Masking Depth in Clinical Practice: A Follow Up to Star Face

by Grego (Gergely Földvári)

Budapest, 8 September 2026.

A survey of existing evidence bearing on the STAR FACE census, and a proposal.

Abstract. STAR FACE is a sieve model demonstrated on a polyhedron, introducing the term summetry, the discrete unity of sum and symmetry. The sieve descends from a population well over the number of atoms in the Solar System to 6,223,974 magic readings, the red cells in a single microlitre of blood, and terminates in a singularity: one numbering, unique up to the symmetry of the solid. Each of the twelve four-faced pyramids stands for one diagnostic reading; when two or more pyramids carry the same four values, one reading masks another or more that could be distinct and clear. This paper does four things. It separates the several unrelated senses the word masking carries in medicine. It corrects the statistic that a clinical comparison requires, replacing the census by shape with a census by the depth of the deepest mask. It surveys what the literature already contains at each masking depth, and finds the diagnostic odyssey of rare disease to be the only body of evidence reaching past depth two. And it sketches, for others better qualified to finish it, the survey that would test the model's one genuinely untested claim.

Companion to STAR FACE. Grego's Polyhedral Sieve Model for Medical Diagnostics, DOI 10.5281/zenodo.22259868, grego.hu/star.

DOI 10.5281/zenodo.22659189 · concept DOI 10.5281/zenodo.22659188, which always resolves to the current version

grego.hu/md · grego.hu/md.pdf · https://www.grego.hu/papers

In plain words

the whole paper in two paragraphs, written to be translated

A solid shape has twelve apexes. Four triangles meet at each apex. Every triangle carries a number. The numbers are placed so that the four triangles at every apex always add up to 26. Each apex is therefore one result, and each result is a set of four numbers. There are 6,223,974 ways to place the numbers. In some of these ways, two or more apexes show the same four numbers. Those apexes are then no longer distinct. Their results are identical, so nothing tells them apart. Ambiguity appears. The model lists every possible pattern of this sameness. There are nineteen patterns. At best, all twelve apexes show different results. At worst, only five different results remain. No more than four apexes ever show the same four numbers.

Now read each apex as one medical test result. A patient has twelve results. If all twelve are different, the diagnosis is clear. If two or more are identical, they cannot be told apart, and one illness can then hide behind another. The model says how often this can happen, and how many illnesses can hide at once: one clear answer in about 57 cases out of 100, two possible answers in about 42 out of 100, three in about 7 out of 1,000, and four in about 3 out of a million. Medicine has no such count. Doctors know that one illness can hide behind another, and they have several names for it, but no study has ever asked how many illnesses were hiding at the same time. The model gives numbers that can be checked. The question is whether real patients follow them.

1. The sieve, in one page

Escher’s Solid, the stellated rhombic dodecahedron, carries 48 congruent isosceles triangle faces meeting at 12 outer apexes and 8 inner apexes. Facing any inner apex projects a magic hexagram. Four of the eight hexagram stars are disjoint, sharing no face [36].

Numbering the 48 faces freely gives 48! of about 1.24×10^{61} arrangements, roughly ten thousand times the number of atoms in the Solar System. The sieves reduce this. Requiring every diametric quadruple, every diaquad, to sum to the magic constant 26, and requiring the four disjoint stars to carry different integer sets, leaves 8,366 combinations. Requiring in addition that every one of the twelve four-faced pyramids sum to 26 leaves 6,223,974 pyramid-magic numberings, counted up to the symmetry of the solid.

Why 26

The magic constant is not chosen. Symmetry tells you which sets of faces are equivalent; it says nothing about what they carry. Summetry adds one requirement, that equivalent sets sum alike, and the shape then fixes what the sum must be.

Within one star, three diaquads of four faces each exhaust the twelve faces and the values 1 to 12. Their total is 78, and three equal parts give 26. Across the solid, twelve pyramids of four faces each exhaust all forty-eight faces, and every value appears once per star, so four times in all. That total is 4 × 78 = 312, and twelve equal parts give 26 again. Two different partitions, at two different levels, and the same constant falls out of both.

Elimination confirms it. Requiring equal sums admits three arrangements: six pairs at 13, three quadruples at 26, and two sixes at 39. Only 26 survives contact with the four-faced pyramids, because a pyramid holds four faces and needs a quadruple sum. The other two are retired by the shape, not by preference.

Summetry, then, is symmetry that carries arithmetic: where symmetry says two sets of faces are equivalent, summetry requires them to sum alike, and since those sets exhaust a fixed total, the shape itself fixes what the sum must be.

That figure has a convenient physical size. The adult male red-cell reference interval is 4.6–6.2 million cells per microlitre [1]. The census sits at the top of it. The sieve therefore runs from the Solar System to a single drop, and the drop is the very unit a laboratory counts.

Below the drop, one further step. Among all 6,223,974 numberings, exactly one carries the shape (4,0,1), four masked pairs and one masked quadruple together: the singularity.

2. What masking means in medicine

Before any comparison can be made, one obstacle has to be cleared. The word masking is not one term in medicine. It is at least six different ideas sharing a single spelling, and the confusion is the reason the evidence has never been gathered in one place.

Senses that are this model’s sense

Diagnostic overshadowing. A patient already carrying one diagnosis has new symptoms attributed to that existing label rather than investigated afresh. The concept originates in intellectual disability research of the early 1980s and has since moved to physical illness in people with mental illness, where a 2022 concept analysis and a 2026

mixed-methods systematic review both define it as the misattribution of the symptoms of one illness to an already diagnosed comorbidity [2, 3, 4]. Here the mask is a prior label.

Pharmacological masking. Treatment itself conceals the reading. Beta-adrenergic blocking agents mask the warning signs of acute hypoglycaemia and the tachycardia of thyrotoxicosis, and have been argued to mask hypotension as well [5]. Corticosteroids blunt the signs of infection. Here the mask is the therapy.

Symptom-disease pairing. A benign label conceals a dangerous disease, the canonical example being dizziness and stroke. This is the subject of section 4.4 below.

Senses that are not

Masking also means blinding in a clinical trial, so that a registry entry reads Masking: None as a statement about study design. It means the tone played into one ear during audiometry. It means perceptual masking in vision and hearing research. It means the deliberate concealment of autistic or attention-deficit traits, usually called camouflaging. It means a piece of cloth over the face. None of these is a diagnostic reading concealing another diagnostic reading.

Why this matters here

Simply put: doctors already have a name for what happens when a patient’s existing label swallows their new symptoms. The literature explains how it happens and why it is harmful. But nobody ever counts. No study states how many other conditions were hiding behind the label. One, two, three? The question is not asked, so there is no number to set beside the census.

And because the word carries several unrelated meanings, the separate literatures never meet. A paper on overshadowing does not cite the beta-blocker labelling. Neither cites the symptom-disease pair work. A search on masking returns mostly blinded trials. Four bodies of evidence describe one idea under four names, with no shared vocabulary and no shared measurement.

This sharpens the claim this paper makes at its close. It is not that nobody has looked at masking; that would be easy to disprove. It is that four literatures describe one phenomenon, none cites the others, and none asks how many readings coincide.

3. Two ways of counting, and the one the clinic needs

Everything that follows rests on one table, so it is printed here in full. Each row is a reading-shape: a pattern of coincidence among the twelve pyramids. Nineteen shapes occur across all 6,223,974 pyramid-magic numberings, counted up to the symmetry of the solid.

shape	reading	distinct reads	numberings	share
(0,0,0)	all clear	12	3,538,104	56.85 %
(1,0,0)	one pair	11	2,079,465	33.41 %
(2,0,0)	two pairs	10	505,208	8.12 %
(3,0,0)	three pairs	9	56,916	0.91 %
(4,0,0)	four pairs	8	3,013	0.048 %
(5,0,0)	five pairs	7	9	0.0001 %
(6,0,0)	six pairs	6	91	0.0015 %
(0,1,0)	one triple	10	21,287	0.342 %
(1,1,0)	pair + triple	9	14,982	0.241 %
(2,1,0)	two pairs + triple	8	4,132	0.066 %
(3,1,0)	three pairs + triple	7	160	0.0026 %
(0,2,0)	two triples	8	198	0.0032 %
(1,2,0)	pair + two triples	7	314	0.0050 %
(0,3,0)	three triples	6	78	0.0013 %
(0,0,1)	lone quad	9	3	< 0.001 %
(1,0,1)	pair + quad	8	8	0.0001 %
(2,0,1)	two pairs + quad	7	3	< 0.001 %
(4,0,1)	four pairs + quad, the singularity	5	1	< 0.001 %
(1,1,1)	pair + triple + quad	6	2	< 0.001 %
	total		**6,223,974**	**100 %**

How to read the table

Start with the first row. Twelve distinct reads, the all-clear, means every one of the twelve diagnostic instruments' test results is balanced, unambiguous and reliable: a clear diagnosis is evident. This is the commonest case in the model, 56.85 per cent of all numberings. Whether it is also the commonest case in the clinic is the question section 8 exists to ask.

Picture the twelve pyramids laid out in a row, each showing its own four numbers, each set adding to 26.

Suppose two of them show the same four. Those two are a masked pair: two readings behind one appearance. Of the twelve results in the patient's file, only eleven are now

distinct readings, namely ten pyramids each carrying their own 26-sum quadruple, plus the one quadruple that the two masked pyramids share between them. That is the eleventh reading, and that is why the row for one pair shows eleven distinct reads.

Now suppose instead that three of the twelve show the same four values. That is a masked triple: three readings behind one appearance. The file now holds only ten distinct readings, namely nine plus the one shared by the three. Hence ten distinct reads on the row for one triple.

The shape itself is written (pairs, triples, quads). It counts how many separate blocks of each size the numbering contains. So (2,1,0) means two masked pairs and one masked triple somewhere among the twelve, and (0,0,1) means one masked quadruple and nothing else.

Why the shape is not enough for a clinician

The shape is an exact combinatorial description and it stays. But it answers a question the clinic does not ask. A doctor does not need to know how many separate blocks of coincidence exist. The doctor needs one number: at worst, how many conditions could be hiding behind a single appearance?

Call that number the depth. The depth of a numbering is the largest number of pyramids anywhere on the solid that show the same four values.

Blocks and depth are not the same thing, and they can run in opposite directions. Three masked pairs, shape (3,0,0), contains three blocks but has depth two: the worst coincidence anywhere is two pyramids alike. One masked triple, shape (0,1,0), contains one block but has depth three. Fewer blocks, deeper mask.

A third quantity moves independently of both. Read the distinct-reads column for those same two shapes: three pairs leaves nine distinct reads, one triple leaves ten. The shallower numbering leaves the patient with fewer clear readings than the deeper one, because three pairs each swallow one reading while a single triple swallows two. Blocks, depth and distinct reads are three different numbers, and none can be derived from the others. That is why the fold has to be set out rather than assumed.

Folding the nineteen shapes by depth gives four groups. Depth one is the all-clear shape alone. Depth two gathers the six pair-only shapes. Depth three gathers the seven shapes that contain a triple but no quad. Depth four gathers the five shapes that contain a quad. One plus six plus seven plus five is nineteen, and every numbering lands in exactly one group.

The doctor’s table

depth	what the doctor faces	shapes	numberings	share
1	one clear explanation	1	3,538,104	56.846381 %
2	two conditions could coincide	6	2,644,702	42.492176 %
3	three could coincide	7	41,151	0.661169 %
4	four could coincide	5	17	0.000273 %
	total	**19**	**6,223,974**	**100 %**

Read the second row as follows. In 2,644,702 of the 6,223,974 numberings, somewhere on the solid two pyramids show the same four values and nowhere do three. The patient’s worst ambiguity is two-way. Whether that numbering also contains a second or a third masked pair is a question about its shape, and it is answered by the first table, not this one.

This second table is the one that can be set beside clinical data. The first cannot. The shape distribution happens to open on figures of roughly 57, 33, 8 and 1 per cent, which invites comparison with clinical rules of thumb of a similar look. The resemblance is arithmetical coincidence between two different statistics, and is not evidence of anything.

What the table proves, and what it does not

Three results are proven rather than observed. No reading is ever masked more than fourfold, so depth five does not exist. The twelve pyramids never fall below five distinct reads, a floor reached only by the singularity (4,0,1), which is unique. And exactly one combination, number 4366, admits no fully clear reading at all.

A fourth result is a tendency rather than a bound, and it is the one this paper carries into the clinic. A fourfold mask can stand alone: the shape (0,0,1), a lone quad among eight clear reads, occurs in three numberings. But it usually does not. Of the seventeen numberings that reach depth four, three are lone quads and fourteen arrive with other masking alongside. The deepest confusion comes surrounded by other confusion roughly six times in seven.

That tendency is worth testing, because nothing in the clinical literature has ever asked whether deep diagnostic ambiguity travels with other ambiguity or can sit alone inside an otherwise clear picture.

4. What the literature already holds

Four families of published work bear on the census. None was designed to answer it.

4.1 How often is anything unclear

Diagnostic uncertainty affects at least 20 % of primary care consultations, and is a recognised driver of additional investigation and referral [6]. A systematic critical review of how primary care physicians manage that uncertainty found the empirical base thin and the responses varied, with a majority of general practitioners acknowledging uncertainty within the consultation itself [7]. Physician tolerance of uncertainty is itself measurable and consequential: low tolerance is associated with inappropriate antibiotic prescribing [8].

Medically unexplained symptoms account for a further 10–15 % of consultations, with 3–10 % of adult consulters reporting persistent or recurring unexplained symptoms [9, 10].

The census places 43.15 % of numberings in the not fully clear category. The direction agrees. The observables do not match: 20 % is a floor on recognised uncertainty, not a count of how many readings were compatible.

4.2 Readings that never resolve

In a Finnish primary care register of 503,001 recorded diagnoses, 13.5 % were symptom codes rather than disease codes, roughly one appointment in eight [11]. A Dutch longitudinal cohort followed 12,532 episodes of care that began as a symptom diagnosis: 9.4 % became a disease diagnosis, at a median of 22 days; 85.8 % resolved within a year; 4.3 % persisted as symptom diagnoses [12].

This measures the absence of a reading, not the coincidence of several. It is the clinical counterpart of a face left blank, not of two faces carrying the same four values.

4.3 The instrument ceiling

Physicians typically carry about three working hypotheses in daily practice, and several randomised studies of diagnostic decision support explicitly required participants to record no more than three differential diagnoses [13, 14]. Checklist studies record the number of hypotheses considered but with the same practical compression [15].

This matters more than it first appears. The model’s proven ceiling is four. No instrument in the current literature can see a fourth competing reading, because the instruments stop at three by design. The ceiling is not contradicted; it is invisible.

4.4 Masking proper

The closest structural analogue in medicine is Symptom-Disease Pair Analysis of Diagnostic Error, in which a benign label conceals a dangerous disease, the canonical pair being dizziness and stroke [16]. Across 28 studies covering more than 91,000 patients, the median diagnostic error rate for the dangerous conditions studied was 13.6 %, ranging from 2.2 % for myocardial infarction to 62.1 % for spinal abscess, with a median serious harm rate of 5.5 % [17]. Extended to national incidence, the weighted mean rates were 11.1 % error and 4.4 % serious harm per disease case, giving an estimated 795,000 Americans permanently disabled or dead each year, with roughly half of all serious harm arising from just fifteen diseases [18]. Stroke is misdiagnosed at around 17.5 %, against roughly 1.5 % for myocardial infarction. Missed stroke after a benign dizziness label is measurably less frequent in specialty care than in general practice [19].

This is genuine masking, measured. But it is a pair method. It compares one wrong label against one right one. It cannot report that three readings coincided, and it never has.

5. The diagnostic odyssey: a depth-3 observable

One literature does reach past depth two, and it has not previously been read this way.

People eventually diagnosed with a rare disease are given, on average, two or three misdiagnoses before the correct one [20]. Other summaries put it at three misdiagnoses and five doctors [21]. The EURORDIS Rare Barometer surveyed 6,507 people living with 1,675 rare diseases across 41 countries and found an average total diagnosis time of 4.7 years, with 56 % of respondents diagnosed more than six months after first medical contact [22]. The same programme reports that 60 % were initially misdiagnosed with a different physical disease and 60 % were misdiagnosed with a psychological condition or dismissed, while 25 % required eight or more consultations before confirmation [23]. A survey of 3,471 patients and carers across 436 rare diseases found 46 % had received a misdiagnosis, after an average of four physicians, seven tests and 4.4 years [24]. In a German expert centre series of 78 patients, the correct diagnosis had been reached without the expert in only 21 % of cases; patients had seen an average of six physicians before referral, which itself took four years [25]. A German Delphi survey of rare-disease centres describes the same pattern, with stigmatisation as psychosomatic a typical experience [26].

A patient carrying two or three successive wrong labels before the true one is a masked reading of depth three or four. This is the zebra of the book’s own aphorism, and here it is counted.

The honest limitation

Census masking is simultaneous: several pyramids carry the same four values at once. The odyssey is sequential: labels replace one another over years. These are not the same observable. The defensible bridge is weaker than an identity and still worth stating: a sequential chain of k labels demonstrates that k readings were each compatible with the presentation, which entails that a simultaneous mask of depth at least k existed. The bridge is an inference, not a measurement, and is offered as such.

Two adjacent findings support the same reading. Among 7,838 patients with twelve rarer cancers, 23.4 % had three or more pre-referral consultations, rising above 30 % and reaching 60 % for small intestine, bone sarcoma, liver, gallbladder, cancer of unknown primary, soft-tissue sarcoma and ureteric cancer [27]. Higher morbidity burden is associated with longer diagnostic intervals and more emergency referrals, that is, with a more crowded field of candidate explanations [28].

6. Why the denominators differ, and why the low figures surviv

At first reading the literature appears to contradict the census. The census puts depth-3 masking at 0.66 %. The rare-disease literature puts multiple misdiagnosis at a majority of patients. The contradiction is an artefact of selection, and recognising it is the central methodological result of this survey.

source	conditioned on
rare-disease odyssey	a confirmed rare disease
second-opinion series	referral to a quaternary centre
autopsy series	death, then on autopsy being requested
SPADE	a subsequent adverse event

Every dataset that displays deep masking is drawn from cases already known to have been hard. The census is unselected: it enumerates all 6,223,974 numberings on equal terms.

The correction is estimable. Rare diseases collectively affect on the order of 6 % of people, some 30 million in Europe and 300 million worldwide [22]. If a majority of that group experiences a chain of three or more candidate labels, the unconditional population rate of a depth-3 masking falls to the order of a few per cent. That is the same order of magnitude as the census figure of 0.661169 %, and it is the first quantity in this project that is comparable rather than merely analogous.

Two further bodies of work confirm that unselected comparison is unavailable rather than unfavourable.

Second opinions

Of 286 patients referred from primary care to a quaternary internal medicine division, the referring diagnosis was confirmed in 12 % of cases, better defined or refined in 66 %, and distinctly different from the final diagnosis in 21 % [29]. A clean three-way split, but a measure of agreement between two readers, not of how many diagnoses were compatible with the presentation.

Autopsy

The Goldman criteria classify antemortem against post-mortem diagnosis, Class I being a missed major diagnosis that might have altered therapy or survival. Major Class I and II discordance is reported across the literature between 16.6 % and 59 % [30]; a decade review of one centre's autopsies found 77.5 % concordance, 19.4 % discordance and 3.1 % inconclusive, about one autopsy in five [31]. A systematic review and meta-analysis of adult intensive care autopsy series pools these rates formally [32], and the canonical trend analysis established that the rate has fallen far more slowly than technology alone would predict [33].

Autopsy is nevertheless binary: one label against one truth. It cannot report depth. Even diagnostic accuracy in a well-characterised condition such as Parkinson's disease, estimated at 80.6 %, is expressed as a single accuracy figure rather than a distribution of competing candidates [35].

The newborn intensive care series

One autopsy study deserves separate treatment, because it is the only source located in this survey that speaks to the shape of masking rather than merely to its rate. In a series of 545 consecutive neonatal intensive care deaths, of which 338 were autopsied, discordance was found to track prior clinical uncertainty rather than to occur at random: autopsy was performed most often where certainty had been lowest, and the major discordance rate varied inversely with the degree of diagnostic certainty [34].

Simply put, the study works as follows. In a newborn intensive care unit the clinicians recorded how confident they were about each infant's diagnosis. After death, the autopsy showed what had actually been wrong, and the two were compared. Where the clinicians had been confident, the autopsy usually agreed. Where they had been unsure, the autopsy more often revealed something major that had been missed. The serious misses did not happen quietly; they happened in cases that already looked untidy.

Neonatal denotes the first 28 days after birth. These are newborns, born alive and admitted to intensive care, not stillbirths. The distinction matters, and it is the reason this particular population is unusually well suited to the question. These are patients with almost no history. They have no symptoms they can describe, no prior records, no lifestyle, no accumulated conditions. Diagnosis rests almost entirely on

what the machines and the examination show. The masking observed is therefore not confounded by a patient who forgets, minimises or misdescribes. What the clinicians were uncertain about was the biology itself.

A second detail of the same study points the same way. Cases were sorted into six clinical groups: anomalies, cardiac anomalies, hypoxic ischaemic encephalopathy, prematurity and its complications, infections, and other. Discordance was most frequent and most consequential in prematurity and in other, which are precisely the two least specific categories. The vaguest labels concealed the most. In the model's vocabulary, the readings that were hardest to tell apart were the ones carrying the least distinguishing information.

Three cautions belong with all of this. It is a single study, in newborns, from 1998. It measures certainty against error rather than counting how many diagnoses fitted at once. And the autopsies were themselves requested more often where clinicians were unsure, which is a selection effect in its own right. It hints; it does not prove. Set against those cautions is the narrowness that makes it useful: newborns are a small slice of medicine, and nothing here generalises to adult primary care on its own.

7. The proposal

Nothing found in this survey contradicts the census. Nothing confirms it either. The position at each depth is:

census claim	status
56.85 % clear	not contradicted; the 20 % uncertainty figure is a compatible floor
42.49 % two-deep	a real partner in SPADE, on a different denominator
0.661169 % three-deep	first empirical neighbour, after correcting for selection
0.000273 % four-deep	no counterpart of any kind exists
ceiling of four	untestable with current instruments, which stop at three
lone deep masks rare	untested, and never asked

The last line is the reason for this paper.

The model shows that a fourfold mask rarely occurs among otherwise clear readings: three of the seventeen deepest numberings stand alone, fourteen come with other masking. Translated: a four-way diagnostic confusion is usually, though not always, surrounded by other ambiguity. If that proportion holds clinically, it is directly useful, because it says the deepest and most dangerous confusions announce themselves through the untidiness around them rather than hiding inside apparent clarity. The newborn intensive

care findings above are a first and indirect echo of exactly this: discordance rose where certainty was already low, and concentrated in the least specific diagnostic categories [34].

A physician survey is required

Its categories should be taken from the census rather than invented. For a consecutive series of presentations, each clinician records whether the presentation admitted one clear explanation, a two-way ambiguity, a three-way ambiguity, or a four-way or deeper ambiguity, and separately whether the deepest ambiguity present was accompanied by other ambiguity or stood alone. The first item yields a distribution directly comparable with 56.85 / 42.49 / 0.66 / 0.00027. The second and third ask whether deep ambiguity stands alone or travels with other ambiguity, a question no existing study addresses.

Such a survey would not be a metaphor for the census. It would be a test of it. Section 8 sets it out in full.

The clinical prize is worth stating plainly, and carefully. A survey cannot make a rare disease less rare; prevalence is not a variable a questionnaire can move. What it can change is how often a rare disease is missed. If deep masking is systematically flagged by the ambiguity surrounding it, and if clinicians are taught to read that surrounding untidiness as the signal, then the average of two or three wrong labels and 4.7 years may be shortened. Making rare conditions less rarely found is the achievable goal, and on the evidence assembled here it has never been specifically researched.

8. The survey

A necessary caution, to be read before anything below

The author is a musician, not a physician. It is way beyond his competence to even remotely suggest that such a survey would be complete or best worded, and it would be a grave mistake not to emphasise that it is a mere sketch, a skeleton, a kind of kaleidoscope to take as a basis: a viewpoint from which a proper and professional study may be developed.

Every item below should be rewritten by those qualified to write it. Wording, response categories, recall window, sampling frame, ethical approval and validation are all outside what one person working alone on a polyhedron can settle. What is offered here is the shape of the question, not the finished instrument.

What follows is a first sketch of the instrument, not a description of one. It can be copied and used as a starting point. Two cards: one completed for each consecutive patient, one completed once by the clinician.

One instruction matters more than any other and belongs at the top of the form. Question A1 does not ask how many diagnoses were written down. It asks how many remained genuinely compatible with the presentation. The existing literature caps recorded differentials at about three, often by study design, which is exactly why the fourth reading has never been seen. A form that inherits that cap will inherit its blindness.

Part A · the encounter card, one per consecutive patient

A1 At the close of this encounter, how many distinct conditions remained genuinely compatible with the presentation?

- ☐ One. A single clear explanation
- ☐ Two
- ☐ Three
- ☐ Four or more

A2 If more than one remained, could they be told apart on the information available at the close of the encounter?

- ☐ Yes, they could be ranked with confidence
- ☐ Partly
- ☐ No, they were indistinguishable on the evidence to hand
- ☐ Not applicable, only one remained

A3 Was the rest of the picture otherwise clear?

- ☐ Yes. The ambiguity was isolated in an otherwise crisp presentation
- ☐ No. Other features of the case were also ambiguous
- ☐ Not applicable, only one condition remained

A4 If other features were also ambiguous, how many separate further ambiguities were present, apart from the deepest one?

- ☐ None. The deepest ambiguity stood alone
- ☐ One other
- ☐ Two others
- ☐ Three or more

A5 Was any competing condition concealed by something other than itself?

- ☐ By a diagnosis the patient already carried
- ☐ By current treatment
- ☐ By a benign-looking label already applied to this episode
- ☐ No, nothing concealed anything

A6 Presentation category. Record the presenting complaint, not the final diagnosis.

A7 Optional, at 30 days. What became of the encounter?

- ☐ A single diagnosis was confirmed, as expected at the time
- ☐ The working diagnosis was revised
- ☐ Still unresolved
- ☐ Lost to follow-up

Part B · the clinician card, completed once

B1 Setting.

- ☐ Primary care
- ☐ Emergency department
- ☐ Inpatient
- ☐ Specialist clinic

B2 Years since qualification.

- ☐ Under 5
- ☐ 5 to 14
- ☐ 15 or more

B3 Specialty.

Analysis

A1 alone yields a distribution directly comparable with the census: 56.85, 42.49, 0.66 and 0.00027 per cent. A3 and A4, read against A1, ask whether deep ambiguity stands alone or travels with other ambiguity. The model says standing alone should be uncommon rather than impossible, roughly three cases in seventeen at depth four, and A4 recovers

the shape by counting how many further ambiguities accompany the deepest one. A5 sorts the observed masking into the senses separated in section 2, and is the first instrument to ask for that separation.

The arithmetic of sample size is unforgiving and should be stated before anyone begins. Expected counts at the census rates:

encounters	expected 3-deep	expected 4-deep
1,000	6.6	0.003
3,000	19.8	0.008
10,000	66.1	0.027
1,000,000	6,612	2.7

Depths one, two and three are testable. Depth four is not. At the census rate a four-deep mask appears roughly once in 366,000 encounters, so no survey of feasible size will observe one, and the model's rarest prediction stays out of reach. This should be conceded in advance rather than discovered afterwards.

A practical pilot is thirty clinicians recording forty consecutive encounters each: 1,200 cards, roughly eight expected 3-deep events. That is enough to establish whether the one and two-deep split lands anywhere near 57 and 42, and enough to see whether A3 ever returns the isolated answer at depth three. It is not enough to settle the three-deep rate, which needs several thousand.

The model predicts that the isolated answer at A3 will be uncommon wherever the depth is three or four, and A4 predicts that when ambiguity does accompany it there will usually be more than one further block. If clinical data show the same lean, the analogy holds at the point where it is most specific. If isolated deep ambiguity turns out to be the ordinary case, the analogy fails there, and that too is worth knowing.

Why it is worth doing

The zebra is hiding, and patients are dying because of the masking depths. It is time to employ the data processing powers available to mankind: to screen, to sieve, to save them.

References

1. Normal and Abnormal Complete Blood Count With Differential. StatPearls, NCBI Bookshelf NBK604207, 2024.
2. Hallyburton A. Diagnostic overshadowing: an evolutionary concept analysis on the misattribution of physical symptoms to pre-existing psychological illnesses. Int J Ment Health Nurs 2022. doi 10.1111/inm.13034. PMC9796883.
3. Diagnostic overshadowing in mental health: a mixed-methods systematic review of its impact on health inequities and system-level responses. J Public Health 2026. doi 10.1007/s10389-026-02797-x.
4. Diagnostic overshadowing reviewed and reconsidered. PMID 11531461. Review of twelve studies; the concept originates with Reiss, Levitan and Szyszko, early 1980s.
5. Beta blockers can mask not only hypoglycaemia but also hypotension. PMID 35593361. See also standard product labelling: beta-adrenergic blocking agents may mask signs of acute hypoglycaemia and the tachycardia of thyrotoxicosis.
6. Additional Investigations and Referrals by Primary Care Physicians in Uncertain Clinical Situations: Cross-Sectional Study Using Virtual Patient-Based Scenarios. JMIR Form Res 2026;e71717. PMID 41911436.
7. Managing diagnostic uncertainty in primary care: a systematic critical review. BMC Primary Care 2017. doi 10.1186/s12875-017-0650-0.
8. Coping With Diagnostic Uncertainty in Antibiotic Prescribing: A Latent Class Study of Primary Care Physicians in Hubei, China. PMC8695689.
9. The General Practitioner’s Consultation Approaches to Medically Unexplained Symptoms: A Qualitative Study. PMC4041244.
10. Medically unexplained symptoms and symptom disorders in primary care: prognosis-based recognition and classification. PMC5297117.
11. Symptomatic diagnoses in primary care: an observational cohort study. BJGP Open 2024;8(4):BJGPO.2023.0234. PMC11687245.
12. The course and management of symptom diagnoses in general practice. Scand J Prim Health Care 2026. doi 10.1080/02813432.2026.2678543.
13. Effects of a Differential Diagnosis List of Artificial Intelligence on Differential Diagnoses by Physicians. Int J Environ Res Public Health 2021;18:5562. PMC8196999.
14. Efficacy of Artificial-Intelligence-Driven Differential-Diagnosis List on the Diagnostic Accuracy of Physicians: An Open-Label Randomized Controlled Study. PMC7924871.
15. Kämmer JE, et al. Differential diagnosis checklists reduce diagnostic error differentially: a randomised experiment. Med Educ 2021. doi 10.1111/medu.14596.
16. Liberman AL, Newman-Toker DE. Symptom-Disease Pair Analysis of Diagnostic Error (SPADE): a conceptual framework and methodological approach for unearthing misdiagnosis-related harms using big data. BMJ Qual Saf 2018;27(7):557–566. PMC6049698.
17. Newman-Toker DE, Wang Z, Zhu Y, et al. Rate of diagnostic errors and serious misdiagnosis-related harms for major vascular events, infections, and cancers: toward a national incidence estimate using the Big Three. Diagnosis (Berl) 2021;8(1):67–84. doi 10.1515/dx-2019-0104.
18. Newman-Toker DE, Nassery N, Schaffer AC, et al. Burden of serious harms from diagnostic error in the USA. BMJ Qual Saf 2024;33(2):109–120. doi 10.1136/bmjqs-2021-014130. PMC10792094.
19. Stroke hospitalization after misdiagnosis of benign dizziness is lower in specialty care than general practice: a population-based cohort analysis of missed stroke using SPADE methods. PMID 34147048.
20. Educational needs in diagnosing rare diseases: a multinational, multispecialty clinician survey. PMC11613602, 2023.
21. Diagnosing rare diseases and mental well-being: a family’s story. PMC9990187.

22. Time to diagnosis and determinants of diagnostic delays of people living with a rare disease: results of a Rare Barometer retrospective patient survey. PMC11369105, 2024.
23. EURORDIS-Rare Diseases Europe. Major survey reveals lengthy diagnostic delays for rare disease patients. 2024.
24. Rare disease diagnostic odyssey survey, 3,471 patients and carers across 436 rare diseases. Rare Patient Voice case study, 2023.
25. Rare diseases: why is a rapid referral to an expert center so important? BMC Health Serv Res 2023. doi 10.1186/s12913-023-09886-7. PMC10463573.
26. Diagnostic needs for rare diseases and shared prediagnostic phenomena: results of a German-wide expert Delphi survey. PMC5325301.
27. Pre-referral GP consultations in patients subsequently diagnosed with rarer cancers: a study of patient-reported data. PMC4758496.
28. Morbidity and measures of the diagnostic process in primary care for patients subsequently diagnosed with cancer. PMC9295610.
29. Van Such M, Lohr R, Beckman T, Naessens JM. Extent of diagnostic agreement among medical referrals. J Eval Clin Pract 2017. doi 10.1111/jep.12747.
30. Usefulness of Pathological Autopsy for Patients With Malignant Tumors. Cureus 2024. PMC11867707.
31. Enhancing the Quality and Delivery of Healthcare: A Decade Review of Autopsy Data. AHFE Open Access, ISBN 978-1-958651-27-8, chapter 12. Ten-year series, 2002–2011.
32. Clinical vs. autopsy diagnostic discrepancies in the intensive care unit: a systematic review and meta-analysis of autopsy series. Intensive Care Med 2024. doi 10.1007/s00134-024-07641-y. PMID 39287650.
33. Shojania KG, Burton EC, McDonald KM, Goldman L. Changes in rates of autopsy-detected diagnostic errors over time. JAMA 2003;289:2849–2856.
34. Dhar V, Perlman M, Vilela MI, Haque KN, Kirpalani H, Cutz E. Autopsy in a neonatal intensive care unit: utilization patterns and associations of clinicopathologic discordances. J Pediatr 1998;132(1):75–79. doi 10.1016/S0022-3476(98)70488-3. PMID 9470004. Division of Neonatology, Department of Paediatrics, Hospital for Sick Children, Toronto. Full text was not available to us; the operational definitions of the three certainty levels are cited from the abstract and remain unverified.
35. Should Physicians Be More Open About Clinical Uncertainty? Medscape, 2026, reporting diagnostic accuracy in Parkinson’s disease at 80.6 %.
36. Grego (Földvári G). STAR FACE. Grego’s Polyhedral Sieve Model for Medical Diagnostics. 2026. DOI 10.5281/zenodo.22259868. grego.hu/star.