

When You Hear Hoof Beats: Four Principles for Separating Zebras From Horses

C. Scott Smith, MD, and Douglas S. Paauw, MD

Background: When a patient comes to the clinic with a new complaint, the often wide array of possible causes creates uncertainty about the optimal evaluation and treatment. Selecting an approach to evaluation involves values that range from ruling out all disease processes (all zebras) regardless of cost to limiting cost by looking only for those processes that are likely (assuming all hoof beats are created by horses). Neither extreme is an optimal approach. We do not want to spend money on unnecessary tests, but we also do not want to miss a rare but potentially serious and treatable disease.

Methods: We offer four principles and their accompanying corollaries that make it possible to separate more easily hoof beats for horses from those for zebras.

Results and Conclusions: By applying these principles and the accompanying corollaries, a physician can more efficiently determine the most efficient and cost-effective approach to taking care of patients. (J Am Board Fam Pract 2000;13:424–9.)

In the frenzy of a busy clinic, it can be difficult to know what to do when a rare disease comes to mind during patient care. As an example, consider the following patient.

Mr. Jones, a 34-year-old new patient, complains of a recent history of palpitations, choking sensations, and nervousness. His blood pressure is 176/103 mm Hg.

Mr. Jones could have a common problem, such as anxiety or essential hypertension; a less common problem, such as hyperthyroidism or recreational drug use; or a rare problem, such as pheochromocytoma or renal vascular disease. Simply ordering a test and moving on is, at best, often a waste of resources and can lead to confusion, further testing, and iatrogenic complications.

Methods

In this article, we offer four principles, and their accompanying corollaries, which make the job of separating hoof beats of horses from those of zebras as efficient and cost-effective as possible.

Submitted, revised, 20 March 2000.

From the Northwest Regional Faculty Development Center (CSS), Veterans Administration Medical Center, Boise, Idaho; and the Division of General Internal Medicine (CSS, DSP) and Department of Medical Education (CSS), University of Washington, Seattle. Address reprint requests to C. Scott Smith, MD, Department of Medicine (111), VA Medical Center, 500 W Fort St, Boise, ID 83702.

Principle 1. Common Diseases Occur Commonly and Rare Diseases Occur Rarely

Corollary: Recognize distinguishing features (those that truly make the likelihood of a rare disease considerably higher or lower) and be cautious with trigger features (those that make us think of a rare disease but do not seriously change likelihood).

What is the first thing that comes to mind when you hear a patient mention night sweats? Many physicians think of tuberculosis or lymphoma. There are, however, far more likely causes, such as menopause, medication side effects, or simple infections. Night sweats is an example of a trigger feature—a feature that make physicians think of a rare disease but does not truly change the likelihood of the rare disease. Trigger features are not bad if used correctly. They can lead a physician to ask additional, more specific questions and focus the physical examination toward possible rare diseases. Does the patient have a history of cough and tuberculosis exposure? Does the patient have adenopathy or bone pain when drinking alcohol?

Trigger features lead physicians astray when, if they consider only the rare disease, these features obscure a more likely cause (upper respiratory tract infection), or when a physician proceeds to diagnostic evaluations based only on a trigger feature (eg, abdominal computed tomographic [CT] scan for lymphoma). Contrast the clinical signs of this

Table 1. Examples of Several Discriminating Features of Rare Disease.

Features	Disease
<i>History</i>	
Unexplained weight loss with labile hypertension (if screening test for alcoholism is negative)	Pheochromocytoma
Patient <30 years (especially female) with hypertension	Renovascular disease caused by fibromuscular dysplasia
Bone pain after consumption of alcohol	Tumor (especially Hodgkin), osteomyelitis
Platypnea (dyspnea worse when sitting)	Pulmonary shunting (Osler-Weber-Rendu, severe liver disease), patent foramen ovale
Flushing and diarrhea (not medication or alcohol related)	Carcinoid, medullary thyroid carcinoma, mastocytosis
<i>Physical finding</i>	
Orthostatic vital signs in a patient with hypertension	Pheochromocytoma
Saddle nose deformity	Wegener granulomatosis, cocaine abuse, relapsing polychondritis
Inflammation of the external ear sparing the lobe	Relapsing polychondritis
Blue sclera	Osteogenesis imperfecta, pseudoxanthoma elasticum
Livedo reticularis (not caused by cold exposure)	Systemic lupus erythematosus, polyarteritis nodosum, atheromatous emboli syndrome
<i>Laboratory finding</i>	
Unprovoked hypokalemia in a patient with hypertension	Primary aldosteronism, renovascular disease
Dipstick positive for blood but no red blood cells on urinalysis	Rhabdomyolysis, hemolytic anemia

possibility with some clinical features that Mr. Jones might have.

Further evaluation of Mr. Jones shows findings from his physical examination and family history to be unremarkable. He is taking no medications. His screening laboratory studies are notable for a potassium level of 3.3 mEq/L.

Renovascular disease and primary hyperaldosteronism are unusual causes of hypertension in the general population (2% or less in most studies).¹ Spontaneous hypokalemia is rare, however, in uncomplicated hypertension.² This clinical picture increases the likelihood of renovascular disease to 10% to 20%³ and the likelihood of primary aldosteronism to nearly 50%.⁴ At these rates, it is important to screen for these reversible and serious causes of hypertension. Unprovoked hypokalemia in a patient with hypertension is a good discriminating feature because it is rare in the usual causes of hypertension, whereas it is relatively common in the unusual causes. It changes the pretest likelihood of these diseases by a factor of 10 or more.

In deciding whether a clinical feature is a trigger feature or a distinguishing feature, ask yourself three questions (Table 1 lists several distinguishing features).

1. How reliable is the feature?
2. How often is it encountered in the rare disease?
3. How often is it encountered in more common diseases?

Principle 2. The Mind Is an Imperfect Likelihood Estimator

Corollary: Beware of a story “too good to be true” and recent experience. These trigger features are particularly compelling, leading physicians to estimate incorrectly the likelihood of a rare disease.

Clinical reasoning involves the iterative generation and refinement of hypotheses until a working diagnosis is reached. This working diagnosis should be coherent, adequate to describe all the clinical features, and parsimonious.⁵ Two processes dominate the search for a working diagnosis, induction and heuristics. Induction is a generalized rule arrived at by repeated observations. “The sun will come up tomorrow because it has come up every other day that I can remember.” Heuristics are mental shortcuts or rules of thumb used when there are not enough observations to use induction.⁶ One such heuristic is Sutton’s Law, named after a fa-

mous bank robber who explained that he robbed banks because “that’s where the money is.” Because heuristics occur at a preconscious level, they are subject to unrecognized bias.

Mr. Jones relates to you that his hands, left more than right, get tingly with these episodes.

When given this history, one learner ordered a chest CT scan to look for aortic coarctation, because asymmetric perfusion and hypertension are classic features of that disease. This approach is an example of a *representativeness* heuristic bias—overestimation of likelihood because of resemblance to a prototype, while ignoring the true pretest likelihood. Another learner ordered an electrocardiogram (ECG) to look for atrial fibrillation. This physician had just seen a patient with atrial fibrillation who had an embolus to the left arm and who had symptoms a lot like those of Mr. Jones. This approach is an example of the *availability* heuristic bias—where likelihood is judged by the ease with which something is remembered. Several things that do not change disease likelihood can affect the ease with which something is remembered (eg, amount of exposure in the curriculum, a striking recent experience, or a lack of familiarity).

When a cue leads to consideration of a rare disease, the physician can reduce the possibility of heuristic bias by asking three questions.^{7,8} If the answer to any of these questions is yes, then consulting someone with greater experience would be helpful.

1. Is the story too good to be true (and did I lead the witness)?
2. Has a recent experience affected my judgment?
3. Is my experience with this entity too limited?

Principle 3. The Unusual Presentation of a Common Disease Is Generally More Likely Than the Usual Presentation of an Uncommon Disease

Corollary 1: *Screening everyone for a rare disease is not useful.*

Corollary 2: *When faced with an unusual clinical feature, ask first whether it can be explained by something other than a rare disease.*

Principle 3, stated another way, says that screening everyone for a rare disease is likely to result in more new versions of normal (false-positive results) than rare diseases (true-positive results). Even if we

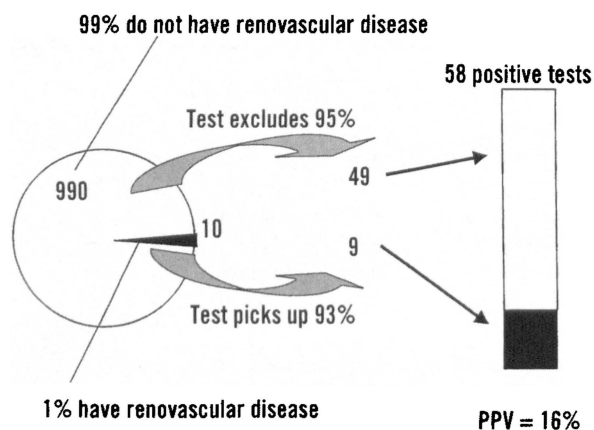


Figure 1. The effect of screening 1,000 persons for renovascular disease (likelihood 1%) with a captopril-stimulated isotope renogram. This test has a sensitivity of 93% and a specificity of 95%.⁹ The black bar represents true-positive results. The white bar represents false-positive results. The test has a positive predictive value (the probability that a positive test represents true disease) of only 16% in this population.

find every instance of a zebra, we will have very few examples because the disease, by definition, is rare. At least a few normal patients will have an abnormal screening result even with the best of tests, and this absolute number will be larger than the number of zebras. This argument is described graphically in Figure 1.⁹ If it is possible to increase the prescreening likelihood of the rare disease even several percentage points by using discriminating features (see principle 1) then, as shown in Figure 2, screening is much more effective. Now consider the following twist in Mr. Jones’ case.

Mr. Jones is given nifedipine without additional evaluation. He returns 2 months later complaining that his choking and nervous spells are getting worse, and that he now has flushing. He denies diarrhea. In addition, he is currently taking some over-the-counter medications for an upper respiratory tract infection. Mr. Jones’ medications are extended-release nifedipine 60 mg/d, acetaminophen 650 mg twice a day, and sustained-release pseudoephedrine-guaifenesin twice a day. Because of the new complaint of flushing, a 24-hour urine 5-hydroxyindoleacetic acid (5-HIAA) level is obtained. It is slightly elevated at 12 mg/24 h. The most likely explanation is . . . ?

This scenario is a good example of one in which corollary 2 above should be applied—that one

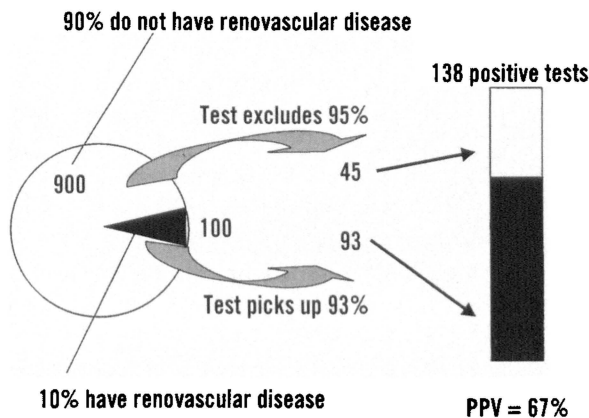


Figure 2. When discriminating features are used to increase the pretest likelihood of renovascular disease to 10%, screening 1,000 people becomes much more effective. The captopril-stimulated isotope renogram now has a positive predictive value of 67% in this population.

should search for a simple answer to unusual clinical findings before attributing them to a rare disease. This case has two unusual clinical features, facial flushing and an elevated urinary 5-HIAA level. These findings would lead one to wonder about carcinoid as the cause of the patient's signs and symptoms. Both features, however, are more likely to be side effects of a drug.

Although most patients with carcinoid have flushing, most patients with flushing do not have carcinoid. The most common causes of flushing include menopause, medications, food additives, neurophysiologic responses (blushing and anxiety), alcohol, and acne rosacea.¹⁰⁻¹² The patient's nifedipine is well known to cause flushing.^{11,13}

The 24-hour urine test was ordered to look for carcinoid because of the flushing (see what kind of trouble you can get into when you chase a trigger feature). Unfortunately the acetaminophen and guaifenesin were not discontinued, and both are known to cause a falsely elevated 5-HIAA reading.^{14,15} In fact, false-positive elevations of 5-HIAA levels, with values in this range, are common.¹⁶

The carcinoid syndrome is very rare, with a crude incidence of less than 1/100,000 population.¹⁷⁻¹⁹ When it does occur, solitary flushing is rarely the only clinical finding.^{20,21} Flushing and diarrhea together make carcinoid more likely. Adding a rare feature of carcinoid, such as asthma, pellagra, or telangiectasia, to flushing and diarrhea makes carcinoid much more likely. This patient has

none of these additional features. Most patients with carcinoid have 5-HIAA levels of greater than 150 $\mu\text{mol}/24\text{ h}$, not the borderline elevation seen here.¹³ Given all these arguments, the flushing and elevated 5-HIAA levels are most likely due to effects of medication. A safe and cheap way to test this hypothesis would be to discontinue the nifedipine for a brief period to see what happens to the flushing before proceeding to an expensive and possibly harmful evaluation.

Principle 4. Not Everything We Are Taught Is Correct

Corollary 1: Whenever possible, we should seek evidence-based approaches to clinical problems. This advice is especially important for commonly encountered diseases or where the stakes are high.

The word of a professor or senior resident is a powerful trigger feature, and unfortunately, medical myths abound.²²⁻²⁴ These myths often develop as plausible (but unproved) theory and are then spread by the informal system of communication. Superiors hold great power in the eyes of trainees because they control evaluation. Trainees might not be aware of the economic, political, and historical dimensions that influence what they are being told. These dimensions include the tacit punishment and reward systems for teaching and research,^{25,26} the hidden curriculum that determines the learning environment (eg, policies, evaluation, and resource allocation),²⁷ and the effect of research funding.²⁸ If Mr. Jones were 20 years older, consider the following nuance in his case.

Mr. Jones relates that he had a myocardial infarction 2 months ago. He also has moderate chronic obstructive pulmonary disease (COPD) (forced expiratory volume [in 1 second] vital capacity ratio 60% of predicted), and modest improvement with bronchodilators. You would like to give him a β -blocker for its cardioprotective effects but worry about exacerbating his COPD.

β -Blockers provide considerable improvements in reinfarction and mortality rates for at least 2 years after a myocardial infarction.²⁹ In one trial, β -blockers were given to 50 patients with COPD, most of whom had a bronchodilator response on pulmonary function tests. These patients had substantial cardiovascular effects, but no patient had dyspnea or wheezing, and there was no noteworthy group effect on pulmonary function.³⁰ Other stud-

Table 2. Questions for Applying the Four Principles in Practice.

1. What is my estimate of the probability of the rare disease in this patient population and how will testing change this estimate?
2. Could I have a biased estimate?
 - Was I taught this as a clinical pearl without support in the literature?
 - Has a recent experience affected my judgment?
 - Could there be an unfamiliar cause (especially a drug side effect)?
 - Am I overvaluing a great story (and did I lead the patient)?
3. Is the feature that led to the consideration of a rare disease is a distinguishing feature or a trigger feature?
 - How reliable is the feature?
 - How often is it seen in rare disease?
 - How often is it seen in common diseases?

ies have shown similar results for selected β -blockers.³¹⁻³³ Despite these facts, symptomatic COPD and a bronchodilator response are still thought to be contraindications to β -blocker therapy at many institutions. A healthy skepticism can be a great asset in a trainee and can encourage faculty to seek evidence for their statements.

Conclusions

It is difficult to separate the rare zebra from a large herd of horses in a busy clinic. Time pressures tend to force physician to order laboratory tests, imaging procedures, and other diagnostic technology quickly whenever the question of a rare disease comes up. The questions in Table 2 apply to the four principles listed above and can help physicians use resources efficiently and cost-effectively.

References

1. Kaplan NM, Lieberman E, Neal WW. Clinical hypertension. 6th ed. Baltimore: Williams & Wilkins, 1994.
2. Double-blind control study of antihypertensive agents. Veterans Administration Cooperative Study on Antihypertensive Agents. *Arch Intern Med* 1962; 110:230-6.
3. Simon N, Franklin SS, Bleifer KH, Maxwell MH. Clinical characteristics of renovascular hypertension. *JAMA* 1972;220:1209-18.
4. Fishman LM, Kuchel O, Liddle GW, Michelakis AM, Gordon RD, Chick WT. Incidence of primary aldosteronism uncomplicated "essential" hypertension. A prospective study elevated aldosterone secretion and suppressed plasma renin activity used as diagnostic criteria. *JAMA* 1968;205:497-502.

5. Kassirer JP, Kopelman RI. Learning clinical reasoning. Baltimore: Williams & Wilkins, 1991.
6. McDonald CJ. Medical heuristics: the silent adjudicators of clinical practice. *Ann Intern Med* 1996; 124(1 Pt 1):56-62.
7. Tversky A, Kahneman D. Judgment under uncertainty. Heuristics and biases. *Science* 1974;185: 1124-31.
8. Travis CB, Phillippi RH, Tonn BE. Judgment heuristics and medical decisions. *Patient Educ Counsel* 1989;13:211-20.
9. Mann SJ, Pickering TG. Detection of renovascular hypertension. State of the art: 1992. *Ann Intern Med* 1992;117:845-53.
10. Wilkin JK. Flushing reactions: consequences and mechanisms. *Ann Intern Med* 1981;95:468-76.
11. Ray D, Williams G. Pathophysiological causes and clinical significance of flushing. *Br J Hosp Med* 1993; 50:594-8.
12. Mohyi D, Tabassi K, Simon J. Differential diagnosis of hot flashes. *Maturitas* 1997;27:203-14.
13. Conlin PR, Williams GH. Use of calcium channel blockers in hypertension. *Adv Intern Med* 1998;43: 533-62.
14. Maton PN. The carcinoid syndrome. *JAMA* 1988; 260:1602-5.
15. Young DS, Pestaner LC, Gibberman V. Effects of drugs on clinical laboratory tests. *Clin Chem* 1975; 21:398D.
16. Nuttall KL, Pingree SS. The incidence of elevations in urine 5-hydroxyindoleacetic acid. *Ann Clin Lab Sci* 1998;28:167-74.
17. Buchanan KD, Johnston CF, O'Hare MM, et al. Neuroendocrine tumors. A European view. *Am J Med* 1986;81:14-22.
18. Norheim I, Oberg K, Theodorsson-Norheim E, et al. Malignant carcinoid tumors. An analysis of 103 patients with regard to tumor localization, hormone production, and survival. *Ann Surg* 1987;206:115-25.
19. Woods HF, Bax ND, Ainsworth I. Abdominal carcinoid tumours in Sheffield. *Digestion* 1990; 45(Suppl 1):17-22.
20. Thorson A. Studies on carcinoid disease. *Acta Med Scand* 1958;334(Suppl):S7.
21. Kahler HJ, Heilmeyer L. Klinik und Pathophysiologie des Karzinoids und Karzinoid syndromes unter besonderer Berücksichtigung der Pharmakologie des 5-Hydroxytryptamins. *Ergeb Inn Med Kinderheilk* 1961;16:292.
22. Paauw DS. Did we learn evidence-based medicine in medical school? Some common medical mythology. *J Am Board Fam Pract* 1999;12:143-9.
23. Toolis F. Malodorous wounds. *Lancet* 1996;348: 1737.
24. Wall PD. The generation of yet another myth on the use of narcotics. *Pain* 1997;73:121-2.

25. Davidoff F. Why is teaching valued less than research? *ACP Observer* 1996;8. (Au, please add issue information, ie, Month or No.)
26. Carmel S, Glick SM. Compassionate-empathic physicians: personality traits and social-organizational factors that enhance or inhibit this behavior. *Soc Sci Med* 1996;43:1253-61.
27. Hafferty FW. Beyond curriculum reform: confronting medicine's hidden curriculum. *Acad Med* 1998;73:403-7.
28. Stelfox HT, Chua G, O'Rourke K, Detsky AS. Conflict of interest in the debate over calcium-channel antagonists. *N Engl J Med* 1998;338:101-6.
29. Hennekens CH, Albert CM, Godfried SL, Gaziano JM, Buring JE. Adjunctive drug therapy of acute myocardial infarction—evidence from clinical trials. *N Engl J Med* 1996;335:1660-7.
30. Gold MR, Dee GW, Cocca-Spofford D, Thompson BT. Esmolol and ventilatory function in cardiac patients with COPD. *Chest* 1991;100:1215-8.
31. van Zyl AI, Jennings AA, Bateman ED, Opie LH. Comparison of respiratory effects of two cardioselective beta-blockers, celiprolol and atenolol, in asthmatics with mild to moderate hypertension. *Chest* 1989;95:209-13.
32. Paterson BF, Grammer LC, Martin GJ, Patterson R. The management of coexisting asthma and cardiac disease. *N Engl Reg Allergy Proc* 1987;8:309-15.
33. Fogari R, Zoppi A, Tettamanti F, Poletti L, Rizzardi G, Fiocchi G. Comparative effects of celiprolol, propranolol, oxprenolol, and atenolol on respiratory function in hypertensive patients with chronic obstructive lung disease. *Cardiovasc Drugs Ther* 1990;4:1145-9.