Naming the Problem That Underpins "Rule-out Sepsis"

The Need for Bayesian Thinking

Joseph Schulman, MD, MS
Director, NIQU Quality Measurement and Improvement
California Children's Services
Department of Health Care Services
Sacramento, CA
Joseph.Schulman@dhcs.ca.gov

Many providers appear to consider "Rule-out Sepsis" as a simple categorical matter

- Yes, infection/No – end of investigation.
- If the culture does not grow a pathogen, providers may consider some array of clinical signs and study results nonetheless to indicate "Yes" ("Culture-negative sepsis") – with little consideration of alternative explanations.
- We generally don't accept such an approach to diagnostic reasoning for other pathological entities.
- It is crucial to objectively – and when possible, quantitatively – evaluate alternative possible explanations for a particular array of clinical signs and study results.
    - Today, we will examine what we mean by evaluating possible explanations objectively and quantitatively.

Differential diagnosis underpins reliably accurate diagnostic assignment

- Providers may feel that once they decide to initiate antibiotics for a symptomatic baby, they and the baby are “covered.”
  - Such confidence may be warranted only when bacterial infection is objectively the most likely explanation.
  - Absent confirmatory culture results, providers may not actually determine “the ***most likely*** explanation” from systematic consideration of alternative explanations.
    - “Most likely” should amount to a comprehensive and quantitative assessment.
  - Other explanations for the clinical presentation may spontaneously resolve without medical intervention, but perhaps sub-optimally.

Clinical/Lab/Imaging Information From Previous Vignettes

- Maternal temperature 103 F shortly before delivery
- Difficulty with first oral feed
  - ?Aspiration?
- Increasing respiratory distress at about 4 hours after birth
- CXR with areas of consolidation
- Blood culture negative, or organism of unclear pathological role

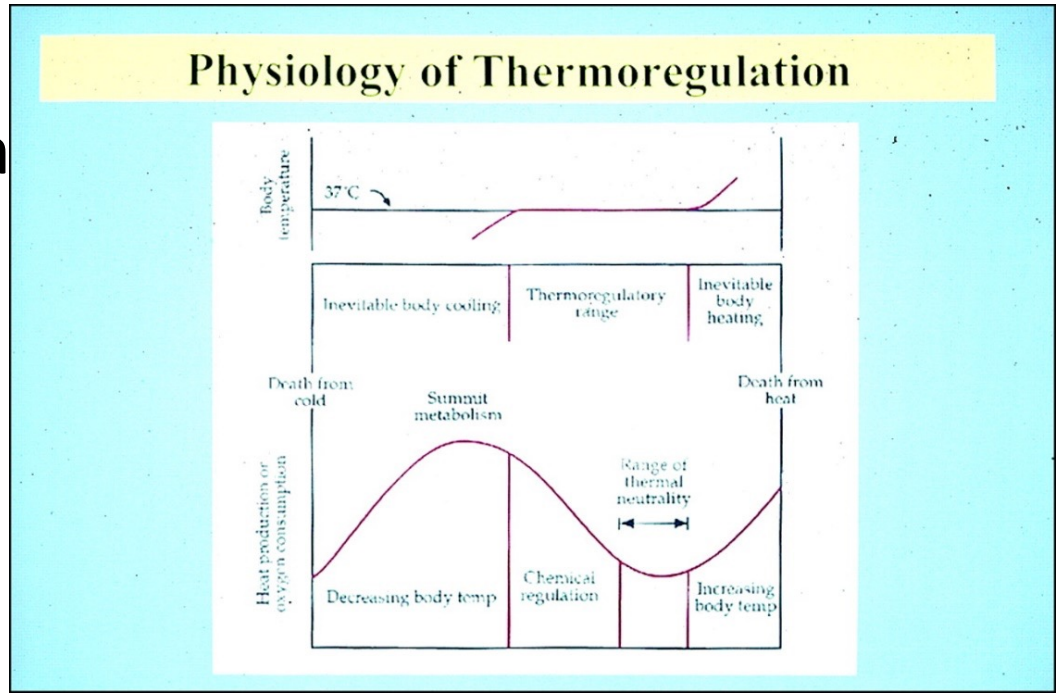
For each information element just presented, what explanation comes to mind as most likely?

How many alternatives explicitly come to mind?

Here are just a few possibilities

Not listed in rank order (varies with the individual baby's particulars

- Thermal stress
  - Environmental
  - Maternal temp – either low, or elevated – effect on neonatal metabolic rate vs nutritional supply
- Retained fetal lung fluid
- Delayed perinatal transition
  - Circulatory
  - Unequal distribution of ventilation
- Hypoglycemia
- Aspiration
- Bacterial infection
- Viral infection



Here are just a few possibilities

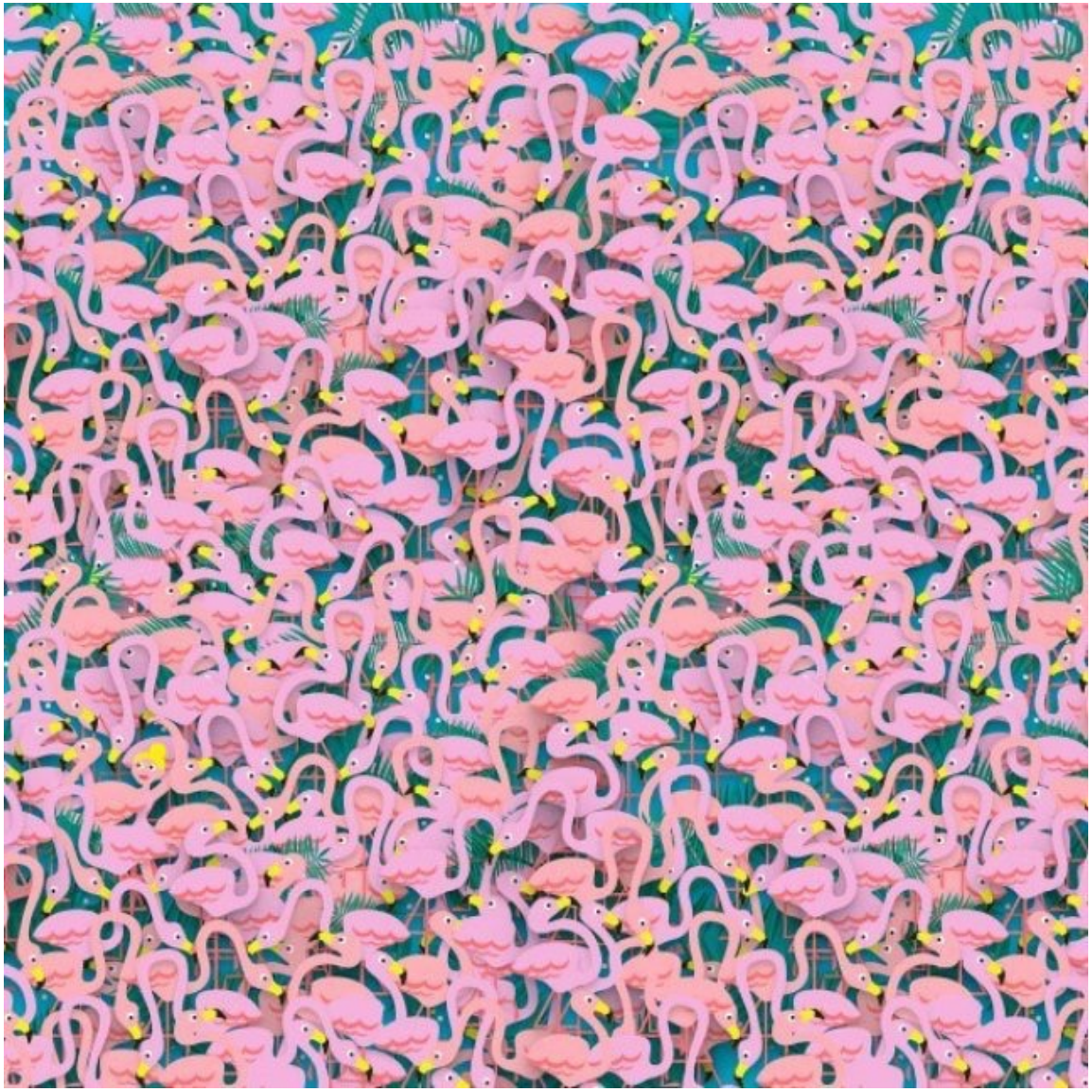
Not listed in rank order, as this varies with the individual baby's particulars

- Thermal stress
  - Environmental
  - Maternal temp – either low, or elevated – effect on neonatal metabolic rate vs nutritional supply
- Retained fetal lung fluid
- Delayed perinatal transition
  - Circulatory
- Hypoglycemia
- Aspiration
- Bacterial infection
- Viral infection

If Aspiration, or Pneumonia, what evidence is there these can resolve clinically and radiographically in 2-3 days?

Chemical pneumonia (especially meconium aspiration) typically lasts for weeks. The inflammatory process of bacterial or viral pneumonia plausibly does too (remains radiographically evident), but these questions have not been rigorously studied.

Too often, we only see what we look for



It’s hard to see the ballerina in this picture if you’re used to only looking for flamingos.

AVERY'S DISEASES OF THE NEWBORN

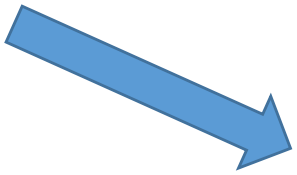
EIGHTH EDITION

H. William Taeusch, M.D.
Professor and Vice Chair of Pediatrics
University of California, San Francisco
San Francisco General Hospital
San Francisco, California

Roberta A. Ballard, M.D.
Professor of Pediatrics and Obstetrics and Gynecology
University of Pennsylvania School of Medicine
The Children's Hospital of Philadelphia
Hospital of the University of Pennsylvania
Philadelphia, Pennsylvania

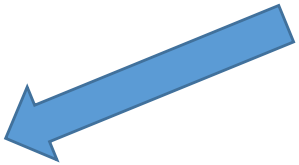
Christine A. Gleason, M.D.
W. Alan Hodson Professor of Pediatrics
Head, Division of Neonatology
University of Washington
Children's Hospital and Regional Medical Center
Seattle, Washington

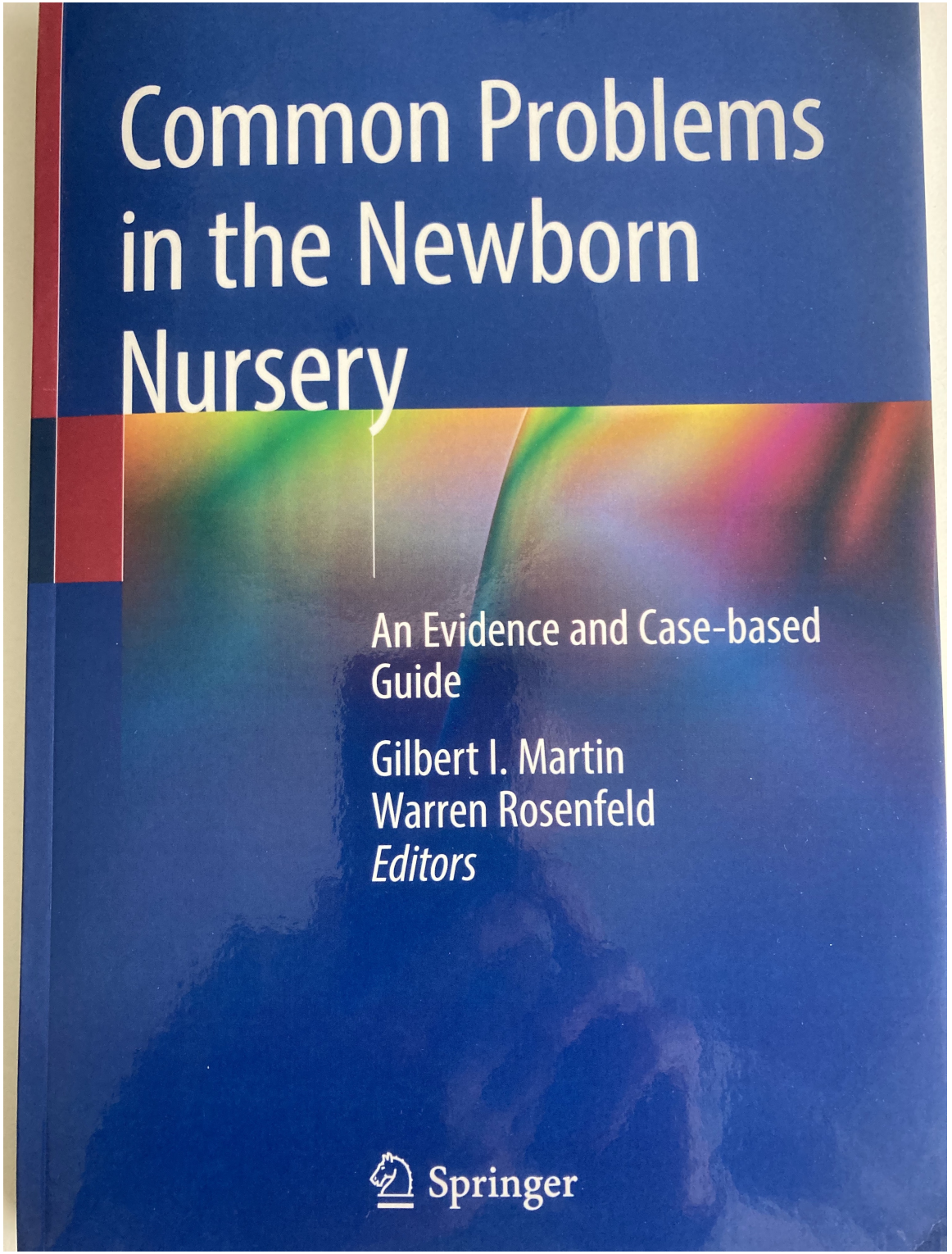
Where is "Aspiration"?



B

Where is a general discussion of "Pneumonia"?




Common Problems in the Newborn Nursery
An Evidence and Case-based Guide
Gilbert I. Martin
Warren Rosenfeld
Editors
Springer

2019

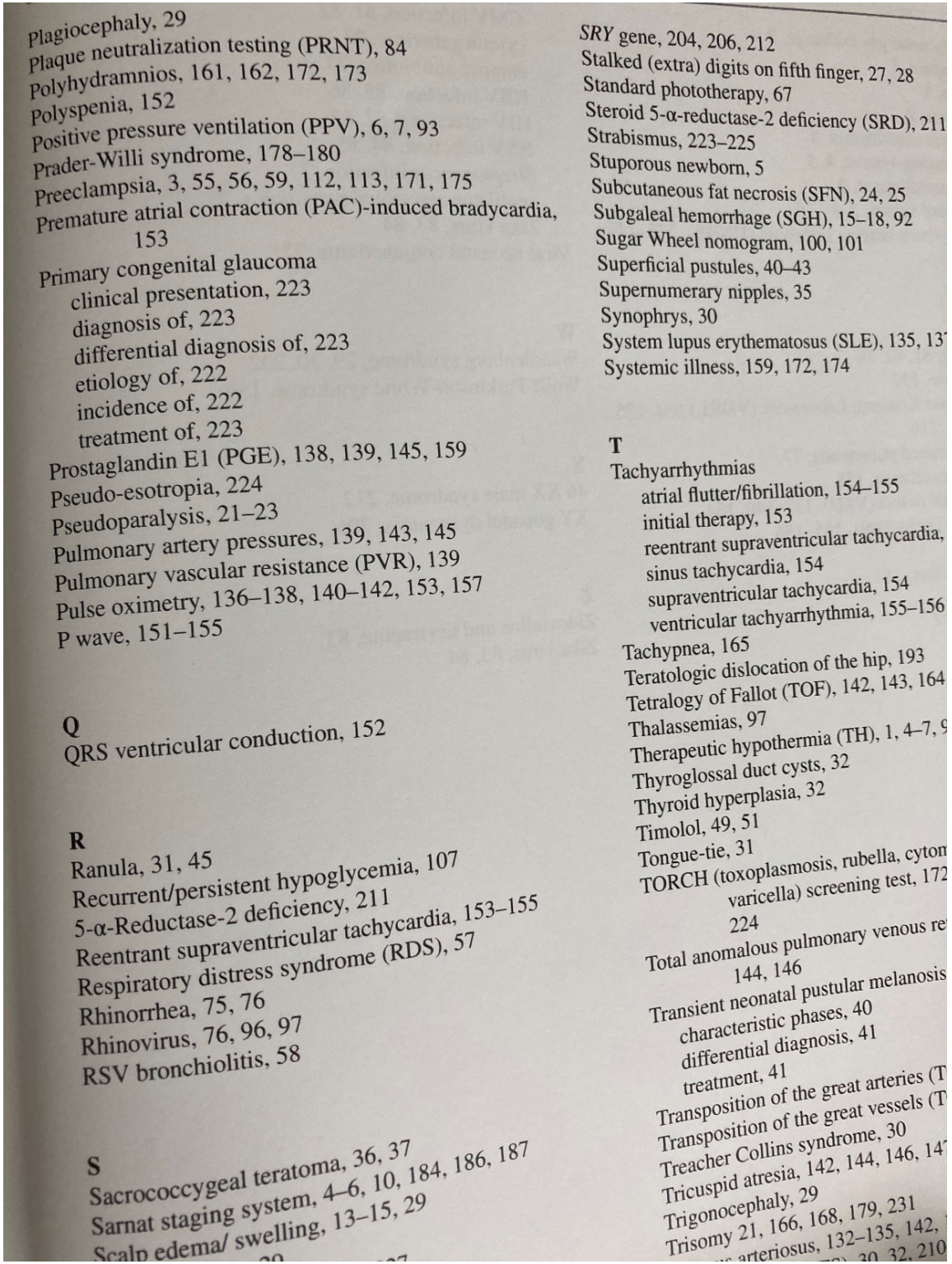
Contents

Is this really a "common problem in the newborn nursery?"

Or, do we just commonly think of it?

Aspiration is not listed

Pneumonia?



Where are Transient tachypnea, Retained fetal lung fluid?

Are We Locked Into *Unrepresentative* Categories for Thinking?

TABLE 1 Distribution of EOS and LOS Rates, Percentage of All Live Births Who Received a Newborn Antibiotic Exposure and Sepsis Diagnostic Efficiency

	Hospital-Level Mean (SD)	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90the Percentile	Lowest	Highest	Statewide
Percentage of births exposed to antibiotics	8.53 (6.27)	3.67	4.69	7.35	9.55	14.14	1.59	42.54	8.43
Diagnostic efficiency, EOS + LOS	66.35 (91.70)	16.54	26.06	41.25	69.50	122.00	7.25	781.00	34.26
EOS									
Rate (cases per 1000 live births)	0.72 (0.69)	0	0	0.53	1.17	1.70	0	2.89	0.75
Diagnostic efficiency	95.08 (71.14)	33.44	46.87	69.52	122.84	178.54	11.45	335.75	88.82
LOS									
Rate (% of admissions with high illness acuity)	3.18 (3.10)	0	0	2.99	4.69	7.25	0	18.75	3.67
Diagnostic efficiency	19.60 (24.02)	3.88	7.09	12.18	22.36	36.96	2.02	164.01	10.35

Few of us are guided by an objective evidence base derived from our own experience.

Schulman J, Benitz WE, Profit J, et al. Newborn Antibiotic Exposures and Association With Proven Bloodstream Infection. Pediatrics. 2019;144(5):e20191105

Basics of Medical Bayesian Logic

One can't interpret a test result without considering pre-test probability.

- Most tests are imperfect; they do nothing more than adjust probability – which may or may not "rule in" or "rule out" the disease.
    - Depends on the situation: risk of not treating when you should have; risk of treating when you shouldn't have.

How often do we actually consider an explicit pre-test probability estimate at the bedside?

- We tend to charge ahead ordering tests without explicitly considering what the new information may be reasonably expected to contribute.

Likelihood Ratio

- LR tells you how likely it is a patient has a disease or condition.
- The higher the ratio, the more likely a patient has the disease or condition.
- A low ratio means that they very likely do *not*.

$$\text{Likelihood Ratio} = \frac{\text{probability a person with the condition has a certain test result}}{\text{probability a person without the condition has a certain test result}}$$

- **Positive LR**: Tells you how much to increase the probability of having a disease, given a positive test result.
- **Negative LR**: This tells you how much to decrease the probability of having a disease, given a negative test result.

T+ Adjusts probability
upward **LR(+)**
a number > 1

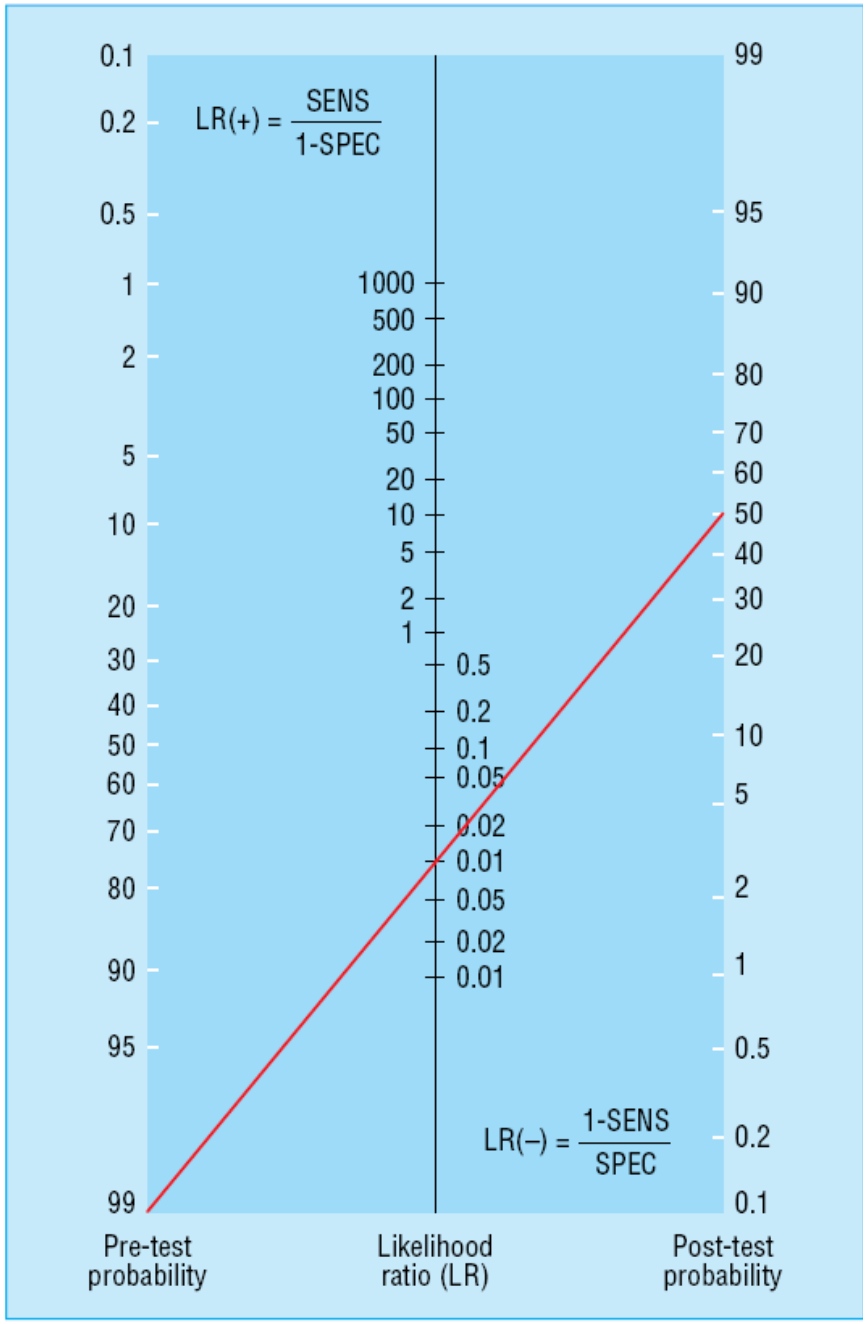


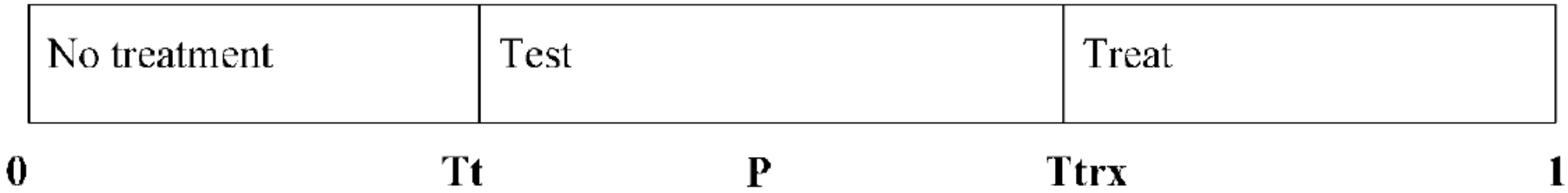
Fig 1 Nomogram (adapted from www.CEBM.net with permission) to convert pre-test probability to post-test probability using the likelihood ratio. The line refers to a text example

T- adjusts probability
downward **LR(–)**
a fraction < 1

Test Results Are Useful In Relation to Conceptual Thresholds for Action

- Test-treatment, or treatment threshold
  - P above which dx sufficiently likely to warrant treatment
  - Pre-test P > treatment threshold
    - Confirmatory test to increase P(D) ***does not contribute.***
- No test-test, or test threshold
  - P below which dx warrants no further consideration
  - Pre-test P < test threshold
    - Exclusionary test to further decrease P(D) ***does not contribute.***

Richardson WS, Wilson MC, Guyatt GH, Cook DJ, Nishikawa J. Users' guides to the medical literature: XV. How to use an article about disease probability for differential diagnosis. Evidence-Based Medicine Working Group. *JAMA* 1999;281:1214-9.



Test may be diagnostically useful when pre-test P(D+) high enough to test for, not high enough to treat, and if the test can move the P(D+) across either threshold

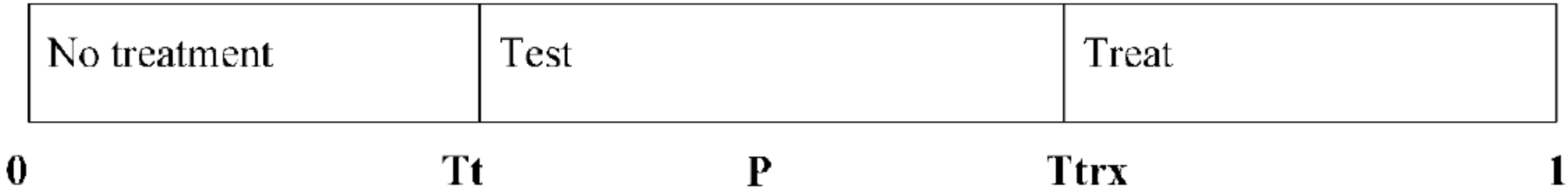
Did you notice, this is the conceptual approach behind the Kaiser sepsis calculator?

Table 4. The likelihood ratios of clinical findings for neontal bacterial infections.

Clinical finding	Likelihood ratio
Common signs	
Pallor	14.4
Poor feeding	8.7
Tachycardia/arrhythmia	5.6
Decreased peripheral perfusion	5.4
Unstable blood pressure	4.0
Abdominal distention	3.5
Apnea	3.1
Lethargy	2.3
Hyperbilirubinemia	2.0
Retractions	1.7
Grunting	1.6
Abnormal tone	1.6
Tachypnea	1.3
Cyanosis	0.3
Temperature instability	0.7
Uncommon signs	
Purpura	47.0
Omphalitis	32.5
Vasomotor instability	8.1
Bleeding	6.5
Pustules	6.1
Bulging fontanel	5.4
Splenomegaly	4.1
Rash	4.0
Diarrhea	3.6
Seizures	2.3

If one is starting with a low probability of bacterial infection, most of these will not substantially change the consideration.

Current Opinion in Infectious Diseases 1995, 8:191–199



- **The error in post-test P attributable to a physician's estimate of pre-test P *might be more important* than the error involved in many medical tests**
- Error or bias in *P* estimates could mean many hypotheses cross the test or test-treat threshold, demanding more tests be performed and more patients be treated, some unnecessarily.
- Some say it is unnatural for people to give numerical estimations, and that **using verbal estimations (such as 'pretty sure' or 'unlikely'), may yield more reliable answers**

Probabilistic reasoning and clinical decision-making:do doctors overestimate diagnostic probabilities? A. CAHAN, D. GILON, O. MANOR and O. PALTIEL, Q J Med 2003; 96:763–769

(BMJ 2006;333:445)

If something always happened, what percentage frequency would you assign to that event? Presumably 100%. And if something never happened? Presumably 0%. Well, not everyone shares that opinion… The table shows combined results of seven studies of what people mean (*Drug Safety* 2005;28:851-70)…

For comparison, …definitions from the *Oxford English Dictionary*. Look, for example, at "occasionally," "infrequently," and "seldom"… according to the dictionary they all mean roughly the same thing. …perhaps when we use words like this we should remember what the German conductor Hans Richter supposedly once said: "Up with your damned nonsense will I put twice, or perhaps once, but sometimes always, by God, never."

Interpretations of words used to indicate frequencies

Word	Interpretation (range of mean percentages)	Definition in the *Oxford English Dictionary*
Invariably/always	91-100	At every time, on every occasion, at all times, on all occasions. Opposed to sometimes, occasionally
Almost always	85-94	—
Normally	71-81	Under normal or ordinary conditions; as a rule, ordinarily
Usually	70-84	In a usual or wonted manner; according to customary, established, or frequent usage; commonly, customarily, ordinarily; as a rule
More often than not	64	—
Common(ly)	56-69	As a usual circumstance; as a general thing; in ordinary cases; usually, ordinarily, generally
Often	42-71	Many times; at many times; on numerous occasions; frequently; for a significant amount or proportion of the time
Frequent(ly)	36-72	At frequent or short intervals, often, repeatedly
Not infrequently	24-35	Rather frequently
Occasionally	17-21	Now and then, at times, sometimes; irregularly and infrequently
On occasion	12	As need or opportunity arises; now and then, occasionally
Infrequently	12-14	Not frequently; somewhat rarely, seldom
Sometimes	11-33	On some occasions; at times; now and then
Seldom	7-8	On few occasions, in few cases or instances, not often; rarely, infrequently
Almost never	2	Scarcely ever
Very rare(ly)	0.8-3	—
Rare(ly)	0.5-9	Seldom, infrequently, in few instances
Exceptionally	0.4-1	Uncommonly, unusually
Never	0-2	At no time or moment; on no occasion; not ever

	BMJ	6W
Invariably/always	91-100	98-100
Almost always	85-94	75-99
Normally	71-81	50->90
Usually	70-84	50-90
More often than not	64	25-100
Common(ly)	56-69	10-80
Often	42-71	50-80
Frequent(ly)	36-72	50-80
Not infrequently	24-35	33-85
Occasionally	17-21	10-40
On occasion	12	10-30
Infrequently	12-14	5-20
Sometimes	11-33	4-40
Seldom	7-8	<2-20
Almost never	2	1-10
Very rare(ly)	.8-3	.5-20
Rare(ly)	.5-9	.1-20
Exceptionally	.4-1	.01-10
Never	0-2	0

Neonatal MRI to Predict Neurodevelopmental Outcomes in Preterm Infants

Woodward, Anderson, Austin, Howard, and Inder

N Engl J Med 2006;355:685-94

Methods

We studied 167 very preterm infants (gestational age at birth, 30 weeks or less) to assess the associations between qualitatively defined white-matter and gray-matter abnormalities on MRI at term equivalent (gestational age of 40 weeks) and the risks of severe cognitive delay, severe psychomotor delay, cerebral palsy, and neurosensory (hearing or visual) impairment at 2 years of age (corrected for prematurity)…

Conclusions

Abnormal findings on MRI at term equivalent in very preterm infants strongly predict adverse neurodevelopmental outcomes at two years of age. These findings suggest a role for MRI at term equivalent in risk stratification for these infants.

Conclusions

Abnormal findings on MRI at term equivalent in very preterm infants ***strongly*** predict adverse neurodevelopmental outcomes at two years of age…

What do they mean by “strongly”? “Almost always”; “often”; “sometimes”? Does it depend on whether you’re speaking to someone at your own NICU or in Boston?

- Using incidence data provided in the article for
    - i. severe cognitive delay
    - ii. severe motor delay
    - iii. CP
    - iv. neurosensory impariment

and based on the test characteristics in the following Table, how much does the post-test probability of certain outcomes change?

Table 5. Sensitivity and Specificity of Findings on MRI and Cranial Ultrasonography in Predicting Severe Neurodevelopmental Impairment at a Corrected Age of Two Years.*

Outcome	Moderate-to-Severe White-Matter Abnormalities (N=35)		Any White-Matter Abnormalities (N=120)		Abnormalities on Cranial Ultrasonography† (N=13)	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
	percent					
Severe cognitive delay						
Value	41	84	89	31	15	95
95% CI	23–61	76–89	70–97	23–39	4–35	89–98
Severe motor delay						
Value	65	85	88	30	18	95
95% CI	39–85	78–90	62–98	22–38	5–44	89–97
Cerebral palsy						
Value	65	84	94	31	18	95
95% CI	39–85	76–89	69–100	24–39	5–44	89–97
Neurosensory impairment						
Value	82	82	89	30	16	95
95% CI	48–97	75–88	65–98	23–38	4–40	89–97
Any neurodevelopmental impairment						
Value	38	89	84	34	11	95
95% CI	25–51	80–94	71–92	25–44	4–23	89–98

* CI denotes confidence interval.
† Abnormalities on cranial ultrasonography were defined as grade III or IV intraventricular hemorrhage or periventricular leukomalacia.

Neonatal MRI to Predict Neurodevelopmental Outcomes in Preterm Infants
Lianne J. Woodward, Ph.D., Peter J. Anderson, Ph.D., Nicola C. Austin, M.D., et al NEJM 2006;355:685-94

Likelihood Ratios

	Moderate to Severe White Matter Abn		Any Abnormality		Abnormality on Cranial Ultrasound	
	LR +	**LR -**	**LR +**	**LR -**	**LR +**	**LR -**
Severe Cognitive Delay	2.56	0.70	1.29	0.36	3	0.89
Severe Motor Delay	4.33	0.412	1.26	0.4	3.6	0.863
Cerebral Palsy	4.06	0.417	1.36	0.19	3.6	0.86
Neurosensory Impairment	4.56	0.22	1.27	0.37	3.2	0.88
Any Neurodevelop Impairment	3.45	0.7	1.27	0.47	2.2	0.94

Remember,

Positive LR: Tells you how much to increase the probability of having a disease, given a positive test result.

Negative LR: This tells you how much to decrease the probability of having a disease, given a negative test result.

	Moderate – Severe White Matter Abnormalities		Any White Matter Abnormalities		Grade III or IV IVH or PVL on HUS	
	LR+ Sens/1-Spec	**LR-** 1-Sens/Spec	**LR+**	**LR-**	**LR+**	**LR-**
Severe cognitive delay Pre-test *P* **17%** Post-test *P*	2.56	0.7	1.29	0.35	3	0.89
	~30%	**~10%**	**~21%**	**~6%**	**~30%**	**~13%**
Severe motor delay Pre-test *P* **10%** Post-test *P*	4.33	0.41	1.26	0.4	3.6	0.86
	~28%	**~4%**	**~12%**	**~4%**	**~25%**	**~7%**
CP Pre-test *P* **10%** Post-test *P*	4.06	0.42	1.36	0.19	3.6	0.86
	~32%	**~4%**	**~12%**	**~2%**	**~24%**	**~8%**
Neurosensory (hearing/vision impaired) Pre-test *P* **11%** Post-test *P*	4.56	0.22	1.27	0.32	3.2	0.88
	~31%	**~2%**	**~13%**	**~3%**	**~27%**	**~10%**

Let's Name The Problem

- Too often, we appear to be locked into ***unrepresentative*** categories for thinking.
- Most of the babies we treat with antibiotics represent indistinct diagnostic categories, for which our evidence base is insufficient to objectively assign probability of disease.
- We often devote insufficient effort exploring differential diagnoses because the underlying pathophysiology resolves spontaneously – so, "it doesn't seem to matter" that diagnosis is less than definitive.
  - If we *rule-out* sepsis, we should *rule-in* the condition that explains the baby's problem.

Let's Name The Problem

- Our EMRs must help us compute the unintuitive, quantitative aspects of our decision making for possible bacterial infection and related differential diagnoses.
- We must move beyond vague, undefined thresholds for action when "ruling out sepsis."
  - At what estimated **probability** value that a patient has a bacterial infection do we **test**, do we **treat**?