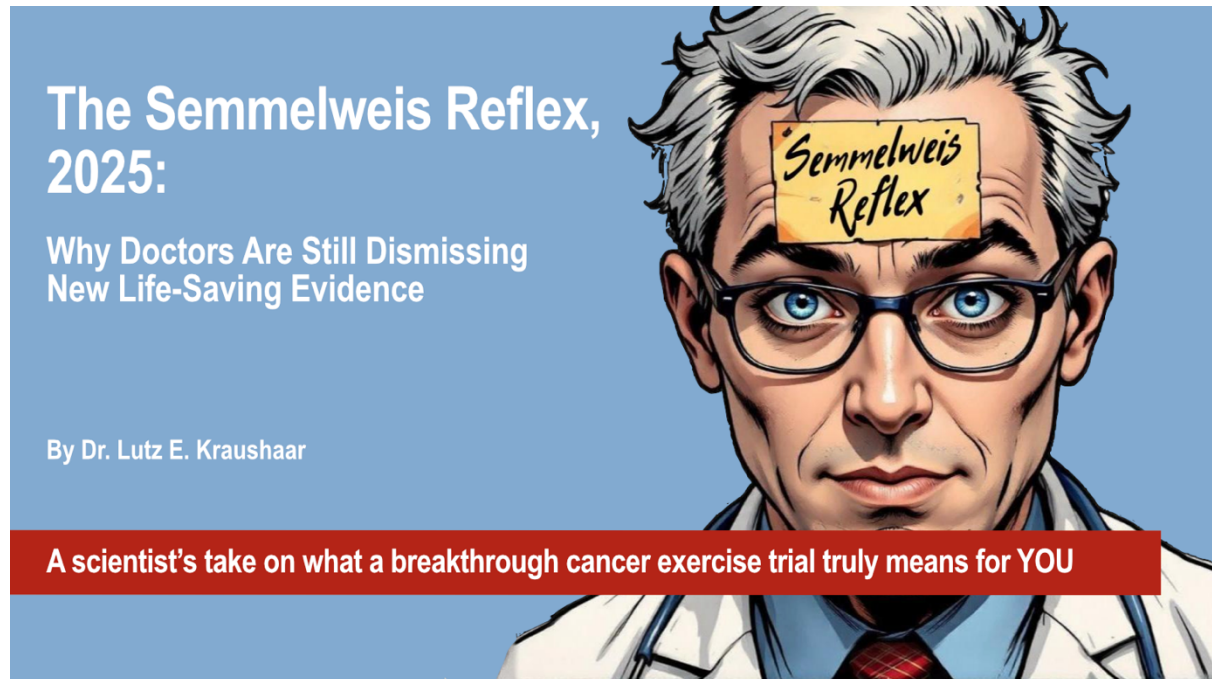


The Semmelweis Reflex, 2025: Why Doctors Are Still Dismissing Life-Saving Evidence

A breakthrough colon cancer exercise study, its flawed critique, and how to protect yourself from bad science



What Is New?

Exercise is an ignored rival to cancer drugs. That's the verdict of the first large-scale phase 3 clinical trial about the effects of exercise on cancer recurrence and survival.

Why It Matters

The medical establishment is known for its reflex-like tendency to dismiss new evidence because it contradicts established norms or beliefs. It's called the Semmelweis reflex. It has a history of making more, rather than fewer, dead people. And you shouldn't suffer from that.

Your Takeaway

A blow-by-blow rebuttal of a prominent critic's dismissal of what could save your life and improve its quality. And a practical guide to critically screening the critics.

An Echo from 1847

It is June 1847: Ignaz Semmelweis demonstrates to his peers and patients that women's deaths from childbed fever (one in five births ended with the mother's death) could be cut by 90% to two in 100. His method: sanitizing the surgeon's hands *before* laying them on a woman in labor.

A simple intervention; no negative side effects.

In the mid-1800s, nobody understood the concept of disinfection (bacteria hadn't made an entrance on the stage of science).

The obstetrician guild's alpha male of the time, Charles Meigs, dismisses Semmelweis' discovery with a compelling rationale:

“Doctors are gentlemen, and gentlemen's hands are clean.”

He and his minions scorned handwashing, continued killing women, and drove the ‘savior of mothers’ out of his job and into madness [1].

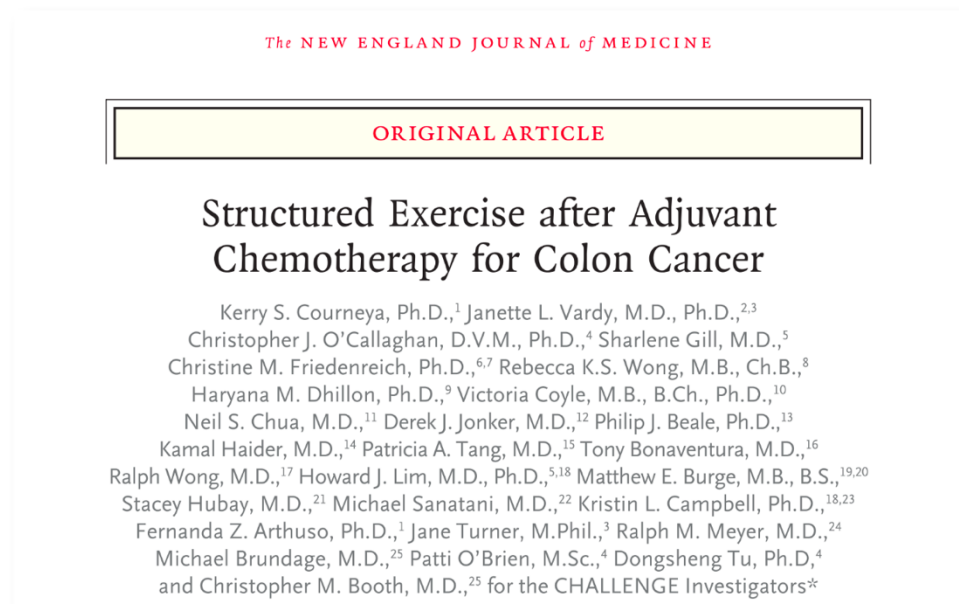
It was 40 years later when Louis Pasteur's discovery vindicated Semmelweis' method.

Fast forward to June 2025: A Modern-Day Challenge

Imagine your doctor offering you a new intervention after colon cancer surgery that could, in his words, slash your risk of death by roughly 40%, and has no side effects. Not exactly a head-scratcher.

When that “intervention” turns out to be exercise, you'll probably have one of two reactions: (A) “Great,” because driving yourself out of breath has been part of your lifestyle all along, or (B) “Aargh,” because exercising is as alien a concept to you as philanthropy is to Putin.

Your doctor backs up his prescription with the data of a large study (appropriately named the CHALLENGE trial, published just 4 days prior to this writing), the first phase III randomized clinical trial (RCT) investigating the effects of an exercise program on survival in patients like you [2].



A simple intervention; no negative side effects.

Then, within 48 hours, a blistering critique of this study insists that exercise is not a cancer therapeutic and that

“you have to let go of your love of great stories to see the details.”

To you, it might sound credible given that this critique appeared in the newsletter of a respected group of doctors.

To me, it represents the *Semmelweis effect*, a fundamental incomprehension of methods and statistics, potentially spawning a health threat. First, because it may discourage *you* from adopting exercise, and second, because in the grand scheme of things, it stands in the way of what we should aim for in medicine: less death.

It serves as a teaching aid and eye-opener for common misinterpretations and misconceptions that I encounter frequently and that *you* should be aware of too, not only in the context of cancer studies.

The CHALLENGE Trial: A Summary

First, The Facts: What the CHALLENGE Trial Did and Found

The Goal: CHALLENGE was a phase 3 randomized controlled trial (RCT). It was designed to uncover whether a structured, long-term exercise program could improve disease-free survival (DFS) and other outcomes in colon cancer patients who had already completed surgery and adjuvant chemotherapy within the 2–6 months prior to enrollment.

This trial is outstanding not only for being the first truly randomized exercise trial in a cancer patient cohort and its rigorous methods, but also for its

- 889 participants (that’s a lot)
- 3-year intervention (that’s long)
- 15-year trial duration (that’s extremely long)

The image below illustrates who participated.

The CHALLENGE Trial Participants

Colon Cancer Patients Who Had:

- *Complete surgical removal (resection) of either stage III or high-risk stage II adenocarcinoma*
- *Completed their post-surgery chemotherapy within the previous 2 to 6 months*
- *Exercised less than 150 minutes per week at a moderate-to-vigorous intensity before joining the study*
- *A basic level of fitness to undertake an exercise program*

Enrollment:

From 2009 through 2024, a total of 889 patients from 55 centers (mostly in Canada and Australia, but also in the US, the UK, and France) were randomly assigned in a 1:1 ratio to one of two groups for a 3-year period:

Health-Education Group (Control—444 patients):

- Received general health-education materials that promoted physical activity and healthy nutrition.
- Continued with standard cancer surveillance (follow-up monitoring).

Structured Exercise Program Group (Intervention Group—445 patients):

- Received the same as the control group.
- **PLUS:** Support from a certified physical activity consultant for 3 years. Patients had to commit to compulsory behavioral support sessions, which were delivered biweekly (first year) or monthly (2nd and 3rd year), either in person or remotely via phone/video.
- **Plus Exercise Goal:** To increase recreational aerobic exercise from their baseline by at least 10 MET-hours per week during the first 6 months and then to maintain or further increase this amount for the remaining 2.5 years.

Baseline Snapshot:

The median age of patients was 61 years (ranging from 19 to 84). 51% were women, and 90% had stage III disease. At the start, patients reported engaging in about 11.5 MET-hours per week of moderate-to-vigorous physical activity.

Brief on METs

MET is shorthand for ‘metabolic equivalent’, which is a measure of energy expenditure. One MET is equivalent to the energy spent at rest. A brisk walk, for example, carries the energy cost of 4 METs, whereas jogging increases that to around 10 METs, depending on intensity, of course.

And just so that you don’t get any wrong ideas, no matter how intense your between-the-sheets performance, you won’t break the 3 METs bar; and your average Friday night run-through is on par with the exhausting activity of watering flowers [3].

CODE	METS	MAJOR HEADING	SPECIFIC ACTIVITIES
13050	2.0	self care	showering, toweling off, standing
14010	2.8	sexual activity	active, vigorous effort
14020	1.8	sexual activity	general, moderate effort
14030	1.3	sexual activity	passive, light effort, kissing, hugging

You can [download](#) Ainsworth's 17-page list of physical activities and their MET cost, covering virtually everything humans do (which makes you wonder how the hell they tested those in the image above, and why you were never asked to serve as a volunteer participant).

Back to the subject. 10 MET-hours is the equivalent of jogging for 1 hour (or three and a half hours of intense ...you know what).

What Was Found?

The image below summarizes the results.

The CHALLENGE Trial Outcome

Disease-Free Survival (DFS)

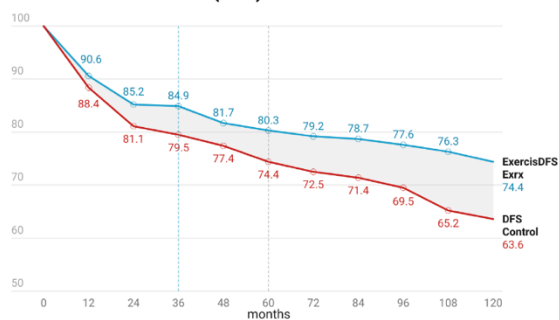


Chart: Dr. Lutz E. Kraushaar • Source: Courneya KS, et al. Structured Exercise after Adjuvant Chemotherapy for Colon Cancer. N Engl J Med 2025 • Created with Datawrapper

Overall Survival

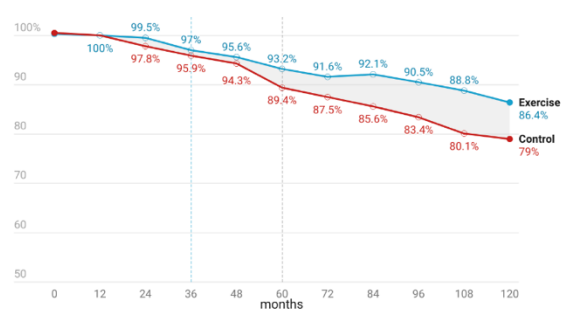


Chart: Dr. Lutz E. Kraushaar • Source: Courneya KS, et al. Structured Exercise after Adjuvant Chemotherapy for Colon Cancer. N Engl J Med 2025 • Created with Datawrapper

Courneya KS, et al. Structured Exercise after Adjuvant Chemotherapy for Colon Cancer. N Engl J Med 2025

After a median follow-up of 7.9 years:

- **Disease-free survival (DFS, the primary outcome)** was significantly longer in the exercise group, with a hazard ratio of 0.72. This means the exercise group had a 28% lower risk (hazard) of suffering an event compared to the health-education group.
- **Overall survival (OS, a secondary outcome)** was also longer in the exercise group, with a 0.63 hazard ratio. That's a 37% lower risk of dying from any cause (the roughly 40% I mentioned earlier).
- **Patient-Reported Physical Functioning** (another secondary outcome) improved significantly more than in the control group.

Intervention group participants also spent more MET hours on exercise (5.2–7.4 MET hrs) per week, and they increased their fitness level (estimated maximum oxygen uptake) by up to 2.7 ml per kg per minute compared to control participants.

They also increased their 6-minute walk distance (between-group difference of 13 to 30 meters).

Interestingly, there were minimal differences in body weight or waist circumference between the groups, suggesting weight loss wasn't the primary driver of other observed benefits.

In Conclusion:

“A 3-year structured exercise program, started soon after adjuvant chemotherapy for colon cancer, resulted in significantly longer disease-free survival and findings consistent with longer overall survival.”

That's a fair assessment. But let's also acknowledge that, at first blush, the increases in fitness and walking distance sound a little meager. I'll address these points when we dissect the critic's comments. Before we do that, let's look at the media's take on this study.

The Media Exuberance

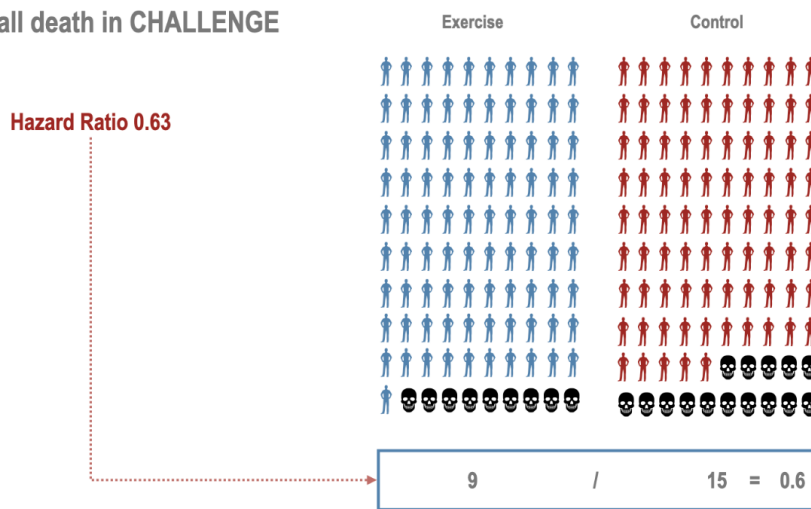
Within 24 hours of the study being published, dozens of media outlets euphorically celebrated its close-to-40 % reduction in the risk of overall death (dying from any cause, not just cancer).



Earth to media: Why don't you tell us that it's just a 7% absolute risk reduction, from 83% survivors in the control group to 90% in the intervention group?

The control group had 66 deaths within the 8-year follow-up; the exercise group had 41. The image below translates those numbers into groups of 100 each: 9 deaths in the exercise group vs. 15 in the control group.

Overall death in CHALLENGE



In other words, for every ten non-exercising colon cancer patients, roughly 8 survived, whereas 9 exercisers made it to the finish line alive. You judge for yourself whether these numbers sound more sober than the 40% risk reduction.

Now let's dissect our critic's objections.

The Critic's Pushback

The critique was published in the Substack Newsletter "Sensible Medicine," which presents itself as

"a shared site featuring the voices of leading physicians, scientists, and thinkers".

I have no beef with the publication itself; many articles that appear there are good reads. Not this one, though.

Sensible Medicine

Exercise is Great but It's Not a Cancer Drug

The CHALLENGE trial ignited social media with its great story of exercise as a cancer therapeutic. But liking a conclusion is not a reason to stop thinking.



JOHN MANDROLA
JUN 02, 2025

Its author raises seven points to justify his overall dismissal of the study results and the media's enthusiasm for its story:

“This is a classic case when you have to let go of your love of great stories to see the details.”

So let's have a look at those details.

Critique 1: “It’s Just Not Plausible”

First, before looking at the methods, the 37% reduction in all-cause mortality is implausible and rivals many proven cancer therapies. For example, this is similar to the mortality reduction with Trastuzumab (Herceptin) in HER2+ breast cancer—a revolutionary finding.

Critic’s Argument: “The 37% reduction in all-cause mortality is implausible.”

The critic offers no evidence for his suggested implausibility of the effect size.

In reality, there is ample evidence in favour of exercise's benefits from observational studies and small intervention trials. For example, an up to 40% risk reduction for cancer death among stage-3 colon cancer patients [4], and improved survival of a similar magnitude in small intervention trials [5]. Adding just 10 weekly minutes of vigorous intensity activity associated with a 50% reduced cancer incidence and mortality in another study [6].

The fact that association never proves causation was what prompted the Canadian Cancer Trials Group to launch the CHALLENGE trial in the first place.

Your takeaway

Plausibility is indeed an acid test for the credibility of any health claim. Ignorance of evidence, however, does not equate absence of evidence.

Critique 2: “The Exercise Didn’t Actually Do Anything”

Second, if you posit exercise provides a massive mortality benefit, there should be some effects of exercise. CHALLENGE reported almost no between-group differences in typical exercise parameters. There was zero differences in body weight, waist circumference and a mere 30 meters longer distance in the 6-minute walk test.

Critic’s Argument: “Almost no between-group differences in typical exercise parameters.”

The assertion is a selective interpretation and largely inaccurate.

First, Significant Increase in Actual Physical Activity (MVPA):

The exercise group achieved and maintained larger improvements in moderate-to-vigorous physical activity (MVPA), ranging from 5.2 to 7.4 MET-hours per week, than those in the health-education group. That translates into roughly 1.5 to 2 hours of brisk walking per week.

Second, Improved Cardiorespiratory Fitness (Predicted VO₂max):

Compared to the control group, the exercise group increased their predicted maximum oxygen uptake volume by 1.3 to 2.7 ml/kg/min. While these improvements are small, they are within the range observed for exercise interventions in different cancer patient populations [7].

Third: Body Weight and Waist Circumference—Context is Key

The critic is correct that there were “Minimal between-group differences... for body weight or waist circumference.” The study authors themselves acknowledge this, suggesting that:

“weight loss is an unlikely explanation for the observed effects of exercise on cancer outcomes”.

This lack of weight change doesn't negate the benefits of exercise. On the contrary, it strengthens the argument that the positive impacts of exercise on cancer outcomes might be mediated through known exercise-stimulated mechanisms *other than* weight loss, such as:

- Reduced inflammation
- Improved immune function
- Enhanced insulin sensitivity (which can occur without weight loss)
- Direct effects on the tumor microenvironment or circulating factors

In essence, the critic is selectively focusing on a few outcomes that didn't change dramatically (or that he deems “mere”) while ignoring clear evidence that the exercise intervention *did* lead to significant (albeit moderate) increases in physical activity and cardiorespiratory fitness—the very goals of such a program and established predictors of better health.

Critique 3: “The Study’s Timeline Looks Suspicious”

Third, the Kaplan-Meier survival curves for disease-free survival begin separating at 12 months while the death curves take 4 years to separate. I am not a cancer doctor, but a tiny difference in exercise dose (as evidenced by the lack of objective measures) is not enough to reduce cancer recurrences that rapidly. This finding suggests suboptimal randomization, which is not surprising given the fact the ambitious complicated trial took 15 years to enroll.

Critic’s Argument: DFS curves separate at 12 months, OS curves at 4 years. He implies this is problematic.

It isn't. This pattern is common and logical in oncology, specifically observed in the colon cancer scenario [8]. Interventions that reduce cancer recurrence show a separation in DFS curves before a separation in OS curves. That's because cancer recurrence is often the first "event" in DFS, and patients may live for a considerable time afterwards before the event impacts overall survival.

Critic's Argument: Tiny Difference in Exercise Dose" & Inability to Reduce Recurrences "That Rapidly":

He reiterates his claim of a "tiny difference in exercise dose (as evidenced by the lack of objective measures)" and asserts this couldn't reduce recurrences rapidly.

As discussed for his second point, this is a misrepresentation of the study's findings.

He also fails to provide an argument why the observed increases in exercise volume and fitness should *not* impact recurrence risk within a year.

Exercise is known to affect various systems fairly quickly:

- **Immune Function:** Enhanced immune surveillance can target micrometastases.
- **Inflammation:** Reductions in chronic inflammation, a promoter of cancer progression, can occur.
- **Metabolic Health:** Improvements in insulin sensitivity and related growth factor pathways can influence cancer cell proliferation.

Patients entered the study 2–6 months after completing adjuvant chemotherapy. This is a critical period where residual disease might exist (a few cancerous or precancerous cells that escaped the scalpel). An intervention that bolsters the body's defenses or creates a less favorable environment for cancer cell survival could plausibly show effects on recurrence relatively early.

Now, here is one thing we have to acknowledge:

As I explained in the introduction, the relative risk reduction of 40% is the pimped up version of the 7% absolute increase in survival chance, from 83% to 91%. The media, but also researchers, are infatuated with relative risk's shinier numbers for an obvious reason. They attract far more eyeballs than their sober sibling: the absolute change of risk or benefit.

Your takeaway:

In health articles that tout increases or decreases of outcome, scout for the absolute numbers, that is, the difference between affected participants in each group, and compare them using simple arithmetic.

Critic's Argument: "Suboptimal Randomization" Due to Timing, "Dose," and Enrollment Period.

He links the curve separation timing and his (incorrect) assessment of the exercise dose to suggest "suboptimal randomization," additionally blaming the 15-year enrollment period.

This is a large leap in logic and a serious accusation that lacks supporting evidence.

Here is the evidence against his claim:

Evidence of Successful Randomization:

The most crucial evidence against “suboptimal randomization” is the baseline characteristics of the two groups. If baseline characteristics are balanced, the randomization did its job. And it did. While the study’s baseline comparison table does not specify the p -values for the 35 biomarkers, I calculated them all: no significant difference between the two groups for any of the biomarkers.

Randomization Method: The study used a “dynamic minimization procedure” for randomization, stratifying by trial center, disease stage, BMI, and fitness performance status. This is a sophisticated method designed to ensure balance across key prognostic factors, especially in trials with long durations or multiple centers.

Randomization procedures

The purpose of randomization is to create groups that are as similar as possible at the start of the trial. Researchers want to be sure that any differences in observed outcomes at the end of the trial are more likely due to the intervention itself, rather than pre-existing differences between the groups.

Typically, researchers use digitally glorified coin-flip methods or variations thereof to ensure random participant allocation. In trials such as CHALLENGE, which last over long periods and are conducted at different centers, this method may miss its purpose. That’s where dynamic minimization procedures come in.

For each new patient eligible for the trial, a minimization algorithm considers the current distribution of patients already randomized across the different treatment groups, specifically looking at the balance of the pre-defined stratification factors.

It calculates how assigning the new patient to any group would affect the balance of these factors. The procedure then assigns the new patient to the group that would maintain the balance.

Dynamic minimization is exactly what the CHALLENGE trialists did, supervised by an independent Data and Safety Monitoring Committee (DSMC). This randomization process worked perfectly fine over the entire 15-year enrollment period, as evidenced in the balanced baseline data. It’s the hallmark of successful randomization.

The critic finds the outcome of the study surprising and alleges it is due to a fundamental flaw in the randomization procedure for which he can offer no proof, because no such flaw is evident. That’s not the way scientific appraisal works.

You can do better than that!

Your takeaway:

Look at a study's baseline comparison (typically table 1). Search for the presence of significant differences (expressed as p-values). If you find some, it doesn't necessarily mean that randomization was flawed. Differences can emerge just by chance.

Researchers have some tools in their shed to account for such differences, such as statistically controlling for these differences. They will then mention that in the methods or results section.

Critique 4: “But Not Everyone Followed the Protocol!”

Fourth, poor adherence to the exercise regimen further reduces plausibility. Nearly half of the patients in the exercise program did not complete the treadmill protocol at three years and a third did not complete the 6-minute walk test. These patients were included in the intention-to-treat findings—and would have the effect of reducing between-group differences in exercise.

My response to his point: “Exactly!”

Inclusion of the non-adherent exercise-group participants waters down the difference between this group and the control group.

Think of it this way: If your soccer team has 7 great players and 4 like me, you probably still win against another team that has ten like me, but with a smaller difference in goals than if you had 11 crack footballers on your side.

The same thing happened in CHALLENGE (and in every other trial with less than 100% adherence). The effects of exercise become *less visible precisely because* of the non-adherent participants' inclusion. **Which means, exercise probably has an even larger effect on DFS and OS than what we see in the data.**

This type of analysis is called intention-to-treat (ITT), meaning that regardless of whether a participant adhered to the intervention protocol or not, their data are included in the group data. In statistics lingo, ITT biases the results towards the null. That is, it reduces the apparent difference between groups and makes it harder to find a statistically significant effect. This is the preferred conservative approach to reflect real-world effectiveness as closely as possible.

To switch from the predefined ITT to a per-protocol analysis would have been a violation of the research principle *not to* switch analytic methods only because the results fail to meet expectations. Here is why: Switching means jeopardizing your randomization. The subgroup of adherent participants may be significantly different from the control group, which is exactly what you wanted to avoid with randomization.

Your takeaway:

Scout for the terms intention-to-treat and per-protocol. If you sniff a switch, typically from ITT to PP, be very skeptical about the results. More often than not, it's like putting lipstick on a pig. The pig is still a pig.

The declining adherence to supervised sessions over the 3-year period (e.g., from 79% for mandatory supervised exercise in Phase 1 down to 44% for recommended supervised sessions in Phase 3) is very common in long-term behavioral interventions. Life can get in the way, motivation can wane, and the perceived need for structured sessions may diminish.

The critic's entire point 4 defeats his own arguments about the meager effects of exercise.

Critique 5: "They Messed Up the Statistics"

Fifth, the authors originally designed the trial to detect differences at 3 years. This required 380 events to have sufficient statistical power. Due to slow recruitment and a slower-than-expected event rate they changed to a five-year analysis. Yet they still had far less than the expected primary endpoint events (224 vs 380). This reduces statistical power and raises the possibility of false positive findings—which is consistent with the biological implausibility. What's more, the KM curves show most of their separation *after* 3 years. A stronger paper would have included the pre-specified 3-year results, which may have been non-significant.

Critic's Argument: The Original design for 3-year analysis needed 380 events. Changed plan due to slow recruitment/event rate. Still had fewer events (224 vs. 380):

"This reduces statistical power and raises the possibility of false positive findings..."

Utter nonsense. If anything, the scenario raises the possibility of false *negative* findings, that is, not detecting an effect of exercise when there really is one.

A study isn't "underpowered" for a result it successfully detects with statistical significance.

Let's look at what really happened (because the critic's presentation is incorrect).

The study transparently explains *why* the plan was adapted: slow recruitment and a lower-than-expected pooled event rate. This is a common issue in long-term trials. The researchers realized in their interim analysis that by their originally planned clinical cut-off date, they would have accrued less than the anticipated 380 events.

They then moved the goal posts, making the data cutoff dates conditional on having at least 200 events, which would give them a power of 80%. 'Power of 80 %' means that the chance of detecting a significant difference when there truly is one is 80%. To set power at 80% is common practice in biomedical research.

They achieved 224 events, with a DFS hazard ratio (HR) of 0.72 at $p=0.02$. That is, a 28% relative risk reduction of any event in the exercise group compared to the control group. In the trialists' own words:

“Exercise significantly reduced the relative risk of disease recurrence, new primary cancer, or death by 28%.”

“False Positives” accusation: A p -value of 0.02 means that *if the null hypothesis (no effect) were true*, there's only a 2% chance of seeing such a result (or more extreme) just by chance. This is below the conventional 5% chance ($p=0.05$) used in biomedical research.

Your takeaway:

In intervention studies, look out for the power calculation. Nowadays, ethics boards and journal editors want to see it. If a study doesn't have enough power to begin with, it's a waste of resources (why do all the work when you know in advance that you are unlikely to detect an effect). If it's overpowered, it's a waste of resources, too (it could do with fewer participants, consumables, and work), and it unnecessarily exposes the excess participants to possible intervention side effects.

Critic's Argument: “KM curves show most of their separation after 3 years. A stronger paper would have included the pre-specified 3-year results, which may have been non-significant.”

Later Separation is Not a Flaw: The study notes that DFS curves began separating at “about 1 year.” If the benefit becomes more pronounced over longer follow-up (median 7.9 years), that's a feature showing sustained or accumulating benefit, not a flaw. Many oncological interventions show increasing separation of KM curves over time.

Reporting 3-Year Data: Point taken. For maximum transparency, researchers could have explicitly reported the 3-year DFS figures, given that it was part of the initial design consideration.

However, as mentioned in the previous section, researchers defined the final analysis to happen after a certain number of events had accrued, leading to a longer follow-up. The absence of a specific 3-year statistical comparison doesn't invalidate the significant findings from the primary analysis based on 224 events and longer follow-up.

Critique 6: “It's Just a Placebo Effect”

Sixth, while the first five concerns relate to the conduct and the design of the trial, there is also the inherent challenge of strategy trials: different attention in the two groups. In CHALLENGE, the structured exercise group received an incredible amount of intervention in both behavioral modification and exercise. This makes performance bias highly likely, as evidenced by the large differences in the quality of life questionnaires.

Critic's Argument: The exercise group got an “incredible amount of intervention” (behavioral mod + exercise), making “performance bias highly likely, as evidenced by the large differences in the quality of life questionnaires”.

Yes, the exercise group received a comprehensive, multi-component intervention. **That was the entire point of the study**—to test if the “structured exercise program” (the intervention) improves outcomes compared to just receiving health-education materials alone.

The definition of performance bias according to the Cochrane Handbook [9]:

“biases that arise when there are deviations from the intended interventions”

Deviations, not *differences* between interventions. A difference in intervention is the independent variable under investigation. And there were no obvious deviations. The study's supplemental material documents every aspect of the formalized intervention that the researchers applied to every participant.

Moreover, patient-reported quality of life was a secondary *outcome* marker that was *expected* to differ between the groups *because* of the intervention.

It's inherent in RCTs (be it a drug, therapy, or behavioral intervention trial) that the intervention group receives something different and usually more intensive than the control group. This isn't a flaw; it's the experimental design. The critic is trying to frame a fundamental aspect of the CHALLENGE behavioral intervention trial as its fatal flaw.

A caveat: As readers of the study, or of any study, we can never entirely rule out performance bias. What the critic does here is to take a subjectively disagreeable study result (the effect of exercise on DSF) as proof of performance bias.

Your takeaway:

As far as performance bias is concerned, look out whether the groups were treated equally (Besides the Intervention).

Bias is a real and present threat to almost every study. To discover it is not an easy task. It's not just for kicks that the Cochrane Handbook devotes an entire chapter to the different flavors of biases and how they can threaten the interpretation of any study.

Critique 7: “These Results Don't Apply to Real People”

Seventh, there are serious challenges in external validity or generalizability of the CHALLENGE trial. The difficulty in enrolling patients (it took 15 years) speaks to the complexity and intensity of the behavioral and exercise program. The authors don't tell us how many were screened to enroll these 900 patients. I suspect it was a lot. What's more, enrolled patients were young (age 61) non-obese and performed well on baseline measures of function. Even if you accepted the results as presented it would apply to a fraction of patients with colon cancer.

Critic’s Argument: “Difficulty in Enrolling Patients (15 years) Speaks to Complexity/Intensity”:

I don’t dispute that. Enrolling nearly 900 patients into a 3-year intensive behavioral intervention trial across 55 international centers is a massive undertaking.

Put yourself into the shoes of the researchers and think about your task: Find patients who meet all the specific inclusion criteria (stage III or high-risk stage II, completed chemo 2–6 months prior, exercising *less than* 150 min/week, able to do a treadmill test, *and willing to commit to a 3-year program*). This isn’t like recruiting for an Ozempic trial.

Critic’s Argument: “Authors Don’t Tell Us How Many Were Screened”

Yes, it would be nice to have a flowchart showing how many were screened and *not* selected for which reason. But our focus should be on the characteristics of the enrolled patients and whether they represent a group for whom the intervention is relevant, rather than speculating on those not enrolled without data.

Critic’s Argument: “Enrolled Patients Were Young (Age 61), Non-Obese, and Performed Well on Baseline Measures”

“Young”: A little over one-third of patients were >65. A median age of 61 is very common in cancer intervention trials.

“Non-obese” is a misrepresentation. Almost 40% of the participants were obese, and only one quarter were in the normal-weight range (BMI < 25).

“Performed well on baseline measures”: yes, they did, because that was one inclusion criterion. But these people weren’t fit. Their fitness level was below the 40th percentile (estimated maximum oxygen uptake) of the reference population [10].

“Apply to a Fraction of Patients with Colon Cancer”:

The patient characteristics (age, BMI including overweight individuals, baseline activity levels indicating they were not already exercisers, and reasonable but not elite fitness) define a population that is indeed relevant and for whom an exercise intervention is plausible and, quite obviously, highly beneficial.

Selecting a well-defined group is how medical knowledge advances—establishing efficacy in that group first, before exploring the intervention’s wider applicability. To dismiss it because it doesn’t immediately apply to everyone is to misunderstand this process.

My Verdict: Beyond the Details, Why This Matters

The CHALLENGE trial provides crucial, life-saving evidence for the benefits of exercise. First and foremost, as an adjunct intervention to standard colon cancer care. But it goes much further than that.

There is a lot of room to amp up the volume and intensity of exercise compared to the intervention group. Not only to treat cancer but to prevent it in the first place. From

cardiometabolic diseases, we know that the dose-response relationship between exercise and cardiometabolic health is linear over a wide range from sedentary to athletic training levels.

And we have every reason to believe that this relationship will hold in the cancer context. Why? Because cancer is increasingly recognized as a metabolic disease. After all, a hallmark and intervention target of cancer cells is their altered metabolic profile [11].

Hence, the CHALLENGE study is probably just scratching the surface of what *you* can achieve and realize.

Will the exercise “pill” reduce your risk of cancer to zero? Certainly not. Cancer and childbed fever are incomparable on so many levels. What is comparable is the Semmelweis reflex:

The reflex-like tendency to dismiss new evidence because it contradicts established norms or beliefs.

It is still very much alive.

Your Final Prescription

Ignore the critic. Recognize exercise for what it quite possibly is: your most important lifestyle medicine.

Cited References

[1] De Costa CM. “The contagiousness of childbed fever”: a short history of puerperal sepsis and its treatment. *Med J Aust* 2002;177:668–71. doi:<https://doi.org/10.5694/j.1326-5377.2002.tb05004.x>.

[2] Courneya KS, Vardy JL, O’Callaghan CJ, Gill S, Friedenreich CM, Wong RKS, et al. Structured Exercise after Adjuvant Chemotherapy for Colon Cancer. *N Engl J Med* 2025. doi:10.1056/NEJMoa2502760.

[3] AINSWORTH BE, HASKELL WL, HERRMANN SD, MECKES N, BASSETT DRJR, TUDOR-LOCKE C, et al. 2011 Compendium of Physical Activities: A Second Update of Codes and MET Values. *Med Sci Sports Exerc* 2011;43:1575–1581. doi:10.1249/MSS.0b013e31821ece12.

[4] Brown JC, Ma C, Shi Q, Fuchs CS, Meyer J, Niedzwiecki D, et al. Physical Activity in Stage III Colon Cancer: CALGB/SWOG 80702 (Alliance). *J Clin Oncol* 2023;41:243–54. doi:10.1200/JCO.22.00171.

[5] Markozannes G, Becerra-Tomás N, Cariolou M, Balducci K, Vieira R, Kiss S, et al. Post-diagnosis physical activity and sedentary behaviour and colorectal cancer prognosis: A Global Cancer Update Programme (CUP Global) systematic literature review and meta-analysis. *Int J Cancer* 2024;155:426–44. doi:10.1002/ijc.34903.

[6] Ahmadi MN, Clare PJ, Katzmarzyk PT, del Pozo Cruz B, Lee I-M, Stamatakis E. Vigorous physical activity, incident heart disease, and cancer: how little is enough? *Eur Heart J* 2022;43:4801–14. doi:10.1093/eurheartj/ehac572.

[7] del-Rosal-Jurado A, González-Sánchez M, Cuesta-Vargas AI. Effect of therapeutic exercise on peak oxygen consumption in oncological population: a systematic review with meta-analysis. *Support Care Cancer* 2024;32. doi:10.1007/s00520-024-09004-1.

[8] Sargent DJ, Wieand HS, Haller DG, Gray R, Benedetti JK, Buyse M, et al. Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: Individual patient data from 20,898 patients on 18 randomized trials. *J Clin Oncol* 2005;23:8664–70. doi:10.1200/JCO.2005.01.6071.

[9] Higgins J, Green S (Sally E, Cochrane Collaboration. Assessing risk of bias in included studies *Cochrane handbook for systematic reviews of interventions*. *Cochrane Handb. Syst. Rev. Interv.*, 2017, p. 1.

[10] Whaley MH, Brubaker PH, Otto RM, Armstrong LE. *ACSM's Guidelines for Exercise Testing and Prescription*. 2006.

[11] Collier HA. Is cancer a metabolic disease? *Am J Pathol* 2014;184:4–17. doi:10.1016/j.ajpath.2013.07.035.

A breakthrough colon cancer exercise study, its flawed critique, and how to protect yourself