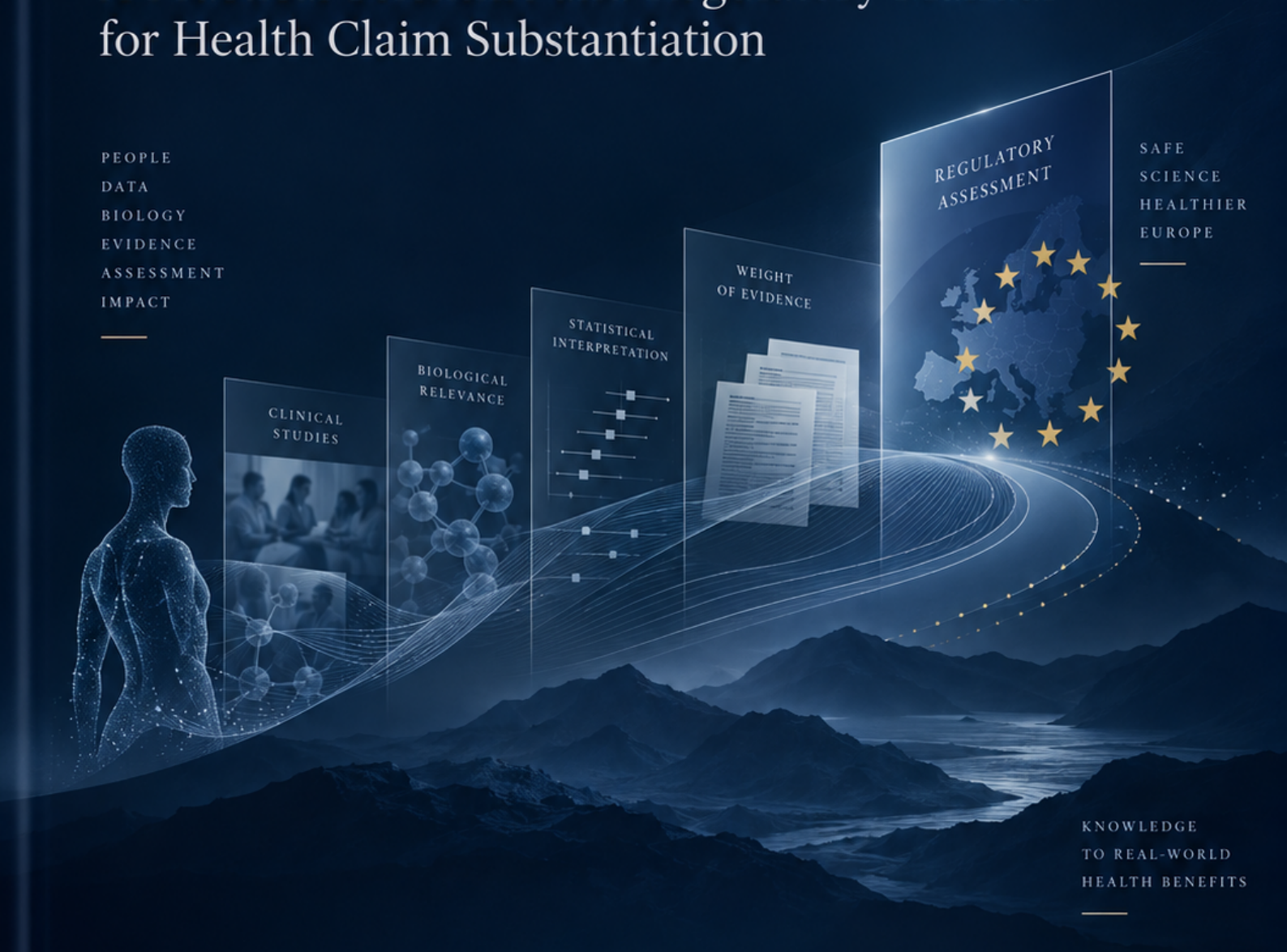
SCIENCE FOR A HEALTHIER TOMORROW

EVIDENCE
PEOPLE
REGULATION
BETTER HEALTH

FROM CLINICAL EVIDENCE TO EFSA

A Practical Scientific and Regulatory Manual for Health Claim Substantiation



Human Intervention Studies · Biological Relevance · Statistical Interpretation · Weight of Evidence · Regulatory Strategy

INTABIOTECH Scientifc & Regulatory Affairs — 2026

FROM CLINICAL EVIDENCE TO EFSA

A Practical Scientific and Regulatory Manual for Health Claim Substantiation

Human Intervention Studies • Biological Relevance • Statistical Interpretation • Weight of Evidence • Regulatory Strategy

José M. López, BSc, MD, PhD, LLD (C) · Technical Director

Belén Amboade Vázquez (Attorney at Law) · Head Regulatory

INTABIOTECH Scientific & Regulatory Affairs

Technical Edition · 2026

Editorial Note

This manual was developed for technical, clinical, scientific and regulatory teams working on food and food-constituent health claims in the European Union. It is deliberately written from the perspective of development decisions: what should be settled before a study begins, what makes a human intervention study persuasive, and what usually fails when scientific evidence is translated into an EFSA dossier.

The expression “nutraceutical” is used occasionally because it remains common in industry. It is not, however, the legal category that governs the assessment. Under EU law, the relevant object will generally be a food, food supplement, food constituent, nutrient or other substance, and the scientific question is whether the proposed relationship between that food or constituent and a beneficial physiological effect is substantiated under Regulation (EC) No 1924/2006.

A further distinction is essential. EFSA assesses the scientific substantiation of a proposed health claim and adopts a scientific opinion. EFSA does not itself authorise the commercial use of the claim; authorisation takes place within the EU procedure involving the European Commission and Member States. Throughout this manual, “EFSA-ready”, “EFSA-oriented” and similar expressions refer to scientific and procedural readiness for assessment, not to a guarantee of a favourable opinion or subsequent authorisation.

SCOPE AND USE

This is a scientific and regulatory working manual, not legal advice and not a substitute for checking the legislation, EFSA guidance and procedural requirements in force on the date of submission. EFSA expressly advises applicants to verify that they are using the latest guidance before applying.

How to use this manual

The main text is intended to be read as a development narrative. It moves from the regulatory question to the construction of evidence, and from individual trial results to the totality of evidence presented in the dossier. The operational annexes are different: they are working documents. They can be adapted into project templates, CRO briefing documents, internal review forms or standard operating procedures.

The practical tools in the annexes—including the readiness score, fatal-flaw screen and Red Team review—are INTABIOTECH decision-support instruments. They are not EFSA scoring systems and should never be described as such.

A note on terminology

This edition uses British English and follows the terminology of EFSA wherever possible: “food/constituent”, “claimed effect”, “target population”, “conditions of use”, “human efficacy studies”, “biological plausibility”, “totality of the evidence” and “cause-and-effect relationship”. Statistical terms are used in their conventional scientific sense, with emphasis on estimates and uncertainty rather than binary labels of “positive” and “negative”.

Executive Perspective

The weakest moment in many health-claim programmes occurs surprisingly early: a study is commissioned before the company has defined the regulatory question it actually needs the study to answer. From that point onwards, even excellent execution can only optimise the wrong design.

A clinical trial can be scientifically respectable, publishable and statistically significant, yet still add little to an EFSA health-claim dossier. The reason is simple. EFSA is not asked whether an ingredient changes something measurable. It is asked whether a sufficiently characterised food or constituent causes a beneficial physiological effect in the intended population under the proposed conditions of use, and whether the totality of the evidence is persuasive enough to substantiate that relationship.

That changes the order in which development decisions should be made. The intended food–health relationship must come before the endpoint. Endpoint qualification must come before the power calculation. Biological relevance should be considered before results are known. The commercial dose should be reconciled with the clinical dose before recruitment starts. The study population should be chosen with generalisability in mind, not simply because a CRO can recruit it quickly. And Regulatory Affairs must be involved before the first participant is enrolled, not after the statistical report has been signed.

THE GOVERNING IDEA

A p-value can tell you something about compatibility with a statistical model. It cannot tell you whether the endpoint is appropriate, whether the magnitude matters physiologically, whether the tested material matches the marketed product, whether the result is reproducible, or whether the evidence is relevant to the proposed claim.

The practical consequence is that the best EFSA development strategy is not the strategy that produces the greatest number of “significant” findings. It is the strategy that progressively removes uncertainty from the food–effect relationship. The strongest programmes are built so that, by the time the pivotal study begins, the product, claim, endpoint, population, dose, duration and statistical estimand already form one coherent scientific proposition.

Contents

Content Index

1. The Question EFSA Actually Assesses

A health-claim programme should begin with a causal proposition, not with a list of biomarkers.

In the food sector, product development often begins with an ingredient. A company has a botanical extract, a peptide fraction, a microbial preparation, a fermentation product or a new formulation and naturally asks what benefits it might support. That is a legitimate commercial starting point. It is not, however, the correct starting point for an EFSA health-claim programme.

For regulatory development, the first question should be narrower and less comfortable: what exactly is the food–health relationship that the applicant wants EFSA to assess? Once the answer is explicit, the rest of the programme becomes easier to discipline. If the answer remains vague, the development plan tends to accumulate endpoints and exploratory analyses in the hope that one of them will eventually support a marketable story.

EFSA's general scientific guidance is built around three linked questions. First, is the food or constituent sufficiently defined and characterised? Second, is the claimed effect sufficiently defined, and is it a beneficial physiological effect for the proposed target population that can be measured in vivo in humans? Third, taking the totality of the available evidence into account, has a cause-and-effect relationship been established between consumption of the food/constituent and the claimed effect?

REGULATORY PRINCIPLE

If characterisation fails, efficacy becomes difficult to attribute. If the claimed effect is not a recognised or adequately justified beneficial physiological effect, a technically impressive trial may still be irrelevant to the claim. Causality is assessed only after those foundations are in place.

A useful internal discipline is to require the project team to express the regulatory hypothesis in one sentence before protocol development begins:

> *Consumption of [defined food/constituent], at [dose and conditions of use], by [target population], produces [defined beneficial physiological effect], demonstrated by [primary outcome variable and method].*

The sentence is intentionally unforgiving. Each bracket forces a decision that will later appear somewhere in the dossier. If the product cannot be defined, if the effect is still being described in marketing language, if the population is unspecified, or if nobody can agree on the primary outcome, the pivotal study is not ready.

Why "nutraceutical" is an industry term rather than the regulatory object

The term "nutraceutical" is useful shorthand in commercial and scientific discussion, but EU food law does not create a general regulatory category by that name. Depending on the product, the legal object may be a food, food supplement, nutrient, other substance or food constituent. For health-claim purposes, the scientific substantiation exercise is governed principally by Regulation (EC) No 1924/2006 and its implementing framework.

This matters because the evidence has to follow the legal proposition. A dossier is not an abstract demonstration that an ingredient is "healthy". It is an application concerning a defined claim, a defined food or constituent, a defined population and defined conditions of use.

The applicant should think claim-first, not assay-first

A common development sequence is to commission a broad trial, measure a generous panel of biomarkers and ask Regulatory Affairs to formulate a claim after the results are available. That sequence may be attractive for exploratory research, but it is a poor confirmatory strategy. It encourages post-hoc claim selection, increases

multiplicity and makes it difficult to distinguish a pre-specified hypothesis from an attractive finding discovered in the data.

For an EFSA-oriented programme, the logic should be reversed: characterise the food; define the physiological relationship; identify the target population; formulate the claim concept; qualify the endpoint; establish what magnitude of change would matter; and only then design the trial.

2. Three Different Ideas: Statistical Significance, Biological Relevance and Regulatory Relevance

Confusing these three questions is one of the most persistent causes of weak health-claim evidence.

Statistical significance answers a narrow question

A p-value is generated within a statistical model. It says something about how compatible the observed data are with a specified null hypothesis under that model. It does not measure the size of the effect, the probability that the hypothesis is true, the physiological importance of the result, the likelihood of replication, or the regulatory acceptability of the endpoint.

This is why EFSA's Scientific Committee has long advised against treating the identification of statistical significance as the principal objective of analysis. Effect estimates and interval estimates—particularly confidence intervals—carry information that a binary threshold cannot provide.

Biological relevance asks whether the change matters

Biological relevance concerns the nature, magnitude and physiological meaning of the observed change. A sufficiently large trial can make a very small difference statistically significant. Conversely, a study with inadequate precision can fail to achieve conventional statistical significance even when the point estimate is compatible with an effect that would matter physiologically.

The Scientific Committee's 2011 opinion on statistical significance and biological relevance is especially important for development planning because it recommends defining, where possible, the nature and size of the biological change regarded as relevant before the study begins. That prospective judgement should influence the hypothesis, the power calculation and the later interpretation of the confidence interval.

DO NOT INVENT A UNIVERSAL THRESHOLD

EFSA does not operate with a generic rule that a 5%, 10% or 20% improvement is "biologically relevant". Relevance is endpoint-specific, population-specific and claim-specific. Where a validated meaningful-change threshold exists, use it. Where it does not, explain the scientific basis for interpreting magnitude rather than creating false precision.

Regulatory relevance is more demanding again

An effect can be biologically real and still be insufficient for the proposed health claim. The endpoint may not represent the claimed benefit. The population may not support extrapolation. The intervention may have been tested at a dose that cannot be commercialised. The material used in the study may not be equivalent to the material for which the claim is sought. Or the finding may conflict with other pertinent human evidence.

Regulatory relevance therefore combines endpoint validity, magnitude, precision, causality, reproducibility, product equivalence, population relevance and realistic conditions of use. Statistical significance is one component of that assessment, not its organising principle.

3. Characterisation: Evidence Begins with the Product

Before asking whether something works, the assessor must know what "something" is.

Product characterisation is sometimes treated as a technical section that can be completed after the clinical programme. That is a strategic mistake. Characterisation determines whether evidence generated with one material can legitimately be used to substantiate a claim for another.

The problem is particularly acute for botanicals, fermentation products, probiotics, postbiotics, complex polyphenol preparations, peptide mixtures and other materials whose biological activity may depend on composition, manufacturing conditions or strain-level identity. Two products can share a common commercial description and still be materially different for the purposes of scientific substantiation.

What should be locked before the pivotal study

The degree of characterisation required depends on the nature of the product, but the development team should normally have a defensible description of identity, relevant composition, manufacturing process, analytical specification and stability before the clinical material is released. Botanical identity and plant part, extraction solvent, chemical fingerprint, marker compounds, microbial strain identity, active or characteristic constituents, molecular properties and batch-to-batch variability may all become relevant.

The analytical methods used to verify the defining characteristics should be fit for purpose. A specification has little regulatory value if the constituent used to define the product cannot be measured reliably. Likewise, a clinical batch should not be treated as representative of the commercial material merely because it came from the same supplier.

PRACTICAL TEST

Ask whether an independent assessor, using the manufacturing and analytical information in the dossier, could understand why the material used in the human studies is representative of the material intended to carry the claim. If the answer is uncertain, the programme has a product-equivalence problem before it has an efficacy problem.

Stability is part of exposure

If a relevant constituent degrades during the intervention period, the labelled dose may not be the dose that participants actually receive. This is not a peripheral quality issue. It affects causal interpretation. Stability and storage conditions deserve particular attention for live microorganisms, oxidation-sensitive lipids, enzymes, botanicals and other preparations in which activity can change materially over time.

For pivotal studies, it is often useful to retain samples of the clinical batch and to document release testing, storage, packaging, shelf-life and, where scientifically justified, end-of-study verification. Clinical efficacy belongs to the tested material, not simply to the name printed on the sachet.

4. Define the Benefit Before You Choose the Biomarker

A measurable response is not automatically a beneficial physiological effect.

One of the most consequential decisions in the programme is made before the first laboratory assay is ordered: what exactly is the claimed effect? Commercial language tends to be broad—"metabolic support", "healthy ageing", "antioxidant power", "immune optimisation", "brain health". Regulatory development requires a narrower physiological proposition.

The correct sequence is to define the function, explain why a favourable change in that function would be beneficial for the intended population, and only then decide which outcome can demonstrate the change. This ordering prevents the common mistake of choosing a responsive biomarker first and trying to attach regulatory meaning to it later.

Direct outcomes, validated surrogates and mechanistic biomarkers

Not all biomarkers occupy the same evidentiary position. A direct functional outcome may address the claim itself. A validated surrogate or recognised risk factor may, in the appropriate context, provide strong regulatory evidence. A mechanistic biomarker can be valuable because it explains how an effect might occur, but it may not demonstrate that the claimed benefit has occurred. Exploratory markers are useful for hypothesis generation and mechanistic discovery, but they should not be retrospectively promoted to pivotal status simply because they changed.

Evidence role	What it can show	Typical regulatory use
Direct physiological outcome	The claimed function itself	Potentially pivotal when valid and appropriately measured
Recognised surrogate / risk factor	A validated proxy or recognised risk factor	Potentially pivotal in the relevant claim context
Mechanistic biomarker	A pathway, exposure or intermediate response	Supportive evidence for plausibility and interpretation
Exploratory biomarker	A signal not yet established as an outcome for the claim	Hypothesis generation; rarely confirmatory on its own

The antioxidant example: why assay sophistication is not enough

The antioxidant field illustrates the distinction particularly clearly. An intervention can produce highly significant changes in assays such as FRAP, ORAC, TEAC or TRAP and still fail to demonstrate the claimed physiological benefit. EFSA's claim-specific guidance distinguishes global measures of antioxidant capacity from evidence that relevant biological molecules have actually been protected against oxidative damage.

The development question is therefore not "which antioxidant assay gives the largest response?" It is "what molecule or physiological process is claimed to be protected, what constitutes oxidative damage in that context, and which validated measurement can demonstrate that protection in humans?"

LDL cholesterol and blood pressure: examples of context that matters

By contrast, changes in LDL cholesterol and blood pressure can have clear regulatory meaning in appropriate claims. EFSA has recognised reductions in LDL cholesterol within the normal range as beneficial physiological

effects and has also addressed LDL cholesterol as a disease-risk factor in the context of certain Article 14 claims. Maintenance of normal blood pressure is likewise a recognised beneficial physiological effect, with the exact design and interpretation depending on the wording and scope of the claim.

These examples should not be treated as shortcuts. The point is methodological: the regulatory status of an endpoint depends on the physiological relationship being claimed. The same biomarker may have different evidentiary meaning in different contexts.

5. Biological Relevance: Magnitude, Direction and Precision

A result should be interpreted as an estimated effect with uncertainty, not as a traffic light based on p = 0.05.

Effect size comes before the adjective "significant"

Consider two hypothetical trials measuring the same outcome. Study A estimates a treatment difference of −0.2 units with a 95% confidence interval of −0.38 to −0.02. Study B estimates −1.4 units with a 95% confidence interval of −2.0 to −0.8. Both results may satisfy a conventional threshold for statistical significance, but they do not provide the same information about magnitude, precision or physiological importance.

The first line of a regulatory interpretation should therefore state what happened and with what uncertainty. The p-value can follow. This ordering is not cosmetic: it changes the way teams think about results.

Prospective biological relevance

Where science allows it, the project team should write down before the study what magnitude of change would be considered meaningful and why. That value may be informed by previous trials, meta-analyses, recognised clinical or functional thresholds, physiological data, relevant EFSA opinions or well-founded pilot data. If a precise threshold cannot be justified, the plan should instead describe how magnitude and uncertainty will be interpreted.

This prospective exercise protects the programme against two opposite errors. The first is celebrating a tiny effect because a large sample made it statistically detectable. The second is discarding a potentially important effect because an imprecise study produced p = 0.06 rather than p = 0.04.

What confidence intervals add

A confidence interval permits three questions to be asked at once. Is the direction of the estimated effect compatible with benefit? What range of effect sizes remains plausible under the model? And is the estimate precise enough to distinguish a relevant effect from a trivial one?

The distinction between "absence of evidence" and "evidence of absence" often becomes visible only when the interval is examined. A narrow interval centred on a trivial effect can be informative evidence against a meaningful benefit. A wide interval around a promising point estimate is simply uncertain.

A BETTER REPORTING SENTENCE

"The pre-specified analysis estimated a between-group difference of X, with a 95% confidence interval from Y to Z. The magnitude was [consistent / not consistent] with the prospectively defined range considered biologically relevant." This is more informative than "the product was effective because p < 0.05".

6. Designing Human Intervention Studies for a Regulatory Purpose

The pivotal study should be capable of answering the intended claim question without reinterpretation after the fact.

The primary endpoint should be chosen for regulatory meaning

A primary endpoint should satisfy four conditions simultaneously: the measurement should be valid; the change should represent a beneficial physiological effect; the endpoint should align with the intended claim; and the study should be adequately powered for the corresponding treatment effect. A technically advanced assay that fails one of those conditions may be unsuitable as the pivotal outcome.

Secondary outcomes should be organised according to their role rather than their novelty. Supportive physiological outcomes, mechanistic measurements, safety variables and exploratory analyses can all enrich the programme, but the hierarchy should be explicit before unblinding.

Power should be based on an effect that matters

Sample-size calculations should not be reverse-engineered from budget alone, nor built around an implausibly large effect simply to keep the trial small. The assumed effect and variance should be justified from the best available evidence, and attrition should be handled realistically. The relevant question is not “how many participants can we afford?” but “how many participants are required to estimate the effect with sufficient precision to answer the regulatory question?”

An underpowered study can leave a potentially meaningful effect unresolved. An overpowered study can make a negligible effect look compelling if interpretation is reduced to a p-value. Both errors begin with poor planning.

Acute and chronic claims require different time scales

A post-prandial glycaemic response can be studied acutely. A claim about sustained lipid management, maintenance of muscle function or longer-term physiological support cannot usually be answered by the same time scale. Duration should follow the kinetics of the proposed effect, the behaviour of the endpoint and the wording of the claim.

The protocol should state why the chosen intervention period is long enough for the proposed mechanism to operate, the outcome to respond and, where relevant, the response to stabilise or demonstrate persistence. “Eight weeks because that is standard” is not a physiological rationale.

A clinical study is not a fishing expedition

Modern nutritional studies can measure dozens or hundreds of variables. That scientific richness becomes a confirmatory weakness when the protocol has no real hierarchy. A study that measures everything and later promotes whichever endpoint produces the most attractive result cannot carry the same evidentiary weight as a study built around a pre-specified primary question.

Exploratory science remains valuable. The correct approach is to label it honestly. A post-hoc signal can generate the hypothesis for the next trial; it should not be rewritten as if it had been the pivotal hypothesis all along.

7. Statistical Strategy: From p-Values to Estimation and Robustness

A regulatory analysis should be reproducible, pre-specified and resistant to plausible alternative assumptions.

The between-group treatment effect is usually the question

One of the most persistent analytical errors in food intervention studies is to show that the active group changed significantly from baseline while the placebo group did not, and then conclude that the active treatment was superior. That inference is invalid. The relevant analysis generally concerns the difference attributable to treatment—often estimated directly using an ANCOVA, mixed model or another design-appropriate method.

"Significant in one group" and "not significant in the other" is not itself a test of the difference between groups. The statistical model should be selected to answer the actual estimand.

Multiplicity: the cost of asking many confirmatory questions

When multiple primary or confirmatory hypotheses are tested, the analysis should address multiplicity prospectively. The appropriate method depends on the structure of the hypotheses: hierarchical testing, gatekeeping or a suitable adjustment procedure may be appropriate. There is no reason to adjust every exploratory analysis, but there is every reason to be explicit about which findings are confirmatory and which are not.

The worst practice is to decide the hierarchy after seeing the results. That turns a statistical analysis plan into a record of choices rather than a safeguard against them.

Missing data are part of the scientific problem

Missing observations are rarely neutral. Participants may discontinue because of tolerability, lack of perceived benefit, loss of motivation or events related to the outcome. The SAP should therefore define the primary approach to missing data and appropriate sensitivity analyses before unblinding. The purpose is not to select the method that produces the best p-value, but to test whether the conclusion survives plausible alternative assumptions.

Subgroups require interaction, rationale and restraint

A negative overall trial often produces tempting subgroup findings: benefit in women but not men, benefit above a certain age, benefit in participants with low baseline status, benefit in one centre. These findings may be scientifically interesting, but they do not automatically establish efficacy in the subgroup. Credibility depends on pre-specification, biological rationale, sample size, interaction testing, consistency and, ideally, replication.

A subgroup should not become a rescue device for a failed pivotal trial. If it reveals a credible responder population, the usual regulatory response is to design a study capable of testing that hypothesis directly.

Sensitivity analyses should test fragility, not decorate the report

A robust conclusion should not depend on one convenient analytical choice. The protocol and SAP should identify the major assumptions capable of influencing the result—missing-data handling, covariance structure, adherence, influential observations, centre effects, endpoint derivation—and specify sensitivity analyses that probe them. The objective is not to demonstrate that every analysis is significant. It is to understand whether the scientific conclusion changes when reasonable analytical assumptions change.

8. Population, Dose, Duration, Comparator and Adherence

These design variables determine whether a positive result can be translated into the proposed conditions of use.

Efficacy in whom?

The study group should either represent the target population or permit a biologically appropriate extrapolation to it. This is easy to say and often difficult to achieve. Recruiting participants with completely optimal baseline values may leave little room for improvement. Recruiting patients with substantial disease or complex medication may create a result that cannot readily be translated to a generally healthy population.

The best design often lies between those extremes: participants should have enough physiological scope to respond while remaining relevant to the intended consumer. Age, sex, baseline status, medication, nutritional status, BMI, physical activity, training state and baseline biomarker concentrations can all affect both responsiveness and external validity.

Dose is part of the claim

A study performed at a dose that cannot be delivered in the intended commercial product may be scientifically interesting and commercially useless. The same is true when efficacy depends on administration conditions that will not be reproduced in practice—for example, fasting administration when the product will be consumed with meals, or a timing schedule that consumers are unlikely to follow.

The pivotal protocol should therefore be written with a translation test in mind: can the commercial product reproduce the dose, formulation, timing, duration and relevant co-conditions under which the effect was demonstrated?

The comparator is a regulatory design decision

In food studies, a placebo is not always inert in the way a pharmaceutical placebo is intended to be. Removing an ingredient may change energy, macronutrients, sweetness, viscosity, flavour or texture. Those changes can affect the endpoint or reveal treatment allocation. Comparator design therefore deserves the same scientific attention as the active formulation.

The objective is to isolate the effect of the constituent under study while preserving blinding where feasible. Sensory matching, energy matching, macronutrient composition and the physiological activity of replacement ingredients should all be considered.

Adherence closes the causal chain

No efficacy programme can be persuasive if the sponsor does not know whether participants actually consumed the intervention. Returned product counts, diaries, electronic records and exposure biomarkers may all be useful depending on the product. The adherence strategy should be specified before the study, together with the criteria for major non-compliance and the way adherence will be considered in sensitivity or per-protocol analyses.

Per-protocol analyses can be informative, but they should not replace a pre-specified primary analysis simply because excluding inconvenient participants improves the result.

9. Replication, Consistency and the Totality of Evidence

EFSA does not count positive studies; it weighs the body of evidence.

There is no mechanical “two RCT rule”

Neither Regulation (EC) No 1924/2006 nor EFSA’s health-claim guidance imposes a universal rule that exactly two randomised controlled trials are required for every claim. EFSA evaluates the totality and weight of the evidence. Pertinent human efficacy studies carry the greatest relevance for establishing the claimed effect, while consistency, reproducibility, dose–response, biological plausibility and study quality influence the overall causal judgement.

The absence of a mechanical numerical rule should not be misread as evidence that one marginally positive study is usually enough. Replication often transforms an isolated signal into a much more persuasive relationship, particularly where effect sizes, populations, doses and methods are broadly coherent.

Replication means reproducing the biological relationship

Two studies are not consistent merely because both contain a p-value below 0.05. Consistency concerns the direction and approximate magnitude of the effect, the confidence intervals, the tested preparation, population, dose, duration and outcome methodology. A second study that finds no effect on the primary outcome but a favourable secondary subgroup analysis does not replicate the first study in any meaningful sense.

Favourable and unfavourable evidence belong in the same picture

Commission Regulation (EC) No 353/2008 requires pertinent scientific evidence to be presented comprehensively, including evidence favourable and unfavourable to the claimed relationship. This is a critical discipline for sponsors because internal development culture naturally gives more attention to successful studies than to neutral or negative ones.

A credible dossier does not hide contradictory evidence. It asks whether the negative study tested an equivalent material, an adequate dose, an appropriate population and a valid endpoint, and whether the finding weakens the causal hypothesis. Sometimes the explanation is methodological. Sometimes the negative result reveals a real limitation. The latter is uncomfortable but scientifically important.

Systematic review should be part of development, not merely submission

The evidence review should begin before the pivotal study. Search strategy, inclusion criteria, study design, population, intervention, comparator, outcomes and unpublished evidence should be mapped systematically enough to identify what is already known and what uncertainty the next study needs to resolve.

A meta-analysis can help quantify consistency, heterogeneity, dose–response and pooled magnitude, but it cannot make weak underlying studies strong. Product non-equivalence, invalid endpoints and high risk of bias do not disappear when studies are pooled.

10. Mechanistic Evidence, Omics and Biological Plausibility

Mechanism strengthens a coherent human effect; it rarely substitutes for one.

Mechanistic evidence can make a dossier scientifically compelling when it explains why the human effect is plausible, why the effective dose makes sense, why the time course is credible and why particular populations may respond differently. Bioavailability studies, metabolite data, receptor or enzyme interactions, cellular signalling, microbiome-mediated pathways, animal models and ex vivo work can all contribute.

Their role must nevertheless remain proportionate. For claims that depend on demonstrating a beneficial physiological effect in humans, elegant mechanistic work does not compensate for the absence of pertinent human efficacy evidence. Mechanism answers “how might this work?”; the pivotal human study answers “does the claimed effect occur under the proposed conditions of use?”

Omics: powerful discovery tools with a multiplicity problem

Metabolomics, transcriptomics, proteomics and microbiome analysis can generate biologically rich data and help identify mechanisms or responder patterns. They can also generate thousands of statistical comparisons. Unless the regulatory meaning of a derived marker is already established, omics outputs should generally be positioned as mechanistic or exploratory evidence rather than as substitutes for the primary physiological outcome.

The most convincing use of omics is hypothesis-led: the technology is used to interrogate a pathway already relevant to the claim, and the resulting pattern is interpreted alongside the clinical effect rather than instead of it.

The dossier should tell one biological story

Mechanistic evidence is strongest when it forms a continuous chain from exposure to benefit: the constituent is absorbed or reaches a relevant site; it interacts with a plausible target; the pathway changes in the predicted direction; the physiological outcome changes; and the dose and time course are mutually coherent. Contradictions between mechanism and clinical data should be addressed openly rather than separated into different sections of the dossier.

11. Regulatory Routes and the Current EFSA Procedure

Scientific strategy and procedural strategy now need to be planned together.

Article 13(5), Article 14 and Article 19

For innovative products, Article 13(5) of Regulation (EC) No 1924/2006 is particularly relevant where an addition to the list of claims is based on newly developed scientific evidence and/or includes a request for protection of proprietary data. Article 14 covers reduction-of-disease-risk claims and claims referring to children’s development and health. Article 19 provides for modification, suspension or revocation of authorisations in the circumstances set out by the Regulation.

The choice of route affects the scientific question and the dossier architecture. A disease-risk-reduction claim, for example, concerns modification of a risk factor rather than a claim to prevent or treat the disease itself. That distinction must be reflected in both endpoint choice and wording.

EFSA assessment is not the same as authorisation

Current EFSA materials are explicit about the institutional division of roles. EFSA evaluates the scientific basis of the claim and adopts an opinion. The European Commission and Member States then operate the authorisation procedure. Using the phrase "EFSA approval" in technical documents may therefore be convenient shorthand, but it is legally imprecise and is best avoided in a publication intended for regulatory professionals.

The post-2021 pre-submission environment

Regulation (EU) 2019/1381 changed the governance of studies used in regulated-product applications. For health-claim applications, EFSA's current procedure requires studies commissioned or carried out after 27 March 2021 and submitted in support of an application to be notified in Connect.EFSA before they start. Connect.EFSA is also used to obtain the pre-application identification number and to access pre-submission services.

The dossier itself is submitted through the European Commission's E-Submission Food Chain platform (ESFC) to the competent authority of a Member State. Once the application has passed the relevant validity and confidentiality checks and reaches EFSA for scientific assessment, the non-confidential dossier is subject to transparency provisions, including public consultation during the assessment phase.

PRE-SUBMISSION ADVICE HAS LIMITS

EFSA strongly recommends general pre-submission advice for questions about requirements and the content of a future application. However, EFSA states that it cannot design the applicant's study, develop the specific hypotheses or validate a proposed protocol. A sponsor should therefore never describe general pre-submission advice as "EFSA approval of the trial design".

Plan for transparency before the study starts

Transparency and confidentiality strategy should be considered at the beginning of development, not during dossier upload. Contracts with CROs and laboratories should secure access to raw data, study reports, code, analytical documentation and information necessary for disclosure or sanitisation. A sponsor that owns only a final PDF report may discover too late that it lacks the documentation needed to respond efficiently to questions or to reconstruct the analysis.

12. From Clinical Study Report to EFSA Dossier

A dossier is not a collection of papers. It is a structured causal argument supported by traceable evidence.

What the assessor should be able to reconstruct

For each pivotal human study, the dossier should make clear what was planned, what was done, what was analysed and what changed after the study began. Protocols and amendments, the statistical analysis plan, randomisation procedures, analysis populations, variable derivations, missing-data decisions, protocol deviations, sensitivity analyses and complete outcome reporting all contribute to credibility.

A concise journal article may be useful as a scientific communication, but it rarely contains enough detail to serve as the sole regulatory record of a pivotal trial. The dossier should preserve the study at a level of detail that allows the assessor to understand the analysis rather than simply accept the authors' conclusion.

Turn results into an argument, not a list of p-values

A strong regulatory interpretation follows a sequence. First, state the observed treatment effect and its precision. Second, explain whether the magnitude is physiologically meaningful. Third, describe whether the endpoint directly reflects the claimed effect. Fourth, place the result in the context of the other pertinent human studies. Fifth, show that the dose, formulation, population and duration correspond to the proposed conditions of use. Finally, use mechanistic data to explain why the relationship is biologically plausible.

A REGULATORY RESULT IN ONE PARAGRAPH

Ingredient X reduced Outcome Y by 7.4% versus control; the 95% confidence interval was 4.1% to 10.6%. Outcome Y is an accepted measure of physiological function Z. The magnitude was consistent with the prospectively defined relevance framework and with the direction and approximate size observed in the independent confirmatory study. The effect occurred at the intended commercial dose in a population representative of the proposed target group. Human kinetic data and mechanistic evidence support the biological pathway.

Traceability is underrated

Every material proposition in the application should be traceable to a study, dataset, analytical certificate or regulatory source. A simple evidence-to-dossier matrix can prevent important claims from being supported by the wrong study or by a secondary source when primary evidence exists. It also makes contradictory evidence visible rather than allowing it to disappear between sections.

13. Worked Regulatory Examples

The following examples are intentionally simple. Their purpose is to show how regulatory interpretation changes when endpoint validity, magnitude and uncertainty are considered together.

Example 1 — The antioxidant trap

A polyphenol extract is positioned for protection against oxidative stress. An eight-week trial reports ORAC +19% ($p < 0.001$), FRAP +11% ($p = 0.006$), plasma polyphenols +42% ($p < 0.001$) and no change in CRP. Commercially, the study may look impressive: several markers moved and the p-values are small.

Regulatorily, the central question remains unresolved. Global antioxidant-capacity assays do not, by themselves, demonstrate protection of a relevant biological molecule from oxidative modification. The study may provide exposure and mechanistic information, but it has not necessarily measured the claimed beneficial physiological effect.

A stronger programme would define what is being protected, select a validated marker of oxidative damage appropriate to that target, and use global antioxidant-capacity measures only as supporting information if they add mechanistic value.

Example 2 — Statistically significant, biologically uncertain

A validated physiological endpoint produces a between-group difference of –1.1 units, with a 95% confidence interval from –2.15 to –0.05 and $p = 0.041$. The result is conventionally significant. The correct regulatory interpretation, however, depends on whether –1.1 units is a meaningful change, whether that magnitude was considered prospectively, whether the confidence interval includes effects close to triviality, whether the endpoint was pre-specified and whether the result is replicated.

The phrase “significant improvement” therefore says less than the data. A more careful conclusion may be that the trial provides evidence of a small treatment effect but leaves uncertainty about its physiological relevance.

Example 3 — Non-significant does not always mean “no effect”

Suppose the estimated treatment difference is –4.0 units, with a 95% confidence interval from –8.3 to +0.3 and p = 0.067. A simplistic reading would declare that the product does not work. That is not what the estimate says. The data are compatible with both a potentially meaningful benefit and little or no effect.

The study has failed to establish efficacy conclusively, but the reason may be insufficient precision. If the design was otherwise credible and the point estimate aligns with prior evidence, a better-powered confirmatory trial may be justified. This is the practical difference between absence of evidence and evidence of absence.

Example 4 — Muscle function measured through the wrong outcome

A sponsor wants a claim concerning maintenance of normal muscle strength, but the pivotal study measures creatine kinase, inflammatory cytokines, muscle soreness and a metabolic marker. All four outcomes improve significantly. The trial may support an interesting recovery or mechanistic hypothesis, but it has not necessarily demonstrated muscle strength. If the intended claim concerns strength, the programme should include a validated direct measure of strength appropriate to the scope of the claim.

14. Where Programmes Fail — and Why Statistics Cannot Rescue Them

Most fatal weaknesses are created before the database exists.

When a study reaches the statistician with a poorly characterised product, an invalid endpoint or a population unrelated to the intended claim, no analytical technique can repair the underlying regulatory problem. The most valuable contribution of Regulatory Affairs is therefore often preventive: stopping the wrong trial before it starts.

Failure mode	Why it matters	Best time to fix it
Vague claim concept	The endpoint cannot be qualified against an undefined benefit	Before protocol drafting
Inadequate product characterisation	The clinical evidence cannot be confidently attributed to the commercial material	Before clinical-batch release
Mechanistic marker used as pivotal endpoint	The study may show a pathway without demonstrating the claimed benefit	Before power calculation
Too many unranked endpoints	Multiplicity and post-hoc selection weaken confirmatory interpretation	Before SAP finalisation
Sample size driven by budget	The trial may be incapable of estimating a relevant effect precisely	Before CRO contract
Wrong population	A positive result may not generalise to the claim population	Before recruitment
Unrealistic dose	Efficacy cannot be translated to commercial conditions of use	Before formulation lock
Inadequate duration	The study may finish before the physiology can respond or stabilise	Before protocol finalisation
Claim selected after results	The programme becomes exploratory by design	Before first participant in
Regulatory team joins after database lock	Procedural and scientific defects are discovered when they are expensive or impossible to repair	At project inception

Risk of bias can outweigh a small p-value

Randomisation, allocation procedures, comparator suitability, blinding, attrition, endpoint measurement, selective reporting and analytical choices all influence credibility. A very small p-value from a study with substantial risk of bias may carry less evidentiary weight than a more modest but robust estimate from a well-conducted trial.

The practical lesson is not that every study must resemble a pharmaceutical registration trial. It is that the design should control the biases that matter for the specific food intervention and endpoint, and the limitations should be visible rather than hidden.

15. A Development Model That Reduces Regulatory Uncertainty

Good development is less about accumulating data than about knowing which uncertainty the next study is supposed to remove.

The traditional commercial sequence—ingredient, formulation, launch, clinical study, claim search—often places regulatory science too late. For a serious EFSA health-claim programme, the sequence should be almost the reverse. The regulatory target should shape the evidence architecture, which then shapes the clinical programme and ultimately the commercial communication.

This may feel slower during the first months of development because more decisions are made before the trial starts. In practice it is usually faster and cheaper than discovering, after a large clinical investment, that the endpoint was unsuitable, the commercial dose is lower than the tested dose, or the population cannot support the intended claim.

A ten-stage development pathway

1. Define the food or constituent and lock the clinically relevant specification.
2. Define the intended physiological effect and target population in regulatory terms.
3. Map EFSA guidance, relevant opinions and the existing human evidence.
4. Qualify the primary endpoint and measurement method.
5. Define the prospective biological-relevance framework.
6. Align dose, formulation, duration, comparator and population with the proposed conditions of use.
7. Finalise the protocol and SAP; address study-notification and pre-submission requirements.
8. Execute the trial with documented product quality, adherence, blinding and data integrity.
9. Interpret the result through magnitude, precision, robustness, consistency and plausibility.
10. Build the dossier from the totality of evidence, including unfavourable or contradictory data.

The final test before submission

Before filing, the team should be able to answer one question without relying on promotional language:

> *Does this sufficiently characterised food or constituent, consumed at a realistic dose by the intended population, reproducibly produce a beneficial physiological effect of credible magnitude, measured with an appropriate outcome, and supported by a coherent totality of evidence?*

If the answer requires several qualifications that were not anticipated at the start of development, the dossier is probably carrying unresolved uncertainty. If the answer is clear and traceable to the evidence, the programme has moved beyond the search for significance and towards the standard that matters: a defensible causal argument.

References and Core Regulatory Sources

1. **European Parliament and Council.** Regulation (EC) No 1924/2006 on nutrition and health claims made on foods. Official Journal of the European Union. Consolidated text available via EUR-Lex. Source
2. **European Commission.** Commission Regulation (EC) No 353/2008 establishing implementing rules for applications for authorisation of health claims as provided for in Article 15 of Regulation (EC) No 1924/2006. Source
3. **European Parliament and Council.** Regulation (EC) No 178/2002 laying down the general principles and requirements of food law, establishing EFSA and laying down procedures in matters of food safety, as amended. Source
4. **European Parliament and Council.** Regulation (EU) 2019/1381 on the transparency and sustainability of the EU risk assessment in the food chain. Source
5. **EFSA Scientific Committee.** Statistical Significance and Biological Relevance. EFSA Journal. 2011;9(9):2372. doi:10.2903/j.efsa.2011.2372. Source
6. **EFSA NDA Panel.** General scientific guidance for stakeholders on health claim applications (Revision 1). EFSA Journal. 2021;19(3):6553. doi:10.2903/j.efsa.2021.6553. Source
7. **EFSA NDA Panel.** Scientific and technical guidance for the preparation and presentation of a health claim application (Revision 3). EFSA Journal. 2021;19(3):6554. doi:10.2903/j.efsa.2021.6554. Source
8. **EFSA Scientific Committee.** Guidance on the use of the weight of evidence approach in scientific assessments. EFSA Journal. 2017;15(8):4971. doi:10.2903/j.efsa.2017.4971. Source
9. **EFSA NDA Panel.** Guidance for the scientific requirements for health claims related to antioxidants, oxidative damage and cardiovascular health (Revision 1). EFSA Journal. 2018;16(1):5136. doi:10.2903/j.efsa.2018.5136. Source
10. **EFSA NDA Panel.** Guidance on the scientific requirements for health claims related to muscle function and physical performance (Revision 1). EFSA Journal. 2018;16(10):5434. doi:10.2903/j.efsa.2018.5434. Source
11. **EFSA NDA Panel.** Guidance on the scientific requirements for health claims related to the immune system, the gastrointestinal tract and defence against pathogenic microorganisms. EFSA Journal. 2016;14(1):4369. doi:10.2903/j.efsa.2016.4369. Source
12. **EFSA NDA Panel.** Guidance on the scientific requirements for health claims related to appetite ratings, weight management and blood glucose concentrations. EFSA Journal. 2012;10(3):2604. doi:10.2903/j.efsa.2012.2604. Source
13. **European Food Safety Authority.** Health claim application procedure. Last reviewed 6 May 2026. Source
14. **European Food Safety Authority.** Services for applicants: General Pre-Submission Advice and other support services. Accessed September 2026. Source
15. **European Food Safety Authority.** Toolkit for applications: pre-application IDs, study notifications and Connect.EFSA user guidance. Accessed September 2026. Source

CURRENCY OF SOURCES

EFSA updates guidance and application procedures. Before a real submission, applicants should verify the latest versions in EFSA's guidance catalogue, the health-claim application page, the Commission's ESFC guidance and EUR-Lex.

Operational Annexes

The following annexes convert the scientific principles of the manual into working tools. They are designed to be adapted for internal project governance, CRO briefing, regulatory review and management authorisation. They deliberately separate official EFSA requirements and concepts from internal decision rules created to manage development risk.

Annex I. Pre-Trial EFSA Readiness Review

The purpose of this review is to answer one practical question before a company commits to a pivotal human intervention study: if the trial produces the expected result, will that result actually address the proposed health-claim question? The review should be completed jointly by Regulatory Affairs, Scientific Affairs, Clinical Affairs, Biostatistics, Product/Analytical Development and Quality. A CRO should contribute to feasibility, but it should not be asked to define the regulatory hypothesis on behalf of the sponsor.

STOP RULE
If the team cannot express the product, dose, target population, beneficial physiological effect and primary outcome in one coherent sentence, the programme is not ready for pivotal trial design.

A. Regulatory hypothesis

Field	Project definition
Food/constituent	
Regulatory route	Article 13(5) / Article 14(1)(a) / Article 14(1)(b) / Article 19 / other
Target physiological function	
Target population	
Proposed conditions of use	
Preliminary claim concept	
Primary outcome variable	
Primary measurement method	

Complete the working sentence: "Consumption of _____, at _____, by _____, produces _____, demonstrated by _____."

B. Product characterisation

Question	Yes / No / N.A.	Evidence / action
Is the identity of the clinical material unambiguous?		
Are the defining compositional parameters specified quantitatively?		
Are analytical methods fit for purpose?		
Is the manufacturing process sufficiently defined?		
Is batch-to-batch variability understood?		
Is stability adequate for the clinical period?		
Can the clinical batch be linked to the intended commercial specification?		
Are retained samples and certificates of analysis available?		

C. Claimed effect and endpoint qualification

Decision question	Assessment
What physiological function is being evaluated?	
Why would a favourable change be beneficial for the proposed population?	
Does the primary endpoint measure the function directly, or is it a surrogate?	
What is the regulatory status of the endpoint in this claim context?	
Is the measurement validated, reproducible and sufficiently specific?	
What pre-analytical and laboratory variables must be controlled?	
Which EFSA guidance or prior opinions are most relevant?	

D. Biological relevance

Parameter	Prospective definition
Expected treatment effect	
Minimum effect regarded as biologically relevant	
Scientific basis for that judgement	EFSA precedent / RCTs / meta-analysis / physiological literature / validated threshold / pilot data
Interpretation if the CI includes trivial effects	
Interpretation if the CI includes no effect but a meaningful benefit remains plausible	

E. Population and generalisability

Question	Assessment
Who will carry the proposed claim?	
Who will be recruited?	
What baseline status is required to leave room for improvement?	
Are disease status or medications likely to compromise extrapolation?	
Which variables may modify response (age, sex, BMI, nutrition, training, baseline biomarker)?	
If the study succeeds, can the result credibly be extrapolated to the claim population?	

F. Conditions of use

Parameter	Pivotal study	Intended commercial use	Aligned?
Daily dose			
Frequency			
Timing			
Formulation / food matrix			
Duration			
Relevant co-interventions			
Dietary restrictions			

G. Comparator, blinding and adherence

Design element	Questions to settle

Comparator	Is it compositionally and sensorially appropriate? Could it affect the primary endpoint?
Blinding	Can participants, investigators and outcome assessors remain unaware of allocation?
Adherence	How will exposure be measured—counts, diaries, electronic monitoring, biomarkers?
Non-compliance	What constitutes a major deviation and how will it be handled analytically?

H. Statistical readiness

Pre-specified item	Status / decision
Estimand	
Primary hypothesis	
Effect measure	
Sample-size assumptions	
Variance assumption	
Alpha and power	
Primary analysis population	
Primary model	
Covariates	
Multiplicity strategy	
Missing-data strategy	
Sensitivity analyses	
Subgroups	
Interim analysis, if any	

I. Procedural readiness

Control point	Status
Relevant EFSA guidance reviewed	
Relevant EFSA opinions mapped	
Pre-application ID obtained / planned	
Study notification requirement assessed	
Connect.EFSA responsibilities assigned	
GPSA question list considered	
Raw-data ownership and access secured contractually	
Confidentiality / transparency strategy considered	
ESFC dossier ownership assigned	

Annex II. Claim–Effect–Endpoint Regulatory Matrix

This matrix should be completed before protocol drafting. Its value lies in forcing the team to connect the commercial proposition with the exact physiological effect and the outcome that will be used to demonstrate it. Where a cell cannot be completed convincingly, the uncertainty should be treated as a development task rather than hidden in the protocol.

Element	Required project statement
Food/constituent	Exact identity and clinical/commercial specification
Claim concept	Precise food–health relationship
Regulatory route	Article 13(5), Article 14, Article 19 or other
Beneficial physiological effect	Why the effect is beneficial for the target population
Target population	Who is intended to benefit
Primary outcome	What directly demonstrates the effect
Measurement method	Why the method is valid and appropriate
Biological-relevance framework	What magnitude or range would matter
Study population	Why it represents or can be extrapolated to the target population
Dose and administration	Why the study conditions match intended use
Duration	Why the intervention period is physiologically adequate
Comparator	Why causality can be isolated
Key secondary outcomes	Supportive physiology
Mechanistic endpoints	Plausibility, exposure or pathway evidence
EFSA precedent	Most relevant positive and negative opinions / guidance
Principal regulatory uncertainty	The single issue most likely to weaken substantiation

Illustrative endpoint hierarchy

Claim area	Potential pivotal focus	Supportive / mechanistic information	Typical risk
Blood lipids	Appropriately measured LDL-C or other claim-relevant lipid outcome	ApoB, non-HDL-C, absorption/synthesis markers where relevant	Inadequate duration, medication confounding, trivial effect size
Blood pressure	Validated BP measurement appropriate to the claim	Related vascular or mechanistic outcomes	Measurement variability, baseline status, short duration
Oxidative damage	Validated marker of oxidative modification of a relevant biological molecule	Exposure markers, mechanistic antioxidant measures	Using global antioxidant capacity as if it were the claimed benefit
Muscle strength	Validated direct strength assessment appropriate to the claim	CK, soreness, inflammatory or metabolic markers	Mechanistic outcomes substituted for function
Cognitive function	Validated measure of the specific cognitive domain claimed	Mechanistic neurobiological markers	Vague “brain health” construct, multiple unranked cognitive tests

Annex III. Prospective Biological Relevance Memorandum

This memorandum should be finalised before the pivotal study begins. It is not intended to force a numerical threshold where one cannot be scientifically justified. Its purpose is to document, prospectively, how the team intends to distinguish a physiologically meaningful effect from a statistically detectable but trivial one.

Project field	Entry
Product / study ID	
Proposed claim	
Target population	
Primary endpoint	
Direction of benefit	Increase / decrease / maintenance / prevention of deterioration / other
Expected treatment effect	
Prospective relevance threshold or interpretation framework	
Scientific basis	
Date approved	
Approved before first participant enrolled?	Yes / No

Evidence supporting the relevance framework

Evidence source	Effect / threshold	Why it is relevant
EFSA guidance or opinion		
Previous RCT		
Meta-analysis		
Physiological literature		
Validated functional / clinical threshold		
Pilot data		

Interpretation zones

Pattern	Interpretation
Estimate in the favourable direction; CI supports a relevant magnitude	Convincing biological response, subject to study validity and the totality of evidence
Statistically significant; effect smaller than the prospective relevance range	Statistically detectable, biologically weak or uncertain
Point estimate potentially meaningful; CI includes trivial or null effects	Promising but imprecise; efficacy not conclusively established
Narrow CI centred on a trivial effect	Evidence against a meaningful benefit
Favourable post-hoc endpoint or subgroup only	Exploratory unless independently confirmed

Preferred conclusion template

"The pre-specified analysis estimated a between-group treatment effect of _____, with a 95% confidence interval from _____ to _____. The magnitude was [above / below / compatible with] the prospectively defined biological-relevance framework. The result was [consistent / not consistent] with the direction and approximate magnitude of the existing human evidence."

Annex IV. EFSA-Oriented Statistical Analysis Plan Framework

The SAP should translate the protocol into a reproducible analysis before treatment allocation is revealed. The objective is not to anticipate every possible analysis but to settle the choices that could materially change interpretation of the primary question.

1. Administrative information

Item	Entry
Protocol / study ID	
SAP version and date	
Responsible statistician	
Approval date	
Planned database lock	
Amendments and rationale	

2. Objectives and estimand

Component	Definition
Primary objective	
Key secondary objectives	
Exploratory objectives	
Population	
Treatment condition	
Outcome variable	
Summary measure	Mean difference / ratio / risk ratio / odds ratio / other
Intercurrent events	Discontinuation, rescue treatment, prohibited medication, poor adherence, treatment switching, other

3. Analysis populations

Population	Prospective definition and role
Primary / full analysis set	
Per-protocol set	
Safety population	
Other	

4. Primary model and effect reporting

Item	Specification
Primary hypothesis	
Sidedness and alpha	
Primary statistical model	
Baseline adjustment	
Stratification / centre effects	
Repeated-measures structure if relevant	
Primary effect estimate	
95% confidence interval	
Exact p-value	
Descriptive statistics to accompany estimate	

5. Multiplicity

List all confirmatory hypotheses and specify whether the design uses a single primary outcome, hierarchy, gatekeeping or another justified procedure. Supportive and exploratory analyses should be labelled as such rather than incorporated into a post-hoc hierarchy.

Hypothesis	Confirmatory?	Order / procedure	Alpha allocation / note
Primary			
Key secondary 1			
Key secondary 2			

6. Missing data and protocol deviations

Issue	Pre-specified approach
Anticipated missingness	
Primary handling method	
Sensitivity analysis	
Major protocol deviations	
Rules for PP inclusion	
Blinded review of deviations	

7. Outliers, transformations and influential observations

Define how unusual observations will be identified, whether transformations are planned, and how influence will be assessed. Observations should not be excluded merely because they weaken statistical significance. If alternative analyses excluding influential observations are scientifically justified, they should normally be treated as sensitivity analyses unless the rule was pre-specified.

8. Subgroups

Subgroup	Biological rationale	Pre-specified?	Interaction test	Interpretation

9. Sensitivity analyses

- Alternative missing-data assumptions where relevant
- Per-protocol analysis
- Alternative covariance or longitudinal structure
- Adjusted and unadjusted estimates where scientifically useful
- Centre sensitivity
- Influential-observation analysis
- Exposure/adherence sensitivity
- Alternative endpoint derivation if prospectively justified

10. Reproducibility and archive

Archive the analysis datasets, derivation specifications, statistical code, logs, software/version information, randomisation documentation, quality-control records and complete outputs. The regulatory record should permit the primary analysis to be reconstructed without relying on undocumented decisions held by an external vendor.

Annex V. INTABIOTECH EFSA Development Readiness Score

IMPORTANT

This is an internal decision-support tool created for development governance. It is not an EFSA scoring methodology, does not reflect an official EFSA threshold and should never be presented as one.

The purpose of the score is to prevent a programme from advancing because enthusiasm for the ingredient obscures structural weaknesses in the evidence architecture. Scores are useful only after the fatal-flaw screen has been completed.

Domain	Maximum	What should be demonstrated
A. Product characterisation	12	Identity, quantitative specification, analytical methods, batch consistency, stability, clinical/commercial equivalence
B. Claim architecture	12	Precisely defined food–effect relationship, beneficial effect, population, regulatory route, conditions of use
C. Endpoint validity	14	Claim alignment, method validity, EFSA precedent, analytical reliability, time point, pre-specification
D. Biological relevance	10	Prospective relevance framework, scientific justification, realistic expected effect, CI interpretation
E. Population	8	Appropriateness, extrapolation, baseline status, confounding control
F. Intervention and comparator	8	Commercial dose, duration, comparator, blinding, adherence
G. Trial design	12	Randomisation/control, bias control, sample-size justification, pre-specification, operational feasibility
H. Statistical strategy	10	Estimand, model, multiplicity, missing data, sensitivity, SAP lock
I. Regulatory readiness	8	Guidance/opinions, study notification, pre-submission process, confidentiality, raw-data access
J. Totality of evidence	6	Systematic evidence map, negative/unpublished evidence, plausibility, replication, contradictions

Maximum score: 100.

Score	Internal decision	Interpretation
85–100	GO	Programme is structurally mature, subject to absence of a fatal flaw.
70–84	CONDITIONAL GO	Plausible programme, but defined weaknesses should be closed before first participant in.
55–69	REDESIGN	Material risk that the trial will generate evidence of limited regulatory value.
<55	NO-GO	The pivotal investment is not justified in its present form.

Fatal flaws — override the numerical score

Fatal flaw	Consequence
The food/constituent cannot be adequately characterised.	Stop or redefine the product.
The clinical material is not demonstrably representative of the intended commercial material.	Resolve product equivalence before efficacy investment.
The claimed effect cannot be justified as a beneficial physiological effect.	Redefine the claim question.
The primary endpoint does not demonstrate the intended effect.	Redesign endpoint strategy.
The study population cannot support the target population.	Redesign recruitment or claim scope.
The tested dose is incompatible with intended conditions of use.	Reformulate or redesign dose strategy.
The comparator prevents causal interpretation or effective blinding.	Redesign control.
The sample size cannot answer the primary question with useful precision.	Re-power or redesign.
The primary endpoint will be chosen after results are seen.	Treat as exploratory; redesign confirmatory programme.
Applicable study-notification obligations have been overlooked.	Resolve procedural consequences before progression.
The sponsor lacks reliable access to complete underlying data and analysis records.	Renegotiate contracts before study start.
Materially unfavourable evidence is knowingly omitted from internal assessment.	Stop and perform a complete evidence review.

Annex VI. Pre-Submission Governance and Dossier Readiness

A health-claim programme crosses scientific, statistical, quality, legal and procedural boundaries. The sponsor should therefore define ownership before the clinical study begins. When responsibilities are left implicit, critical tasks—study notification, product equivalence, raw-data access, confidentiality, endpoint qualification—tend to be discovered after they have become expensive to correct.

Role	Core responsibility
Regulatory Lead	Claim architecture, regulatory route, applicable guidance, pre-submission strategy, study-notification control, dossier structure and authority interactions.
Scientific Lead	Biological rationale, endpoint qualification, interpretation of relevance, evidence synthesis and mechanistic coherence.
Clinical Lead	Protocol delivery, CRO oversight, centres, participant management, operational quality and adherence.
Biostatistical Lead	Estimand, power, randomisation, SAP, primary analysis, sensitivity analyses and reproducibility.
Product / Analytical Lead	Clinical-material identity, specification, analytical methods, batch release, stability and commercial equivalence.
Quality Lead	Document control, vendor qualification, audit trail, data integrity, deviations and archive.

Minimum regulatory master file before first participant in

File block	Minimum content
Regulatory	Claim strategy; regulatory route; current EFSA guidance; relevant opinions; pre-application ID; study-notification records.
Product	Formulation; specification; manufacturing description; CoA; analytical methods; stability; clinical/commercial batch linkage.
Scientific	Systematic evidence map; mechanism review; endpoint qualification; Biological Relevance Memorandum.
Clinical	Final protocol; comparator rationale; randomisation; adherence methodology; study operational plan.
Statistics	Sample-size calculation; assumptions; SAP; planned outputs; code governance.
Contractual / data	Data ownership; publication rights; raw-data access; code access; audit rights; laboratory documentation; CRO obligations.

Study-notification control point

Control	Record
Responsible person	
Pre-application ID	
Study notification date	
Planned study start date	
Notification/reference number	
CRO/laboratory co-notification responsibilities	
Verification completed before study start	Yes / No

Correct use of General Pre-Submission Advice

Use GPSA to clarify requirements, the content of the future application and procedural issues within EFSA's stated scope. Do not ask EFSA to design the study or validate the sponsor's proposed protocol. Internally, record what was asked, what EFSA answered and what remains the sponsor's scientific responsibility.

Evidence-to-dossier traceability

Dossier proposition	Primary source	Study / document	Location	Favourable / unfavourable	Regulatory weight
Product adequately characterised					
Effect beneficial					
Endpoint valid					
Human efficacy					
Dose effective					
Effect replicated					
Mechanism plausible					
Contradictory evidence					

Human study evidence table

Study	Product equivalence	Population	Dose / duration	Primary outcome	Effect estimate / 95% CI	Main limitation

Annex VII. Red Team Review and Executive Go/No-Go Decision

Before submission, the dossier should be reviewed by people who were not responsible for generating the pivotal evidence. Their initial role is not to defend the programme but to attack it. A credible Red Team review often identifies the question that an external assessor is most likely to ask—and gives the sponsor time to answer it before submission.

Red Team questions

Attack line	Question
Characterisation	Why might the clinical evidence not apply to the product being submitted?
Claim / endpoint	Why might the primary outcome fail to demonstrate the proposed beneficial effect?
Statistics	Could the result depend on multiplicity, missing data, modelling choices, outliers or selective analysis?
Biological relevance	Is the magnitude actually meaningful, or simply statistically detectable?
Replication	Is the result reproducible, or is the programme relying on one favourable trial?
Population	Why might the study group not support the target population?
Dose and use	Can the marketed conditions realistically reproduce the study conditions?
Contradictory evidence	Which unfavourable study most seriously challenges the claim?
Mechanism	Does the proposed mechanism explain the observed dose, timing and magnitude?
Procedure	What procedural or transparency issue could delay or complicate the application?
Final challenge	What is the single question an EFSA assessor could ask that the current dossier would struggle to answer?

Executive pre-trial investment authorisation

Decision field	Entry
Product	
Proposed health relationship	
Regulatory route	
Target population	
Commercial dose	
Primary endpoint	
Biological-relevance framework	
Estimated clinical investment	
Development readiness score	___ / 100
Fatal flaw present?	Yes / No
Principal regulatory uncertainty	
Principal scientific uncertainty	
Principal statistical uncertainty	
Principal commercial-translation risk	

Decision

Decision	Meaning
GO	The pivotal programme is scientifically and regulatorily mature.
CONDITIONAL GO	Proceed only after the specified deficiencies have been closed.
REDESIGN	The concept remains viable, but the proposed study is not capable of producing sufficiently persuasive evidence.
NO-GO	The expected evidentiary value does not justify the investment in the current design.

Required condition before first participant in	Owner	Due date	Closed

Internal sign-off

Function	Name / signature	Date
Regulatory Affairs		
Scientific Affairs		
Biostatistics		
Quality		
Management		

FINAL OPERATIONAL PRINCIPLE

A pivotal trial should begin only when the company knows what result it needs, why that result would matter physiologically, how it will be measured, and how it will support the intended regulatory proposition. That is the practical difference between conducting a clinical study and building an EFSA evidence programme.

FROM CLINICAL EVIDENCE TO EFSA

Statistical significance alone does not substantiate a health claim. In the EFSA context, the central question is whether a sufficiently characterised food or constituent, consumed under defined conditions of use, can be shown to cause a beneficial physiological effect in humans. This manual translates that regulatory standard into a practical framework for industry, showing how to design human intervention studies, interpret biological relevance, distinguish valid endpoints from merely responsive biomarkers, and build a persuasive weight-of-evidence dossier. Written for scientific, clinical and regulatory teams, it offers a rigorous route from study design to EFSA-ready substantiation.

SCIENCE
EVIDENCE
TRUST
BETTER HEALTH

Inside this manual

- Defining the regulatory question before the study begins
- Selecting endpoints that can genuinely support a claim
- Interpreting effect size, confidence intervals and biological relevance
- Avoiding common pitfalls in statistics, population selection and study design
- Building a coherent weight-of-evidence strategy
- Using practical annexes for SAP, dossier planning and go/no-go decisions

FOOD
SCIENCE
HUMAN HEALTH
EUROPE

EVIDENCE
ASSESSMENT
REGULATORY IMPACT

ROBUST SCIENCE
RESPONSIBLE INNOVATION
HEALTHIER TOMORROWS

EFSA does not authorise p-values.
It assesses food–effect relationships.

INTABIOTECH Scientific & Regulatory Affairs — 2026