Call Their Bluff: How to Read Medical Evidence Without a Medical Degree

A *Food Phoenix* companion to the MRI Informed Consent workbook, and to any health decision where someone tells you "the evidence supports it."
The language of evidence is built, half the time, to keep you out. This is the key. You don't need a medical degree, just a handful of questions and the names of the things to look for.

PART 1

Did It Actually Help Anyone? (The Only Outcomes That Count)

When someone says "the evidence supports it," you're entitled to ask: evidence of what? A study only counts if it measured something that would matter to you, a real human, not a number on a screen.

A scan or a marker changing is not a benefit. These are the things that are:

- Did people actually **live longer**? And counted as *all* deaths, not "deaths from this one disease." Disease-specific death counts rest on a guess written on a certificate, are frequently wrong, and are easy to massage. All-cause mortality can't be fudged: you're either alive or you're not.
- Were they **less disabled**, more able to do what a day asks of them?
- Did their **fatigue** ease, their **pain** settle, their **brain fog** lift, measured with a real test, not waved away?
- Did their **sleep, mood and sense of wellbeing** improve, by their own report? (Yes, even happiness has validated measures.)
- Did **fewer end up in hospital or intensive care**?
- Could they **keep working**, or get back to it?
- Did they end up on **fewer prescriptions**, or more? More pills to manage new problems is a harm, not a neutral. Increased prescriptions for psychiatric medications are usually a red flag that all is not well.
- Did it spare them a **more invasive test or procedure**, rather than triggering a cascade of them?
- Were people **followed long enough** to catch harms that surface months or years later, and did it help people **like you**, or only the fittest, who are the least likely to be hurt?

If a study measured none of those, it hasn't shown the thing that matters to you, however impressive it sounds.

You don't have to wade through the jargon to use this. When you're handed "studies show it helps," ask one question:

> *Helps with what, exactly, that I'd notice in my own life? Did people live longer or better, and compared to what?*

If the honest answer is "it makes the picture clearer" or "the numbers looked better," that's a surrogate, not a result. Most of the time, no one has checked the rest.

PART 2

The Real Measuring Sticks (and What to Google)

You don't need to learn any of this. The point is that proper, validated tools exist for every one of these. So when a study, or a doctor, claims something helped, you can check whether they actually measured it with one of these, or just asserted it. If a claim about fatigue, pain or brain fog isn't backed by a recognised measure, that's your cue to be sceptical.

What you want measured	In plain terms	Real tools to look for	What to type into a search
Fatigue	Is the exhaustion actually better?	Fatigue Severity Scale (FSS), FACIT-Fatigue, Chalder Fatigue Scale	`"fatigue severity scale"`, `"validated fatigue questionnaire"`
Pain	Less pain, nerve pain too	Brief Pain Inventory, Visual Analogue Scale; for nerve pain, painDETECT or DN4	`"brief pain inventory"`, `"neuropathic pain questionnaire"`

What you want measured	In plain terms	Real tools to look for	What to type into a search
Brain fog and thinking	Reaction time, processing speed, sustained attention	The sensitive ones are the reaction-time and "fitness-for-duty" impairment tests used on pilots, soldiers and shift-workers: Psychomotor Vigilance Task (PVT), ANAM, Symbol Digit Modalities Test, Digit Symbol Substitution Test, Trail Making, Stroop, Cogstate, CANTAB. Self-report: PROMIS Cognitive Function, Cognitive Failures Questionnaire, Perceived Deficits Questionnaire. A "normal" MoCA or MMSE proves little here: those are dementia screens, far too blunt for brain fog.	`"psychomotor vigilance task"`, `"fitness-for-duty cognitive test"`, `"symbol digit modalities test"`
Sleep	Actually sleeping and rested?	Pittsburgh Sleep Quality Index (PSQI), Insomnia Severity Index; objective: actigraphy	`"Pittsburgh sleep quality index"`, `"insomnia severity index"`
Gut symptoms	Bloating, pain, bowel changes	GI Symptom Rating Scale (GSRS), IBS Symptom Severity Score, Bristol Stool Chart	`"GI symptom rating scale"`, `"IBS severity score"`

What you want measured	In plain terms	Real tools to look for	What to type into a search
Disability and function	Daily life, work, caring for people	WHODAS 2.0, Health Assessment Questionnaire; in cancer, ECOG or Karnofsky	`"WHODAS 2.0"`, `"activities of daily living scale"`
Mood, anxiety, depression	Measured properly, not waved away	PHQ-9 (depression), GAD-7 (anxiety), HADS, DASS-21	`"PHQ-9"`, `"GAD-7"`, `"HADS"`
Happiness and wellbeing	Yes, this is measurable too	WHO-5 Wellbeing Index, Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS), Satisfaction With Life Scale, the ONS4 wellbeing questions	`"WHO-5 wellbeing index"`, `"Warwick-Edinburgh wellbeing scale"`
Quality of life overall	Your whole life, your own rating	EQ-5D, SF-36 or SF-12, WHOQOL	`"EQ-5D"`, `"SF-36 quality of life"`
Work and livelihood	Could you keep earning?	Work Productivity and Activity Impairment (WPAI)	`"work productivity activity impairment"`

Two umbrella terms worth knowing. A **PROM** (patient-reported outcome measure) is any validated questionnaire that asks *you* how you're doing, the gold standard for the things only you can feel. And the **minimal clinically important difference (MCID)** is the smallest change big enough for a real person to notice. A result can be "statistically significant" and still sit below the MCID: real, but too small to feel. Search those two, and you'll punch well above your weight.

PART 3

Sciencese, Translated

The words they'll use, and what they actually mean.

All-cause mortality

Did fewer people die? The honest measure, and the one that can't be fudged, because you're either alive or you're not.

Cause-specific mortality

Did fewer die *of this particular thing*? Sounds precise, often isn't: it rests on a guess written on a death certificate, which is frequently wrong and easy to nudge. Treat with suspicion.

Surrogate (or proxy) endpoint

A stand-in for a real result: a clearer scan, a tidier blood number. Only worth something if it actually leads to you living longer or better, which often goes unchecked.

Absolute vs relative risk

"Cuts your risk by half" (relative) can mean going from 2 in 1,000 to 1 in 1,000 (absolute). Relative numbers look dramatic. Always ask for the absolute.

Statistically significant

Unlikely to be pure chance. It does *not* mean large, or that it matters to you. A tiny, useless difference can be "significant."

Number needed to treat

How many people must have the thing for one to benefit. If it's 200, that's 199 who took the risk for nothing.

Number needed to harm

The mirror of the above: how many are exposed for one to be harmed. Ask for both, and the real trade comes into focus.

Lead-time bias

Finding something earlier can make survival *look* longer without anyone living a day longer. You just knew about it for more of the time.

Length-time bias

Screening catches the slow, lazy tumours far more often than the fast, aggressive ones, simply because the slow ones sit around longer waiting to be found. So the screened group fills up with the least dangerous cases, and the survival figures look wonderful, for reasons that have nothing to do with the screening helping anyone.

Overdiagnosis

Finding something real that was never going to harm you, then treating it anyway, so you carry the risks of treatment for no gain.

Survivorship bias

Judging a treatment by the people still standing at the end, when the ones it failed, or killed, have quietly dropped out of view. If you don't survive the treatment, you don't fill in the follow-up questionnaire, and the average of whoever's left looks rosy.

Competing risks

When one bad outcome hides another by getting there first. If a treatment causes an early death or an early loss, that person can never go on to be counted as one of the failures it was meant to prevent. The harm erases the evidence of the harm.

Loss to follow-up (attrition)

The people who drop out before a study ends. They're rarely a random sample: the sickest, the harmed and the disillusioned tend to leave first, so counting only who's left flatters the result. Always ask what happened to the ones who vanished.

Intention-to-treat (ITT)

The honest way to count: everyone who started in a group stays in that group's results, whatever became of them, dropouts and all. The alternative ("per-protocol") lets a treatment quietly bin its own worst results. If a trial won't report ITT, ask why.

Publication bias

Flattering results get published; disappointing ones quietly don't. So "the studies show" can mean "the ones they chose to print show."

Conflict of interest

Who paid for the research, and who profits if it's positive. It shapes what gets studied, published and spun, so always ask.

RCT vs observational study

A randomised controlled trial assigns who gets what, so it can show cause. An observational study just watches who happened to get what, so it's easily fooled when the sicker, or the healthier, sort themselves into groups. "Early treatment did better" is nearly always observational.

Contraindication

A specific reason you, in particular, shouldn't have it.

Incidentaloma

Something a scan finds by accident that was never troubling you, and may now cost you dearly in tests, worry, and treatment.

Contrast agent (GBCA)

Gadolinium-based dye, injected to make certain things show up brighter.

Linear vs macrocyclic

Two cage designs for the gadolinium. Macrocyclics are sold as the "safer" newer ones; our survey didn't find them gentler.

Retention

Gadolinium staying in the body after the scan (bone, brain and elsewhere), sometimes for years.

Hit-and-run (in pharmacology)

An exposure that sets off lasting harm even after it's left your body, so "it's cleared" doesn't mean "you're fine." A recognised idea in drug safety, and the reason "you'll have peed it out" settles far less than it sounds.

NSF (nephrogenic systemic fibrosis)

A serious scarring condition linked to contrast in people with kidney failure. Often used to imply everyone else is safe, which doesn't follow.

Sensitivity and specificity

How good a test is at catching the thing, and at not crying wolf. No test is perfect at both.

Sciencese

The dialect that makes simple things sound unquestionable and keeps you from asking. You're always allowed to ask for a translation.

PART 4

The Words for What's Being Done to You

These aren't evidence terms. They're the names for the other half of what a bad medical experience does to you: the being disbelieved, the being dismissed, the being harmed by the very people you were told to trust. Naming a thing takes back some of its power. Not one of these is a failing in you.

Testimonial injustice

When your account of your own body is believed less because of who you are: a woman, tired, "anxious," without a diagnosis. Not a flaw in you. A named bias, first set out by the philosopher Miranda Fricker.

Epistemic injustice

The wider family it belongs to: being wronged specifically as a *knower*, treated as though you can't be a reliable witness to your own experience.

Hermeneutical injustice

When medicine has no agreed word yet for what's happening to you, so you can't make yourself understood, and that gets mistaken for nothing being wrong. The gap is in their vocabulary, not your body.

Medical gaslighting

Being nudged to doubt your own perception ("it's stress," "the scan's clear, so you're fine") until you start to distrust yourself. Sometimes deliberate, more often a system reaching for the easy label.

Betrayal trauma

The particular wound of being harmed or let down by the very people and institutions you were meant to be able to trust with your health. It bites harder than harm from a stranger, because it breaks something you relied on. Jennifer Freyd's related idea of *institutional betrayal* fits this even more precisely.

Stockholm syndrome

The survival instinct to defend, even feel loyalty toward, the very system that's harmed you, because turning on something you depend on can feel more dangerous than staying loyal to it. Not weakness or naivety: it's the mind protecting you. Worth naming so you can spot it, in yourself and in the people urging you to keep trusting.

Floccinaucinihilipilification

The act of judging something to be worthless. One of the longest words in the dictionary, for one of the most ordinary things done to patients: your side effect, your missed malnutrition, your toxicity symptoms, weighed, found worthless, and waved away. Now, at least, you can name it.

PART 5

Who's Missing From the Count

Almost any treatment can be made to look good with one quiet move, and it's the move fewest people think to check for. Every result you're handed is a fraction: the good outcomes on top, everyone who got counted on the bottom. You don't have to touch whether the treatment actually works to shift that fraction. You just have to change who ends up in the count. And the people most likely to fall out of it are exactly the ones whose story would have spoiled the headline.

It happens three ways.

Overdiagnosis

Incorrectly identifying a problem that isn't there.

Survivorship Bias

Focusing on successes while ignoring failures.

Missing Denominator

Drawing conclusions from incomplete data sets.

The dead stop being failures (Survivorship bias)

Follow a treatment only through the people still standing at the end, and you've tilted the table before you've counted a thing. The ones it harmed the most, the ones who died, who dropped out too unwell to carry on, who never came back for the final check, take their bad results out of the door with them. Whoever's left averages out beautifully. This is why "the patients who completed treatment did well" tells you almost nothing. Of course they did. They're the ones it didn't sink.

The cured were never ill (Overdiagnosis in screening)

This one hides inside screening, and mammography is the cleanest example. Start scanning healthy women, and you turn up a great many small tumours that were never going to spread, never going to shorten a life, some that would have sat there quietly forever. Treat them all, and every one of those women joins the ranks of the "lives saved by early detection." But they were never going to die of it. You haven't saved them; you've relabelled them, and handed some of them surgery, radiation and fear they'd have been better off without, occasionally shortening the very lives you're claiming to rescue. The survival statistics soar. The number of women actually dying barely moves, and there's no guarantee which direction they edge towards. That gap, between "more survivors" and "fewer deaths", is where the whole illusion lives. It's why you ask for deaths, not survival rates, and never "cases caught early."

The lost are never asked (Missing denominator)

The subtlest version, and the one I keep running into in my own reading of the trials. Take the famous trial that put folic acid in pregnancy on the map. Alongside the celebrated drop in neural tube defects sat something quieter: a cluster of excess pregnancy losses in the folic acid group. Here's the snag. A pregnancy lost early can never be counted as a baby born with a neural tube defect. So if the drug caused even some of those losses, and some of those lost pregnancies would have been affected, the harm didn't merely sit alongside the benefit. It manufactured part of it, by removing the very cases that would have landed in the "failed" column. On my own reading, if only a handful of that excess were drug-caused, the celebrated advantage narrows sharply, and the story changes entirely. The triumph rests on nobody ever asking what really happened to those pregnancies. The metrics look magnificent precisely because the missing cases stay missing.

Notice what all three share. Nobody faked a number. Every figure in the paper is real. The sleight of hand, where there is one, is in the denominator: in who quietly went missing before the counting began.

The single question that catches all of it: *Who's not in the final count, and would the answer change if they were?* Ask what became of the dropouts, the deaths, the losses, the ones "excluded from analysis." Ask whether the treatment could have caused the very thing that removed them. Ask for all-cause mortality instead of survival rates, and intention-to-treat instead of "the patients who completed the course." If the people missing from the bottom line are the ones the treatment hurt, then the beautiful result along the top was never really there at all.

PART 6

How "The Studies Show" Gets Made

The evidence you're handed didn't fall from the sky. A great deal of it is manufactured, by people with a product to sell, long before your doctor (who's busy and trusts the guidelines) ever repeats it to you. Knowing the tricks is how you stop taking "the studies show" as the end of the conversation.

The Barnum effect

Vague, confident claims that feel personally true, the same cold read a fairground psychic runs. When a billion-pound industry runs it on a whole profession with glossy charts, it gets a respectable name: the standard of care.

p-hacking (the Texas sharpshooter)

Fire a machine gun at a barn, then paint the bullseye around the tightest cluster of holes and call yourself a marksman. Slice the trial data enough ways, and some small, convenient number will turn up positive. Watch for a drug that missed the outcome that matters but "significantly improved" a sliver you'd never heard of (usually a surrogate endpoint, see above).

Ghostwriting and KOLs

The paper carrying a famous professor's name was sometimes drafted by an agency the drug company paid, with the "key opinion leader" lending their signature and their university's credibility. The byline says Stanford; the copy came from marketing.

Astroturfed guidelines

Fake grassroots. Fund the societies, committees and campaigns that write the clinical guidelines your doctor is bound to follow, and the "independent consensus" becomes a sales funnel. The opioid "pain as the fifth vital sign" campaign is the textbook case: the bodies that pushed it were funded by the maker of the drug.

Direct-to-consumer (disease-mongering)

The article that looks like a neutral medical encyclopedia, quietly widening what counts as illness, talking down diet and lifestyle as "unproven," and steering you to "ask your doctor about" a brand by name. You arrive at the appointment primed to request the very drug they were lobbied to prescribe.

Your Toolkit at a Glance

The takeaway card. Screenshot it, print it, take it in with you.

Six parts, one purpose: to give you standing in any conversation about evidence. You already have what this takes, not a medical degree, just the right questions and the nerve to ask them.

Ask what was actually measured

Did people live longer, feel better, stay out of hospital? A surrogate isn't a result.

Check whether a validated tool was used

Fatigue, pain, brain fog, wellbeing all have recognised measures. No measure, no claim.

Ask for absolute numbers, not relative ones

"Halves your risk" sounds enormous. "2 in 1,000 down to 1 in 1,000" is the same fact, undressed.

Ask who's missing from the count

The dropouts, the deaths, the losses: where did they go? The denominator is where the illusion lives.

Know the words for what's being done to you

Testimonial injustice. Medical gaslighting. Betrayal trauma. Named things, now yours to name back.

Ask who paid, and who benefits

Industry funding shapes what gets studied, printed and spun. Not conspiracy: documented, named practice.

"The evidence supports it" is where the conversation starts, not where it ends. You're always allowed to ask the questions underneath it:

> **Which evidence? Measuring what? In whom? Funded by whom? And where did the missing ones go?**

That's the question they hope you won't ask. Now it's yours.

A Food Phoenix companion guide.

Nourished and Flourishing.