23/the-catch-in-catching-cancer-early

103. **Mukherjee S. (related New Yorker newsletter piece) "Siddhartha Mukherjee on the Promises and Perils of Early Cancer Detection."** *The New Yorker*. June 17, 2025.
URL:
https://www.newyorker.com/newsletter/the-daily/siddhartha-mukherjee-on-the-promises-and-perils-of-early-cancer-detection