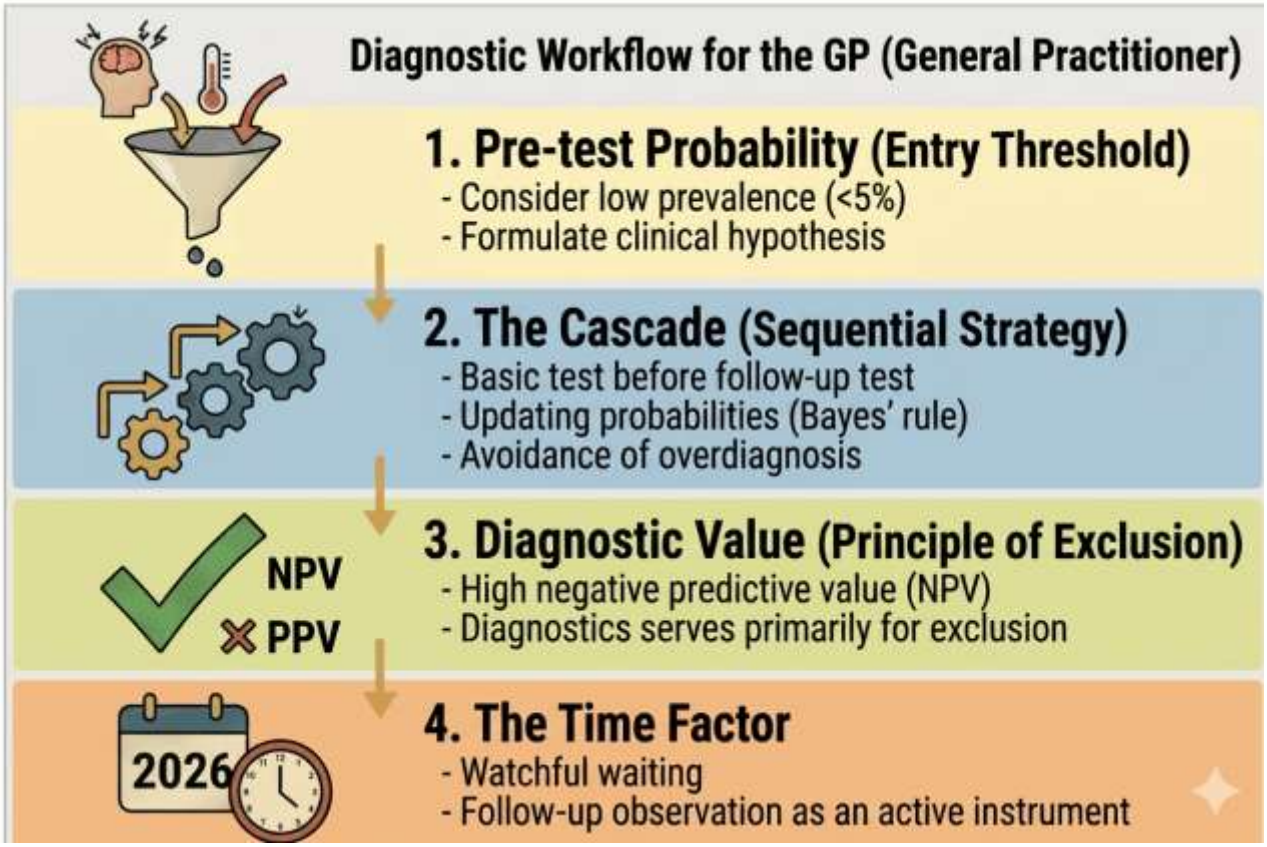
ClinicalOS – Structured decision-making in general practice

Volume 2
Rational Diagnosis



Focus Prevalence-weighted assessment, sequential testing strategies,
instrumental and laboratory diagnostics in GP practice

Key principle: "Don't test blindly; use pre-test probabilities rationally"

PREVIEW - VOLUME 2

Rational Diagnosis

This preview contains curated extracts from the full edition.
It includes the original front matter and contents overview, together with selected pages from Chapter 1, Chapter 7 and Annex A1.
The full edition contains 185 pages; this preview contains 28 PDF pages.
The complete table of contents is shown for orientation and therefore also lists sections that are not included in this preview.

Preview marking

All content pages in this file carry a static 'LESEPROBE' watermark.
The complete paid edition is a separate master file.

Copyright Legal Notice

Note: The information contained in this book has been carefully researched and verified. As AI tools and regulatory frameworks change rapidly, some details may already be out of date by the time you read this. You can find the latest updates at https://clinicalos.de/. The mention of products and manufacturers is without guarantee and does not constitute a recommendation to purchase or use them.

Note on Translation, Content, Use and Liability

This English edition is a translation of the original work written in German. The concept, structure, and core medical content of ClinicalOS were originally developed for German general practice and the German healthcare system, to provide practising GPs, trainee GPs, healthcare assistants (HCAs) and medical students with a structured, practice-oriented learning and reference resource. The content is based on current medical literature, guideline-based sources and practical experience in general practice.

During the linguistic adaptation, references to guidelines, clinical practices and regulatory frameworks in other English-speaking settings were taken into account and highlighted where relevant and feasible; in some areas, differences from German recommendations are also presented explicitly. However, the translation does not constitute a systematic adaptation of every medical, therapeutic or regulatory aspect to all English-speaking countries. Professional guidelines, marketing authorisations, approved dosages, prescribing procedures, reimbursement rules and regulatory requirements may vary between countries;

the guidelines, official prescribing information and regulatory provisions currently applicable in the respective country or healthcare system always take precedence. Dosage information applies to adults without significant renal impairment; individual adjustments are required for elderly patients, those with polypharmacy and during pregnancy. **Current national/regional guidelines (e.g. NICE, ESC, ADA) must be followed as a priority.**

Digital tools, including AI-assisted text generation, were used in the creation of this manual. All content was subsequently checked by the author for technical accuracy, consistency and potential errors. **No guarantee can be given for the correctness, currency or completeness of the information provided;** the complete absence of errors cannot be excluded, including with regard to guidelines or dosage recommendations that may have been updated in the meantime.

The contents of this manual do not replace an individual medical assessment and should not be construed as binding treatment recommendations. Diagnostic and therapeutic decisions must always be made on a case-by-case basis for each individual patient, taking into account the full clinical picture and the currently valid guidelines. **Doctors bear full and ultimate medical responsibility:** any use of content from this manual — whether as a text module, SOP or treatment recommendation — requires independent professional review by the treating doctor. **AI-supported or digitally generated content cannot and must not replace medical judgement.**

Any use of the information presented here is the sole responsibility of the reader. Liability claims relating to damage of a material or immaterial nature arising from the direct or indirect use of the information contained in this manual — even where such information is incomplete or contains errors — are generally excluded.

Contents

Table of Contents

1. Diagnostic fundamentals for GPs

Pre-test probability · Test quality measures · Bayes' rules of thumb · Sequential diagnosis

1.1 Why diagnostics work differently in a GP practice

A GP practice is not a scaled-down hospital. Its diagnostic environment differs fundamentally from that of a hospital, and these differences have direct consequences for the interpretation of each individual test.

The decisive factor: prevalence. In hospital settings, the pre-test probability for the suspected diagnosis is often 30–70 per cent – patients are pre-selected based on admission diagnoses, severity of condition and triage. In GP practices, by contrast, the prevalence of most individual diagnoses is below 5 per cent. A patient presenting with chest pain in A&E has a probability of suspected ACS of 15–25 per cent; the same set of symptoms in a GP practice starts at 1–3 per cent.

This asymmetry has a mathematically inevitable consequence: where prevalence is low, the positive predictive value (PPV) of any test falls dramatically – even with good sensitivity and specificity. At the same time, the negative predictive value (NPV) increases. In GP practice, therefore, diagnostics serve primarily to rule conditions out, not to confirm them.

> **Key point:** In GP practice, the most important diagnostic value of a test is its negative predictive value.
>
> A negative test result at low prevalence is almost always reliable. A positive test result at low prevalence is often a false positive.

The three pillars of GP-specific diagnostics

Prevalence-dependence: Every test has a different predictive value in a GP practice than in a hospital setting – even with identical sensitivity and specificity. Test performance metrics from hospital studies are not directly transferable.

Sequential strategy: Rather than simultaneous test batteries, rational diagnosis follows a cascade: clinical hypothesis → baseline test → result → next step. Each test alters the probability, and the subsequent test builds on the updated probability.

The timeline as a diagnostic tool: Unlike in hospital settings, the GP has the factor of time at their disposal: monitoring the course of the condition, controlled watchful waiting and planned re-evaluation are independent diagnostic strategies, not an expression of passivity.

1.2 Pre-test probability – the key to interpreting any test result

The pre-test probability is the estimated likelihood that a condition is present before a diagnostic test is carried out. It is the single most important factor in interpreting a test result – more important than the sensitivity or specificity of the test itself.

Sources of pretest probability

Epidemiological prevalence: The baseline frequency of a condition in the target population. The 25-diagnosis lists (E0 master list) provide this data for 23 specialist areas. Example: Arterial hypertension in the over-65s = 50–60% (geriatrics list) – high pre-test probability. Macular degeneration = 5–10% – low pre-test probability.

Clinical presentation: History-taking and findings modify the baseline prevalence. A 70-year-old smoker with exertional dyspnoea has a higher pre-test probability of COPD than the general population (baseline prevalence 5–8 per cent, clinically adjusted: 30–50 per cent).

Risk factors and scores: Validated scores (Wells, CHA2DS2-VASc, HEART) systematically quantify the pre-test probability and reduce individual estimation errors.

Calculation examples using real-world prevalence data

Example 1: Depression in GP practice

Prevalence (Psychiatry-25 diagnoses): approximately 10–15% of all consultations concern depressive episodes (F32.9). PHQ-9 as a screening tool: sensitivity 88%, specificity 85%.

With a prevalence of 12% and 100 patients: 12 patients, 88 healthy individuals. PHQ-9 positive: 12 × 0.88 + 88 × 0.15 = 10.6 + 13.2 = 23.8. PPV = 10.6/23.8 = 44.5%. This means: Fewer than half of patients with a positive PHQ-9 screening result actually have clinically significant depression. Consequently, a positive PHQ-9 result always requires a clinical interview for confirmation.

Example 2: Deep vein thrombosis (DVT)

Prevalence (25 cardiology diagnoses): approx. 1–2 per 1,000 patients per year. D-dimer: sensitivity 95 per cent, specificity 40 per cent.

With a prevalence of 0.2% and 1,000 patients: 2 patients, 998 healthy individuals. Positive D-dimer: 2 × 0.95 + 998 × 0.60 = 1.9 + 598.8 = 600.7. PPV = 1.9/600.7 = 0.3%. A positive D-dimer result in the absence of clinical suspicion has a predictive value close to zero. Conclusion: D-dimer is not a screening test. It is only useful following clinical risk stratification (Wells score ≤ 2 points).

Example 3: Type 2 diabetes mellitus in geriatric patients

Prevalence (Geriatrics-25 diagnoses): approx. 15–25% of geriatric patients. HbA1c ≥6.5% as a diagnostic criterion: sensitivity 47%, specificity 98%.

With a prevalence of 20% and 100 patients: 20 affected, 80 unaffected. HbA1c positive: 20 × 0.47 + 80 × 0.02 = 9.4 + 1.6 = 11.0. PPV = 9.4/11.0 = 85.5%. The high prevalence makes even a moderately sensitive test diagnostically valuable. However: the NPV is 88.3% – a negative HbA1c does not definitively rule out T2DM in geriatric patients. Consider a fasting blood glucose test as a second step.

Prevalence–test performance matrix: GP practice vs. hospital

The following table shows how the PPV of identical tests changes at different prevalence rates

Table: Change in PPV of identical tests at different prevalence levels

Illness	Test	Sens.	Specific	Prevalence in the general population	PPV GP	PPV clinical
Pulmonary embolism	D-dimer	95%	40%	0.5%	0.8%	24%
Depression	PHQ-9	88%	85%	12%	44%	78%
Hypertension	Clinic BP	75%	75%	35%	62%	86%
CHD	Stress ECG	68%	77%	5 %	13%	50%
T2DM (geriatric)	HbA1c	47%	98%	20%	85%	95%
Hypothyroidism	TSH	98%	92%	4%	34%	80%

Interpretation guide: Red = PPV <5% (test virtually useless as a confirmatory test). Yellow = 5–70% (results only useful in clinical context). Green = >70% (test diagnostically reliable in this setting).

1.3 Test performance measures – What every GP needs to know

Test performance measures describe the effectiveness of diagnostic tests regardless of prevalence. The four basic measures are: sensitivity, specificity, positive likelihood ratio (LR+) and negative likelihood ratio (LR–).

Sensitivity and specificity

Sensitivity: the proportion of patients with the condition whom the test correctly identifies as positive. High sensitivity = few false negatives = good for ruling out the condition (SnNOut: Sensitivity Negative rules Out).

Specificity: the proportion of healthy individuals whom the test correctly identifies as negative. High specificity = few false positives = good for confirmation (SpPIn: Specificity Positive rules In).

Rule of thumb for SnNOut / SpPIn:

Sensitivity >95%: A negative result rules out the disease with a high degree of certainty.

Specificity >95%: A positive result confirms the disease with a high degree of certainty.

No test has both figures above 95 per cent – which is why sequential diagnosis is used.

Likelihood ratios – the most practical measure

Likelihood ratios (LR) are prevalence-independent and express the extent to which a test result alters the probability of a condition. They are the key tool for Bayesian diagnosis in GP practice.

LR+ (positive LR): = sensitivity / (1 – specificity). Indicates by what factor a positive test result increases the probability.

LR– (negative likelihood ratio): = (1 – sensitivity) / specificity. Indicates by what factor a negative test result reduces the probability.

Table: Examples of the prevalence-independence of likelihood ratios (LR)

LR range	Interpretation	Example (LR+)	Example (LR–)
>10	Strong evidence – significantly alters the diagnosis	Troponin in ACS	
5–10	Moderate evidence – clinically highly relevant	Wells + D-dimer negative	
2–5	Weak evidence – helpful, but insufficient	CRP >50 in pneumonia	
1–2	Virtually worthless – hardly alters the probability	ESR alone	
<0.1	Strongly negative – points against the diagnosis		D-dimer negative in DVT
0.1–0.2	Moderately exclusionary		Normal TSH in thyroid
0.2–0.5	Weakly exclusionary – further tests required		Normal ultrasound in cholecystitis

The Fagan nomogram logic in brief

The Fagan nomogram visually links pre-test probability, likelihood ratio and post-test probability. For everyday clinical practice, the following rule of thumb suffices:

Bayesian rule of thumb for GPs:

Step 1: Estimate the pre-test probability (based on prevalence and clinical presentation).

Step 2: Choose a test with known LR+ and LR–.

Step 3: Pre-test probability × LR = post-test odds.

Or more simply: Pre-test odds = P / (1–P). Post-test odds = Pre-test odds × LR. Post-test probability = Odds / (1+Odds).

Practical shortcut: For low prevalence (<5 %):

LR+ > 10 → a positive test result is interpretable (PPV rises to >30%)

LR+ 2–5 → a positive test alone is not diagnostic

LR– < 0.1 → a negative test reliably rules out the condition

1.4 Sequential diagnostics – the cascade rather than the battery

Sequential diagnostics is the opposite of a simultaneous test battery. Instead of requesting 15 laboratory parameters at once, a baseline test is carried out, the result of which determines the next step. Each test updates the probability, and the subsequent test addresses the remaining uncertainty.

The principle: Test → Result → Decision → Next test

The diagnostic pipeline follows a strict sequence: (1) Formulate a clinical hypothesis. (2) Select the test with the highest diagnostic yield. (3) Interpret the result (calculate the post-test probability). (4) Decision: Is the diagnosis confirmed? Is the diagnosis ruled out? Is further investigation required? (5) Only if uncertainty remains: order the next test.

Case study: Thyroid dysfunction

Clinical suspicion: weight gain, fatigue, intolerance to cold in a 55-year-old female patient. Prevalence of hypothyroidism (Endocrinology-25 diagnoses): approx. 5–10% in women over 50.

Step	Test	Result	Probability of repeat test	Consequence
1	TSH	TSH 8.5 mU/l (elevated)	70–80% for hypothyroidism	Continue to step 2
2	fT4	fT4 0.6 ng/dl (low)	>95% manifest hypothyroidism	Diagnosis confirmed
Alt. 2	fT4	fT4 1.1 ng/dl (normal)	Subclinical hypothyroidism	Follow-up in 3 months

If TSH, fT4 and fT3 had been requested simultaneously, two unnecessary tests would have been paid for if the TSH had been normal. In 80 per cent of cases, TSH is normal – the sequential strategy saves 80 per cent of follow-up tests.

Mathematical cascade: multi-stage Bayesian updating

In sequential diagnosis, the post-test probability from test 1 becomes the pre-test probability for test 2. The maths:

Pre-test odds$_1$ = P_0 / (1–P_0) → Post-test odds$_1$ = Pre-test odds$_1$ × LR_1 → P_1 = Post-test odds$_1$ / (1 + Post-test odds$_1$)

Pre-test odds$_2$ = P_1 / (1–P_1) → Post-test odds$_2$ = Pre-test odds$_2$ × LR_2 → P_2 = final result

Application: Diagnosis of pulmonary embolism

Prevalence in GP practice (25 cardiology diagnoses): approx. 0.1–0.5%. Female patient, aged 68, presenting with sudden dyspnoea and unilateral leg oedema.

Step 1: Wells score = 5 points → moderate clinical probability → pre-test probability adjusted to approx. 25 per cent.

Step 2: D-dimer negative (LR– = 0.05). Pre-test odds = 0.25/0.75 = 0.33. Post-test odds = 0.33 × 0.05 = 0.017. Post-test probability = 0.017/1.017 = 1.7%. Conclusion: PE ruled out with a high degree of certainty.

Step 2 (old): D-dimer positive (LR+ = 1.6). Post-test odds = 0.33 × 1.6 = 0.53. Post-test probability = 34.6%. Conclusion: CT angiography indicated (Step 3).

1.5 The logic behind the 90-second flowcharts

The desk-based tools from Volume 2 (Practice Pearls) use 90-second flowcharts as decision-making aids. These flowcharts are not random decision trees – they encode Bayesian logic into clinically manageable yes/no cascades.

Design principles of the flowcharts

Principle 1 – Red flags first: The first decision node always screens for dangerous clinical courses (exclusion of emergencies). Rationale: Even where prevalence is low, a dangerous clinical course must never be missed. The asymmetric consequence (death versus overdiagnosis) necessitates this prioritisation.

Principle 2 – Highest prevalence first: Following the exclusion of red flags, the differential diagnoses are listed in descending order of prevalence. Rationale: The lists of 25 diagnoses show that the five most common diagnoses in each speciality account for 60–80 per cent of all cases. Checking these five first completes the bulk of the diagnostic work.

Principle 3 – Best available test per node: At each decision point, the test with the highest diagnostic information gain (highest LR) is selected. Not the most expensive or most advanced test, but the one with the greatest shift in probability.

Principle 4 – Clear endpoint: Each pathway ends with a specific course of action: diagnosis + treatment, referral, watchful waiting + re-evaluation, or emergency admission.

Example: Flowchart logic for chest pain

The cardiology desk-side tool flowchart for chest pain follows this Bayesian cascade:

Node	Question	Test / Findings	Probability Shift	Next Action
1	Red flags? (acute, excruciating pain, dyspnoea, hypotension)	Clinical assessment	If yes: ACS/PE/aortic dissection >50%	Emergency doctor / immediate ECG + troponin
2	Exercise-related? Feeling of pressure? Risk factors?	History-taking + Wells/HEART	Probability of CHD 5–30%	Resting ECG + troponin
3	ECG changes? Elevated troponin?	ECG + troponin	Positive likelihood ratio for troponin: >15 in ACS	Induction / Catheterisation Laboratory
4	Respiratory? Pleuritic?	Clinical assessment	Pleurisy/pneumonia 20–40 %	Chest X-ray + CRP
5	Tenderness on pressure? Dependent on movement?	Palpation	Muscular/skeletal 50–70 %	Analgesia + clinical course

The flowchart logic encodes: (1) Asymmetrical presentation (red flags), (2) prevalence-based triage, (3) best test per decision point, (4) clear endpoint.

1.6 Five rules of thumb for rational diagnosis by GPs

Five practical rules of thumb emerge from Bayesian theory, the prevalence data from the 25-diagnosis lists and experience with the desk-based tools:

Rule 1: No test without a hypothesis.

Every test ordered must answer a formulable clinical question. If the question cannot be specifically identified before the test is ordered, the test should not be ordered. The most common breach of this rule: routine check-up laboratory tests without a specific indication.

Rule 2: Prevalence determines the test's value.

The same test with the same sensitivity and specificity has a completely different predictive value in a GP practice than in a hospital. A stress ECG with a 5 per cent prevalence of CHD has a PPV of 13 per cent – in a cardiology outpatient clinic with a 30 per cent prevalence, it has a PPV of 56 per cent. Conclusion: Always take your own patient population into account.

Rule 3: Negative test results are more valuable than positive ones.

In the low-prevalence setting of a GP practice, the NPV is almost always higher than the PPV. Diagnostics serve primarily to rule out conditions. A normal TSH level virtually rules out thyroid dysfunction (NPV >99 per cent at a prevalence of 4 per cent). A positive troponin result in a GP practice, by contrast, has a PPV of only 40–60 per cent – it always requires confirmation.

Rule 4: Sequential testing is better than simultaneous testing.

Stepwise diagnosis is mathematically and economically superior to a battery of tests. Each test updates the probability; the subsequent test specifically addresses the remaining uncertainty. Compared with ordering tests simultaneously, stepwise diagnosis (TSH → fT4 → TPO-antibody) saves two tests in 80 per cent of cases.

Rule 5: Time is a diagnostic test.

Controlled watchful waiting with planned re-evaluation (safety-netting) is an active diagnostic strategy, not an indication of inaction. In cases of acute non-specific back pain without red flags, a 6-week re-evaluation has better diagnostic accuracy than an immediate MRI scan.

Summary of Chapter 1

Diagnosis in GP practice follows Bayesian logic: prevalence → pre-test probability → test (with known likelihood ratio) → post-test probability → decision.

The lists of 25 diagnoses provide the prevalence data, the likelihood ratios provide the test evaluation,

and the flowcharts encode this logic into 90-second algorithms.

The five rules of thumb translate the theory into everyday clinical practice.

Cross-references

Work 1 (GP Manual): Chapter 7 contains the epistemological rationale for hypothesis-driven diagnostics and the diagnostic pipeline. This chapter systematises and quantifies these principles.

Work 2 (Practice Pearls): The flowcharts of the desk-based tools are applications of the Bayesian logic described here. Section 1.5 explains their design principles.

E0 products: The master list of 25 diagnoses (Q31) provides the prevalence data for all calculation examples. The PRISCUS register (Q30) is relevant in Chapter 2 (Laboratory Diagnostics) for PIM-induced laboratory abnormalities.

Subsequent chapters: Chapter 2 (Laboratory strategy), Chapter 3 (Imaging) and Chapters 4–6 (Organ system diagnostics) build on these foundations.

7. Micronutrients & hormone panels – Evidence vs. hype

Diagnostic stewardship in the face of social media-driven pressure to test

Diagnostic stewardship, clarification of indications, expectation management and protection against overdiagnosis

7.1 Introduction: Why this chapter

Every day, GP practices see patients who arrive with screenshots, podcast recommendations or coaching reports and request a ‘complete micronutrient panel’, a ‘hormone profile’ or a ‘silent inflammation panel’. These tests are marketed on social media as the key to energy, prevention and longevity.

The clinical challenge here is not whether micronutrients, hormones or inflammatory markers are biologically important – they undoubtedly are. The question is whether testing and supplementation in the form described are in the patients’ best interests.

✔ Note

- Biologically plausible does not automatically mean clinically useful.
- A test is useful if it influences a decision – not if it merely produces figures.
- Diagnostic stewardship protects against overdiagnosis, testing cascades and a false sense of security.
- This chapter shows when tests constitute high-value care – and when they do not.

7.2 The three major ‘hype’ clusters

Three recurring patterns dominate the list of requested tests in GP practices:

Cluster	Typical narratives	Core risk	GP guidelines
1. Micronutrients	*“Every woman’s ferritin level must be 100.” / “Your panel shows which vitamins you’re lacking.”*	Incidental findings, long-term supplementation without deficiency, cascades, costs	Ferritin + beta-blocker only if deficiency is suspected. B12, vitamin D, zinc: only if there is a clinical concern or risk. Replacement therapy = treatment, not a lifestyle choice.
2. Hormone panels	*“Low T explains everything.” / “DHEA/cortisol saliva test indicates adrenal insufficiency.”*	Hormone panel without clear indication, replacement therapy without diagnosis, diagnostic cascade	Testosterone: only for sexual symptoms + in line with guidelines. DHEA/cortisol saliva test: not standard practice for GPs. Sleep/alcohol/stress first.
3. Silent inflammation	*“Silent inflammation is the cause of 90 per cent of all diseases.” / “The hs-CRP panel indicates the risk of a future heart attack.”*	Panels without actionable results, incidental findings, false reassurance or overtreatment	CRP/ESR: clinical markers for specific clinical suspicion. Cardiovascular risk is better managed using traditional parameters.

7.3 Seven essential questions before every decision to test

These seven questions help to provide a diagnostic framework for any requested test – regardless of the specific substance.

7 essential questions to ask before every decision on a requested test		
1.	**What diagnosis does the patient suspect?**	*Take the hypothesis seriously; use the patient's own words.*
2.	**What is the pre-test probability?**	*Clinical context first – before laboratory tests.*
3.	**What does a positive result change?**	*If nothing changes: no test.*
4.	**What does a negative result change?**	*Can the patient put it behind them?*
5.	**What incidental findings does the test trigger?**	*Anticipate cascades.*
6.	**How do we interpret reference ranges?**	*Normal range ≠ healthy, elevated levels ≠ ill.*
7.	**Who bears the costs – and is that justified?**	*Paying out of pocket ≠ automatically sensible.*

7.4 Traffic-light table: micronutrients, hormones, inflammation markers

The following table summarises the most important test areas. It is structured according to the traffic light principle: GREEN = clear indication, YELLOW = context-dependent, RED = no evidence-based benefit in a routine setting.

All information is based on current guidelines (ERC, EFSA, NICE, Endocrine Society). Self-pay panels and social media recommendations alone do not constitute an indication.

Substance / Test area	Traffic light	Appropriate for	Not for routine use	Key point / GP alternative
— MICRONUTRIENTS				
Vitamin D (25-OH-D)	**GREEN**	Osteoporosis, fragility fractures, chronic malabsorption, bariatric surgery, frail elderly people, those in care	Healthy individuals < 65 years with no risk factors; longevity panels	*"Treating vitamin D deficiency – vitamin D screening in healthy individuals is rarely high-value." Test only if the implications are clear.*
Vitamin D – Replacement therapy	**YELLOW**	Proven deficiency or high-risk profile	Long-term high-dose treatment without medical indication, based on social media recommendations	*"Aim: to correct deficiency – no long-term supraphysiological high-dose therapy." Toxicity possible.*
Vitamin B12	**GREEN**	Neurological symptoms of unclear cause, macrocytosis, vegan diet, metformin, bariatric surgery, malabsorption, aged over 70	Asymptomatic individuals without a risk profile; longevity panels	*"B12 is not a 'wellness' laboratory parameter – test only in cases of risk or clinical suspicion."*
Ferritin + small beta-blocker	**GREEN**	Suspected iron deficiency: fatigue + heavy menstruation, restless legs, signs of anaemia, suspected GI bleeding	"Ferritin optimisation" in cases of normal haemoglobin + asymptomatic; based solely on a social media post	*"Ferritin is a marker of deficiency – not a wellness optimisation value." Always combine with beta-blocker.*
Magnesium (serum)	**YELLOW**	Diuretic therapy, alcohol dependence, arrhythmia, refeeding syndrome risk	Healthy individuals without specific clinical symptoms; sleep/fatigue panel	*"Serum Mg is a rough marker – can only be meaningfully interpreted in*

Substance / Test area	Traffic light	Appropriate for	Not for routine use	Key point / GP alternative
				conjunction with clinical findings."
Selenium	**YELLOW**	Severe malabsorption, enteral/parenteral nutrition, specific oncological issues	Prevention panels, detox check-ups in healthy individuals	*"Selenium is essential – but routine screening in the general population is not high-value care."*
Zinc	**YELLOW**	Malabsorption syndromes, severe malnutrition, specific dermatological issues	Immune check, acne blog panels, longevity kits	*"Zinc deficiency possible – but not a standard screening marker in GP practices."*
Omega-3 index	**RED**	– (Exception: specialist lipid consultation)	Longevity and wellness panels; management of supplement doses	*"Not a routine test in GP practice. Nutritional consultation instead of a fatty acid profile."*
Homocysteine	**YELLOW**	Specific cardiovascular/neurological concerns, rare methylation-related genetic conditions, uncertainty regarding B12/folate	General preventive panels, dementia screening without clear implications	*"Only test if the result has clear therapeutic implications."*
CoQ10	**RED**	If applicable, in cases of statin-associated myalgia + study protocol; not routine	Mitochondrial/longevity panels in healthy individuals	*"Biologically plausible does not automatically mean clinically relevant."*
— HORMONE PANELS				
Testosterone (serum, morning)	**YELLOW**	Men: loss of libido + erectile dysfunction + significant impact on daily life according to EAU/DGGG guidelines	Low-T screening without clear sexual symptoms; saliva tests; biohacking panels	*"Test testosterone only if a guideline-based treatment pathway is possible." Morning measurement is essential.*
DHEA, cortisol saliva profile	**RED**	– (only in specific endocrinological cases)	"Adrenal insufficiency" panel, burnout assessment, daily profiles without validated interpretation	*"Stress is real – but salivary DHEA/cortisol is not a standard tool for GPs."*
Thyroid (TSH)	**GREEN**	Symptoms (fatigue, weight, palpitations, hot flushes,	Routine screening in asymptomatic individuals (no clear	*"TSH as first-line – fT3/fT4/rT3 only where there is a*

Substance / Test area	Traffic light	Appropriate for	Not for routine use	Key point / GP alternative
		cold intolerance, bradycardia)	USPSTF recommendation for healthy individuals)	*specific clinical indication, not routinely."*
fT3, fT4, reverse-T3	**YELLOW**	Pathological TSH or clinically justified suspicion of thyroid disorder	Panel for fatigue without abnormal TSH; reverse-T3 without an endocrinological context	*"Reverse T3 is not a GP parameter – no evidence-based indication for routine testing."*
Oestrogen/progesterone panels (women)	**YELLOW**	Menopausal transition with typical symptoms + need to decide on hormone therapy	Panels without a specific treatment decision; wellness screening for young, asymptomatic women	*"Hormone status: yes – if it facilitates shared decision-making. Not as a general energy screening."*
— INFLAMMATION MARKERS				
CRP / ESR as needed	**GREEN**	Suspected inflammation: infection, IBD, rheumatology, suspected tumour	"Silent inflammation" screening in healthy individuals without clinical concerns	*"CRP is a clinical marker – not an oracle for every non-specific symptom."*
hs-CRP (targeted)	**YELLOW**	Cardiovascular risk calculator (e.g. SCORE2) in borderline cases with clinical implications; only if it changes the clinical decision	Multi-marker wellness panels; solely out of concern for silent inflammation	*"hs-CRP only if it actually alters the management decision."*
Multi-marker panels (interleukins, TNF-α, etc.)	**RED**	–	Longevity packages without actionable results; self-pay wellness kits	*"More markers increase the number of incidental findings – not patient safety."*
— LONGEVITY PANELS				
Longevity/Biomarker Comprehensive Panel	**RED**	–	Comprehensive panels with 30–50 parameters without guideline-based recommendations	*"The best way to extend life: guideline-based prevention, exercise, giving up smoking – not 40-parameter lab test sets."*

7.5 Case studies: Typical consultation scenarios

The following vignettes illustrate the complete process from the ICE discussion to the decision on testing. They can be used directly as a training resource or as a consultation aid.

Vignette: Ferritin / 'Iron deficiency without anaemia'

Vignette: Ferritin optimisation
Female patient, aged 35. Brings a screenshot from a social media post: "If your ferritin isn't at 100, you're chronically tired." Symptoms: tiredness, hair loss. Requests a full iron panel (ferritin, transferrin, transferrin saturation, CRP).

ICE – Ideas / Concerns / Expectations	
Ideas (Perception)	"I have iron deficiency – the post said that almost all women have levels that are too low, even with a normal blood count."
Concerns	"My tiredness is being ignored. I want to rule out iron deficiency as the cause."
Expectations	"I'd like the full iron panel and then to take iron supplements specifically until my levels are just right."

Approach

- **Check for red flags:** Heavy menstruation (in this case: yes, subjectively has become heavier) → pre-test probability of iron deficiency moderately increased. No anaemia symptoms, no gastrointestinal bleeding, no cardiac warning signs.
- **Translate the patient's request:** "The clinical question is: Is there an iron deficiency that explains your symptoms – and do we need to look for a source of bleeding?"
- **Test decision (traffic light system):** GREEN: Ferritin + basic blood count. RED: comprehensive micronutrient panel, additional inflammatory markers with no evidence of systemic disease.

Communication building blocks
"I see a real possibility of iron deficiency here – particularly given your heavier bleeding. It makes sense to measure ferritin and carry out a complete blood count specifically for this. A comprehensive micronutrient panel, on the other hand, offers you little additional benefit and, above all, increases the chance of incidental findings."
"Ferritin normal, haemoglobin normal → no iron supplement. Long-term self-medication without a deficiency offers no benefit but does have side effects." *"Low ferritin + corresponding symptoms → deficiency confirmed; investigate the cause (gynaecological, possibly GI), then provide targeted replacement therapy – with scheduled follow-up."*

Case study: Hormone panel – ‘Low-T’ and adrenal insufficiency

Case study: Men’s health panel

Patient, aged 48, works shifts, drinks 2–3 beers per evening, suffers from sleep deprivation. Brings printouts from an online coach: testosterone (serum + saliva), DHEA, 24-hour cortisol profile, SHBG, oestradiol. Complaints: exhaustion, weight gain, reduced energy levels. Coach: “That’s typical of low T.”

ICE – Ideas / Concerns / Expectations	
Ideas (Expectations)	"I have low testosterone or adrenal fatigue – my coach says it’s a classic condition for men of my age who are under stress."
Concerns	“I’m worried that my performance will keep getting worse.”
Expectations	“I’d like to have the full hormone panel done and then take testosterone or DHEA replacement therapy.”

Approach

- **Check for red flags:** No B-symptoms. Libido slightly reduced, but no clear evidence of erectile dysfunction. No indication of overt endocrinopathy. Shift work, lack of sleep, alcohol = significant confounding factors.
- **Interpreting the patient’s request:** “The clinical question is: Is there evidence of hypogonadism requiring treatment – not simply a desire to optimise levels?”
- **Test decision (traffic light system):** YELLOW: Morning testosterone test only if clear sexual symptoms (loss of libido + erectile dysfunction + impairment of daily life) are present. RED: DHEA panel, salivary cortisol profiles, complete biohacking panel.

Communication building blocks

“I can see many factors that explain your fatigue well – shift work, lack of sleep, regular alcohol consumption. Against this background, a comprehensive hormone panel is a red light – it rarely provides answers, but often yields incidental findings.”

“If you’d like, we can check whether your symptoms meet the criteria for a serious testosterone deficiency. Only then does a targeted test make sense.”

“Testosterone replacement therapy is only indicated following a confirmed diagnosis of hypogonadism. A single, slightly reduced value in the presence of non-specific fatigue is not in itself a clear-cut indication for testosterone therapy.”

⚠ The substitution trap

- Testosterone without confirmed hypogonadism: cardiovascular risk, effects on the prostate, endocrine suppression.
- DHEA supplementation without evidence of patient-relevant outcomes.
- Salivary cortisol profile: no validated GP-related interpretation – creates uncertainty without leading to any action.

ANNEXES

CLINICAL OPERATING SYSTEM FOR PRIMARY CARE

A1. Diagnostic value of physical examination findings

Evidence · Reliability · Practical application in general practice

Based on Evidence-Based Physical Diagnosis (McGee) and JAMA Rational Clinical Examination

General Practice Knowledge Base for Primary Care · 2026

Key clinical message

Physical examination findings in general practice are almost always ‘rule-in’ tools — rarely definitive ‘rule-out’ instruments.

The absence of a finding only rules out a diagnosis if the negative likelihood ratio (LR−) is small enough — and even then, only in the correct context.

The biggest methodological pitfall: the negative predictive value (NPV). In GP practice, it is automatically high due to low prevalence — even without a clear negative finding.

The key factors are: LR+, LR−, pre-test probability (setting!), reliability (Kappa) and — above all — consistency in management.

Core rule: if the findings do not alter the course of action, we do not (yet) need these findings.

Red flags override any statistics. They alter the course of action regardless of LR values.

Methodology: What lies behind the figures

Before findings can be used as clinical tools, three fundamental methodological principles must be understood. They protect against the most common misinterpretations in GP practices.

LR+ and LR−: The actual key figures

Concept	Explanation
LR+ (positive likelihood ratio)	How much more likely is a positive finding in patients with the condition compared to healthy individuals? LR+ > 10 = strong rule-in. LR+ 5–10 = moderate rule-in. LR+ 2–5 = limited rule-in. LR+ < 2 = of little clinical use.
LR− (negative likelihood ratio)	How much more likely is a negative result in healthy individuals compared to those with the condition? LR− < 0.1 = strong rule-out. LR− 0.1–0.2 = moderate rule-out. LR− 0.2–0.5 = limited rule-out. LR− > 0.5 = weak, of little diagnostic value.
Pre-test probability	The same LR can have very different implications depending on the setting. An LR+ of 5 with a pre-test probability of 5 per cent (GP) results in a post-test probability of approximately 21 per cent. The same LR+ with a pre-test probability of 40 per cent results in approximately 77 per cent. The setting is decisive.
Kappa / Inter-rater reliability	Measures how well two independent examiners arrive at the same finding. κ > 0.60 = good. κ 0.40–0.60 = moderate. κ < 0.40 = poor. A finding may be statistically strong (high LR) yet of little clinical use (low κ).

⚠ The NPV trap — the most common misinterpretation in GP practices

The negative predictive value (NPV) indicates how often a negative test result is actually correct.

Problem: In GP practices, many conditions are rare (low prevalence). This automatically results in a high NPV — even with a poor test.

Example: Absence of rectal bleeding in colorectal cancer: NPV very high (~99%), but sensitivity only ~30%. The high NPV is misleading — it results from the low prevalence of cancer, not from the test's accuracy.

GP's diagnosis of 'no pneumonia': NPV 96%, but sensitivity only 29%. The GP is wrong in 71% of genuine pneumonia cases!

Rule: Never use the NPV on its own as evidence of exclusion. Always state the likelihood ratio (LR) and pre-test probability.

Visual legend — symbols in all tables

All findings cards in this chapter use the same symbol system. The legend applies to all modules.

Diagnostic strength

◍◍◍ **strong:** LR+ >10 or LR− <0.1 — significantly alters probability

◍◍ **moderate:** LR+ 5–10 or LR− 0.1–0.2 — relevant change

◍ **limited:** LR+ 2–5 or LR− 0.2–0.5 — contributes to the overall picture

● **weak:** LR+ <2 or LR− >0.5 — of little clinical value

Reliability (Kappa) · Practice · Category

●●● good (κ >0.60)
●●○ moderate (κ 0.40–0.60)
●○○ weak (κ < 0.40)
? insufficient data

☑ simple & easy to use
⚠ Depends on the exercise or context
✕ do not use in isolation

Rule-in: a positive result increases the WSK significantly
Rule-out: negative result reduces WSK significantly
Risk marker: warning sign, not conclusive
Contextual finding: only meaningful within a pattern
Do not isolate: overestimated / no clinical value

Module 1: Dyspnoea — suspected heart failure

Confirm classic signs, but rarely rule out the condition

Clinical key question

Does this patient with dyspnoea have new-onset or decompensated heart failure — and can I rule it out by physical examination?

Setting & pre-test probability

GP practice: Pre-test probability of heart failure in patients with dyspnoea aged > 60 years: approx. 28–30%.

Data source: predominantly McGee's Evidence-Based Physical Diagnosis, outpatient setting. Reliability data (Kappa) for most signs of heart failure are not sufficiently substantiated.

Important: In patients with no pre-clinical findings, the same likelihood ratio (LR) may result in a different post-test probability of heart failure than in a specialist outpatient setting.

Findings sheet

Findings	LR+	LR−	Strength	Reliability	Practice	Category	Consistent management
3. Heart sound present	~21.5	~0.96	●●●	?	⚠	Rule-in (rare!)	heart failure highly probable → NT-proBNP, ECG, echocardiogram, immediate referral to cardiology
Positive abdominal-jugular test	~8	~0.3	●●	?	⚠	Rule-in / Context	If performed correctly: best clinical marker of heart failure → investigate further
Cervical venous pressure	~4.8–5	~0.85	●	?	⚠	Rule-in	Positive finding; negative result does not rule out the condition — depends on the examiner
Basal crackles	~2.95	~0.81	●	?	☑	Context	Supports the suspicion; absence provides little reassurance (LR− 0.81!)
Orthopnoea present	~1.56	~0.87	●	?	☑	Context	A piece of the clinical picture; not a rule-out criterion; not a standalone decision-maker
Absence of dyspnoea when climbing <1 flight of stairs	—	~0.62	●	?	☑	Weak rule-out	LR− 0.62 — not sufficient to rule out heart failure

Traffic light overview: heart failure

3. Heart sounds / abdominal-jugular test Strong rule-in if reliably detected; rare and severe	**Jugular venous distension / crackles** support the suspicion; their absence provides little reassurance	**"No crackles → no heart failure"** Dangerous misinterpretation — LR− 0.81	**Dyspnoea + hypoxia + tachycardia**

Management matrix

Clinical situation	Positive findings	Negative findings → Consequence
Dyspnoea + third heart sound / jugular venous distension / oedema	Heart failure likely → NT-proBNP, ECG, echocardiogram, referral to cardiology if necessary	—
Dyspnoea without classic signs	A single finding may warrant further investigation	Heart failure not ruled out → if there is a relevant suspicion, still perform NT-proBNP/ECG
Dyspnoea + hypoxia / chest pain / syncope	Do not manage as an outpatient — follow the emergency pathway	Do not manage as an outpatient — follow the emergency pathway
Unclear exertional dyspnoea in an elderly patient	Physical examination + NT-proBNP + ECG	Physical examination + NT-proBNP + ECG — not just auscultation
Positive abdominal-jugular test	Relevant heart failure marker → Further investigation	Limited negative predictive value — safety netting in cases of persistent dyspnoea

⚠ Typical misinterpretations of heart failure

❌ "No crackles → no heart failure." — LR− 0.81: not a rule-out criterion.

❌ "Absence of third heart sound → reassuring." — LR− 0.96: practically no rule-out.

❌ "No orthopnoea → probably no heart failure." — LR+ 1.56: too non-specific for a decision.

❌ "Oedema → heart failure." — Oedema is neither sensitive nor specific for heart failure; there are many other causes.

☑ Correct: Classic signs of heart failure tend to confirm rather than rule out the condition. NT-proBNP/ECG remains decisive.

Safety-netting for heart failure

"In cases of dyspnoea, I cannot rely on physical findings alone to rule out the condition. If there is a relevant clinical suspicion, I always carry out an NT-proBNP test and ECG — even if auscultation is unremarkable. Come back immediately if your shortness of breath worsens, you can no longer lie flat, or your legs swell more severely."

Module 2: Chest pain — PE · DVT · ACS

Dangerous patterns first — individual signs are poor predictors

Key clinical question

Is this chest pain dangerous (ACS, PE, pneumothorax, myocarditis) — or is it likely to be of musculoskeletal, reflux-related or functional origin?

Setting & pre-test probability

GP practice: Chest pain — ACS < 5–15 per cent; PE 1–3 per cent; musculoskeletal causes 25–40 per cent; functional/reflux 15–30 per cent.

DVT: Classic signs (Homans' sign, swelling, tenderness) are poor predictors — DVT+ and DVT− cases often look similar.

Key rule: Algorithms (Wells, PERC, HEART) are more reliable than isolated individual findings. Young age alone does not rule out PE, myocarditis or pneumothorax.

Findings chart

Findings	LR+	LR−	Strength	Reliability	Practice	Category	Consistency of management
Absence of sudden-onset dyspnoea (in the context of PE)	—	~0.009	●●●	?	☑	Strongly rules out (PE)	Significantly reduces PE risk — ONLY in the appropriate context; not a substitute for Wells/PERC in cases of high VT risk
Pleural pain + dyspnoea / haemoptysis	Red flag pattern	—	●●●	☑	☑	Rule-in PE / Red Flag	→ PE algorithm / emergency pathway immediately
High Wells score for DVT (combined)	~13.2	~0.2	●●●	☑	☑	Score (no single finding)	→ D-dimer / compression ultrasound
Reproducible on movement of the chest wall	context-dependent	—	○	?	☑	Context / Thoracic wall	Suggests a musculoskeletal cause — rule out red flags first!
Absence of dyspnoea in general (PE)	—	~0.61	●	☑	☑	Weak rule-out	LR− 0.61: do not reassure the patient on this basis alone; continue with the algorithm

Explore the complete ClinicalOS edition

Volume 2 covers prevalence-weighted diagnostics, sequential testing, laboratory and instrumental diagnostics, overdiagnosis and diagnostic heuristics. Full edition and further volumes:

clinicalos.de