AI in Medicine I

AI for Differential Diagnosis

Julia Schnabel

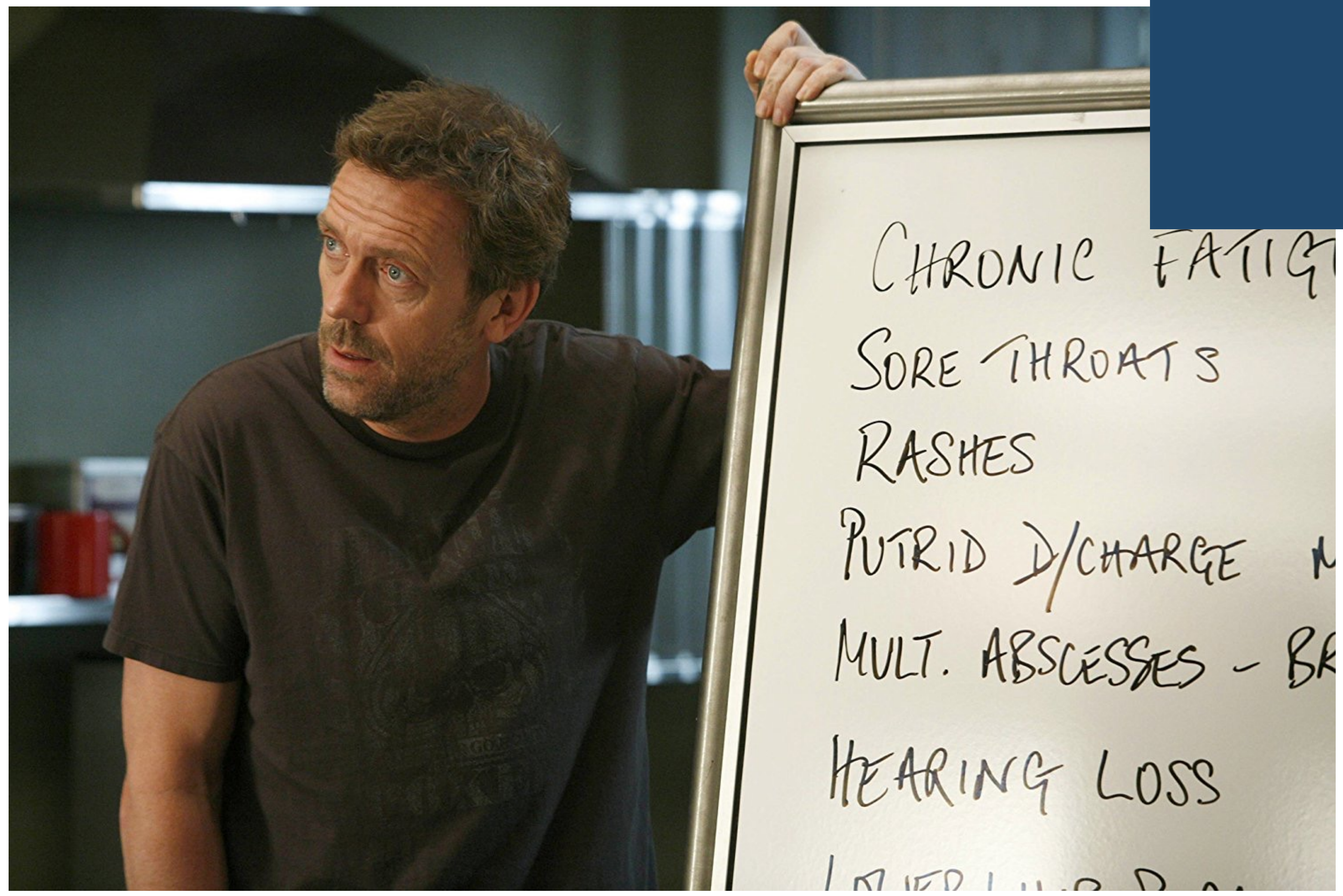
I32 – Chair for Computational Imaging and AI in Medicine

Faculty of Informatics

Tutorial: Haifa Beji

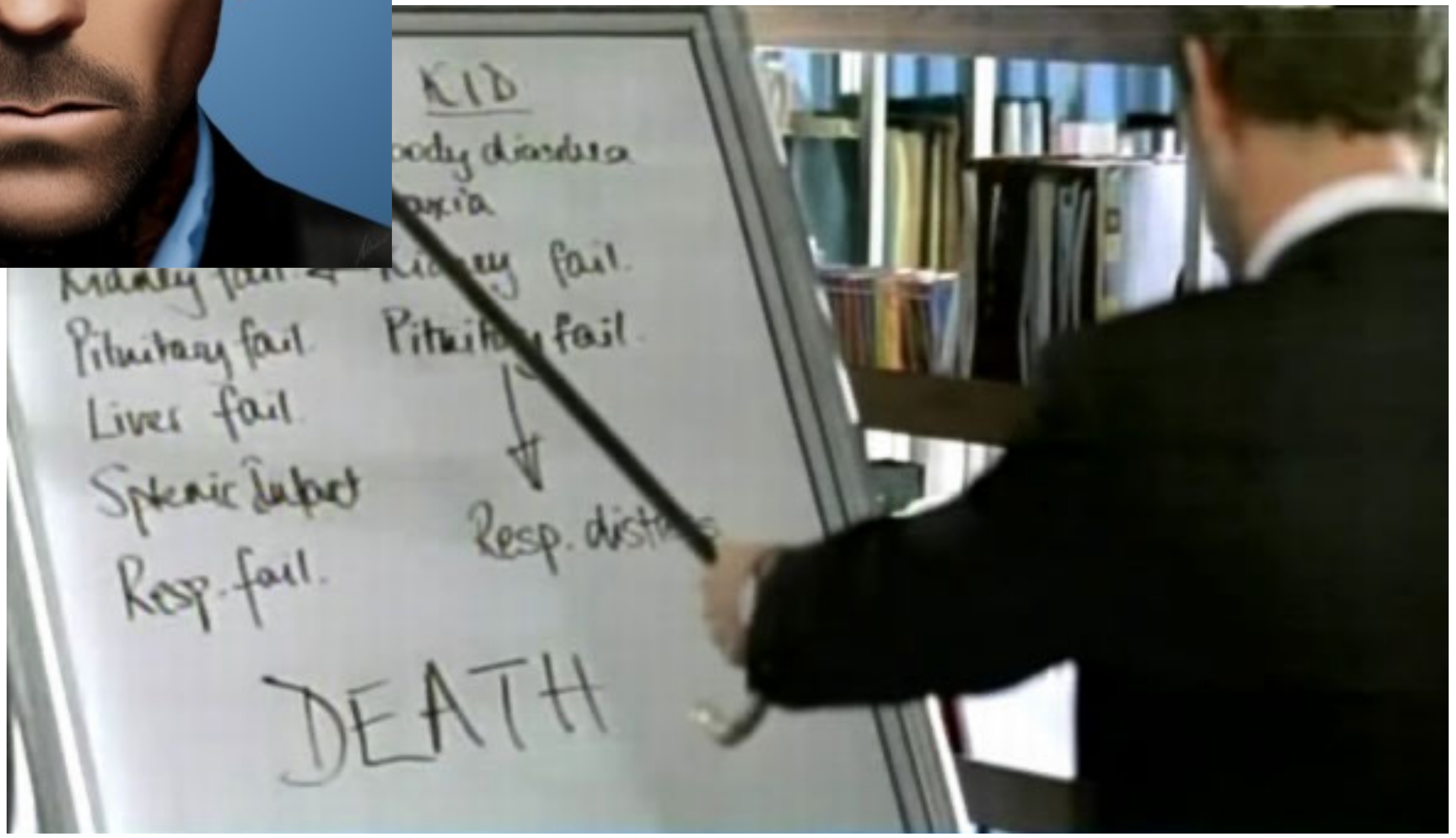
Please provide your online feedback on this lecture series!


SYMPTOMS
CAUSES
TACHYCARDIA
PULMONARY
HEROIN


CHRONIC
SORE THROATS
RASHES
PUTRID D/CHARGE
MULT. ABSCESSES -
HEARING LOSS


HOUSE M.D.


Bloody diarrhea
Ataxia
Kidney fail.
Pituitary fail.
Liver fail.
Bloody diarrhea
Ataxia


KID
Pituitary fail
Liver fail
Splenic Infarct
Resp. fail
Pituitary fail
Resp. dist
DEATH

Overview

- What is Differential Diagnosis?
- Limitations and pitfalls
- Models for Diagnostic Reasoning
- Bayesian inference for differential diagnosis
- The medical test paradox
- Looking beyond Bayes…
- …towards causality

What is Differential Diagnosis?

“In healthcare, a differential diagnosis (abbreviated DDx) is a method of analysis of a patient's history and physical examination to arrive at the correct diagnosis. It involves distinguishing a particular disease or condition from others that present with similar clinical features” *

- *A topical need for differential diagnosis arises in patients with presenting common cold symptoms – which may be quite similar to SARS-COV2 related symptoms.*

*https://en.wikipedia.org/wiki/Differential_diagnosis

What is Differential Diagnosis?

"Differential diagnosis can be regarded as implementing aspects of the hypothetico-deductive method, in the sense that the potential presence of candidate diseases or conditions can be viewed as hypotheses that clinicians further determine as being true or false"**

- *A scientific method that uses empirical claims to populate deductive arguments. This keeps science empirical while yielding conclusions produced by valid, deductive arguments.

* https://en.wikipedia.org/wiki/Differential_diagnosis

What is Differential Diagnosis?

- Differential diagnostic procedures are used by physicians to diagnose the specific disease in a patient, or, at least, to eliminate any imminently life-threatening conditions.

- *"Differential diagnosis is defined as the process of differentiating between probability of one disease versus that of other diseases with similar symptoms that could possibly account for illness in a patient."**

*Machine Learning, Big Data, and IoT for Medical Informatics, 2021

What is Differential Diagnosis?

- Differential diagnostic procedures are used by physicians to diagnose the specific disease in a patient, or, at least, to eliminate any imminently life-threatening conditions.

- Example:
  - In the evaluation of Cough, there might be two possible diagnosis: Acute Bronchitis and Common Cold. Acute Bronchitis will be considered as differential diagnosis even if the final diagnosis is common cold.

What is Differential Diagnosis?

- Clinicians take four steps*:
  1. Gather relevant information about the patient and create a symptoms list
  2. Generate list of possible causes (“**candidate conditions**”) for the symptoms
  3. Prioritise the list by balancing the risks of a diagnosis with the probability.

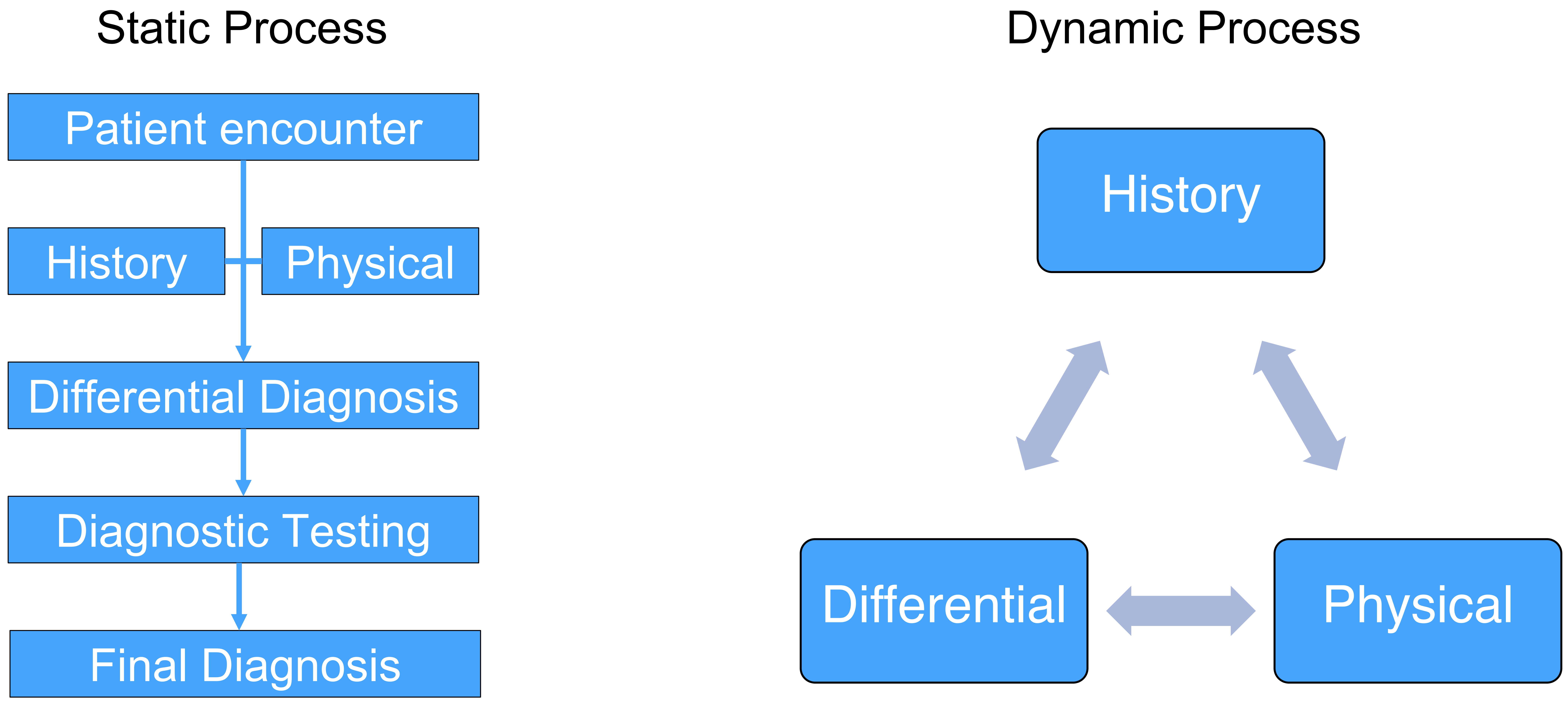
     Normally these are subjective, not objective parameters.
  4. Perform tests to determine the actual diagnosis.

     Known as "**to rule out**".

  Iterate if diagnosis is still not clear, considering the risks, start empirical treatment (“**educated best guess**”) .

* https://en.wikipedia.org/wiki/Differential_diagnosis

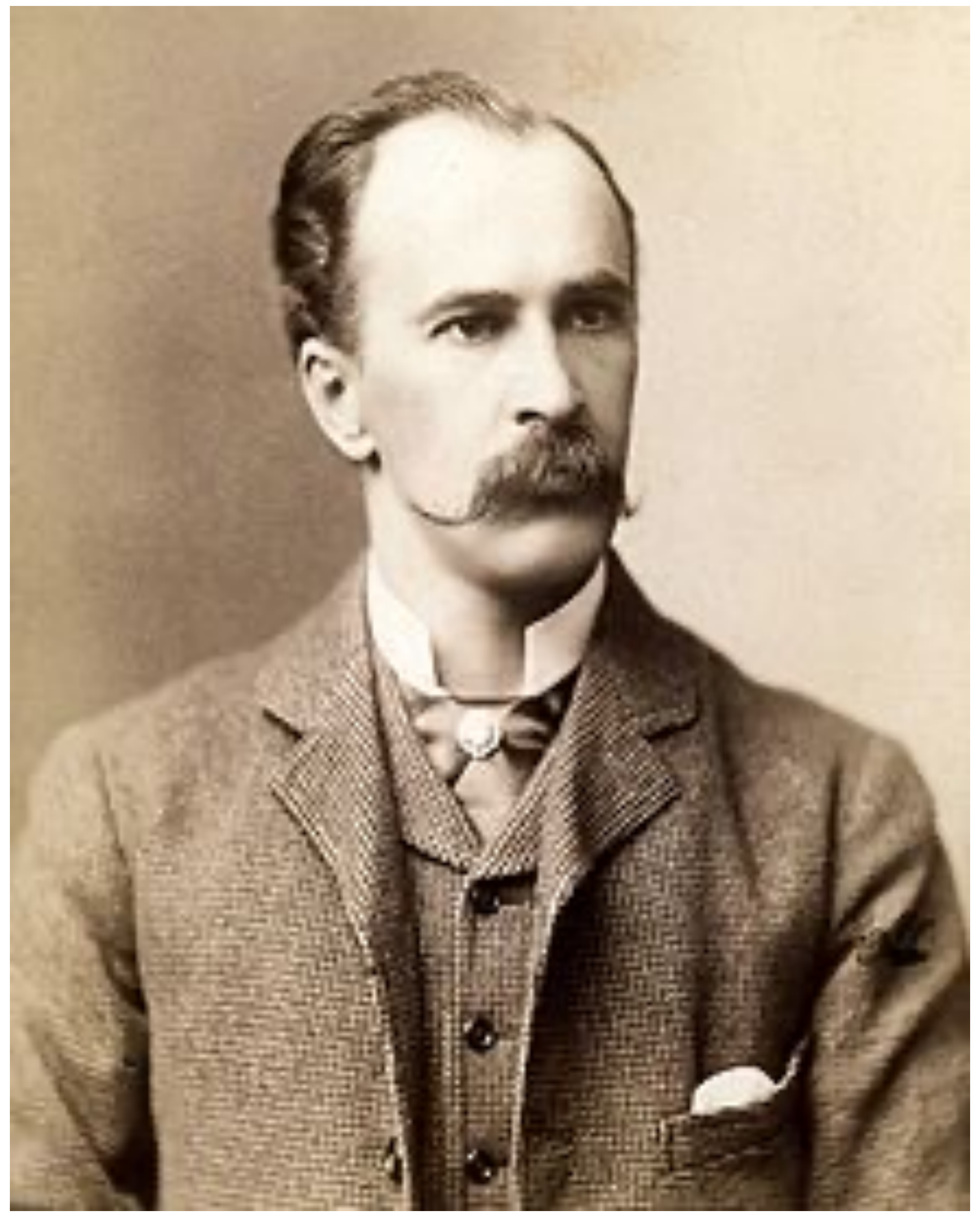
Static vs Dynamic Process



Where is the problem?

- *“If it were not for the great variability among individuals, medicine might as well be a science and not an art.”*

  – Sir William Osler, Johns Hopkins School of Medicine Baltimore (1892)



Osler c. 1880

Osler trained under Rudolf Virchow in Germany. His greatest influence on medicine was to insist that students learn from seeing and talking to patients and the establishment of the medical residency.

Unfortunately he also held views in regards to people’s race that would thankfully today be regarded as completely unacceptable.

Src: https://en.wikipedia.org/wiki/William_Osler

Other quotes:

- *“He who studies medicine without books sails an uncharted sea, but he who studies medicine without patients does not go to sea at all.”*
- *"Listen to your patient, he is telling you the diagnosis"*

Where is the problem?

- *“Most errors in clinical reasoning are not due to incompetence or inadequate knowledge but to the frailty of human thinking under conditions of complexity, uncertainty, and pressure of time.”*
  - Ian Scott, “Errors in clinical reasoning: causes and remedial strategies” BMJ 2009;338:b1860 (2009) https://doi.org/10.1136/bmj.b1860
  - Comments to this article:
    - *“Unfortunately medical schools teach contents rather than a structural reflective approach to a diagnosis”*
    - *“We must insist on applying the same order of stringency to the debiasing strategies or corrective maxims he [Scott] proposes, which are surely no more than heuristic devices by another name”*

Example: Differential diagnosis of chest pain

- There are many illnesses, injuries or conditions that can cause chest pain:
  - Ischaemia
  - Pericarditis
  - Hyperventilation
  - Pulmonary embolism
  - Indigestion
  - Trauma
  - …

According some French colleagues last night, it could also be the effect of watching the 2022 FIFA worldcup final!

Example: Differential diagnosis of chest pain

- Collect patient history:
  - Age
  - Previous episodes
  - Fever
  - Trauma
  - Stress
  - Cardiac disease
  - …

Example: Differential diagnosis of chest pain

- Associated signs/symptoms:
  - Nausea
  - Hypo/hypertension
  - Rash or lesions
  - Pain on palpation
  - ….

Example: Differential diagnosis of chest pain

- Aggravating factors:
  - Breathing
  - Movement
  - Stress
  - After eating
  - Laying down
  - Situational (anxiety)
  - …
- Alleviating factors
  - Rest, decreased movement
  - Position (sitting up, leaning forward)
  - Shallow breathing
  - Diet
  - Medication
  - …

Biases: lots of traps to fall into

- Availability:
  - Past / similar / a lot of it about / common conditions
- Anchoring:
  - First diagnosis, ignoring new evidence
- Representativeness:
  - Best fit: not making differential diagnosis
- Confirmation / attribution bias:
  - Selective listening / make fit
- Premature closure
  - Before all evidence
- Framing effect:
  - Favouring a diagnosis because of context / “Friday afternoon”
- Momentum
  - Drawing others into your believes

Models for Diagnostic Reasoning

- Flowcharts
- Associations between diseases and {signs, symptoms}
- Hypothetico - deductive
- Pattern matching
- Rule based
- Probabilistic

Models for Diagnostic Reasoning

- Flowcharts
- Associations between diseases and {signs, symptoms}
- Hypothetico - deductive
- Pattern matching
- Rule based
- Probabilistic

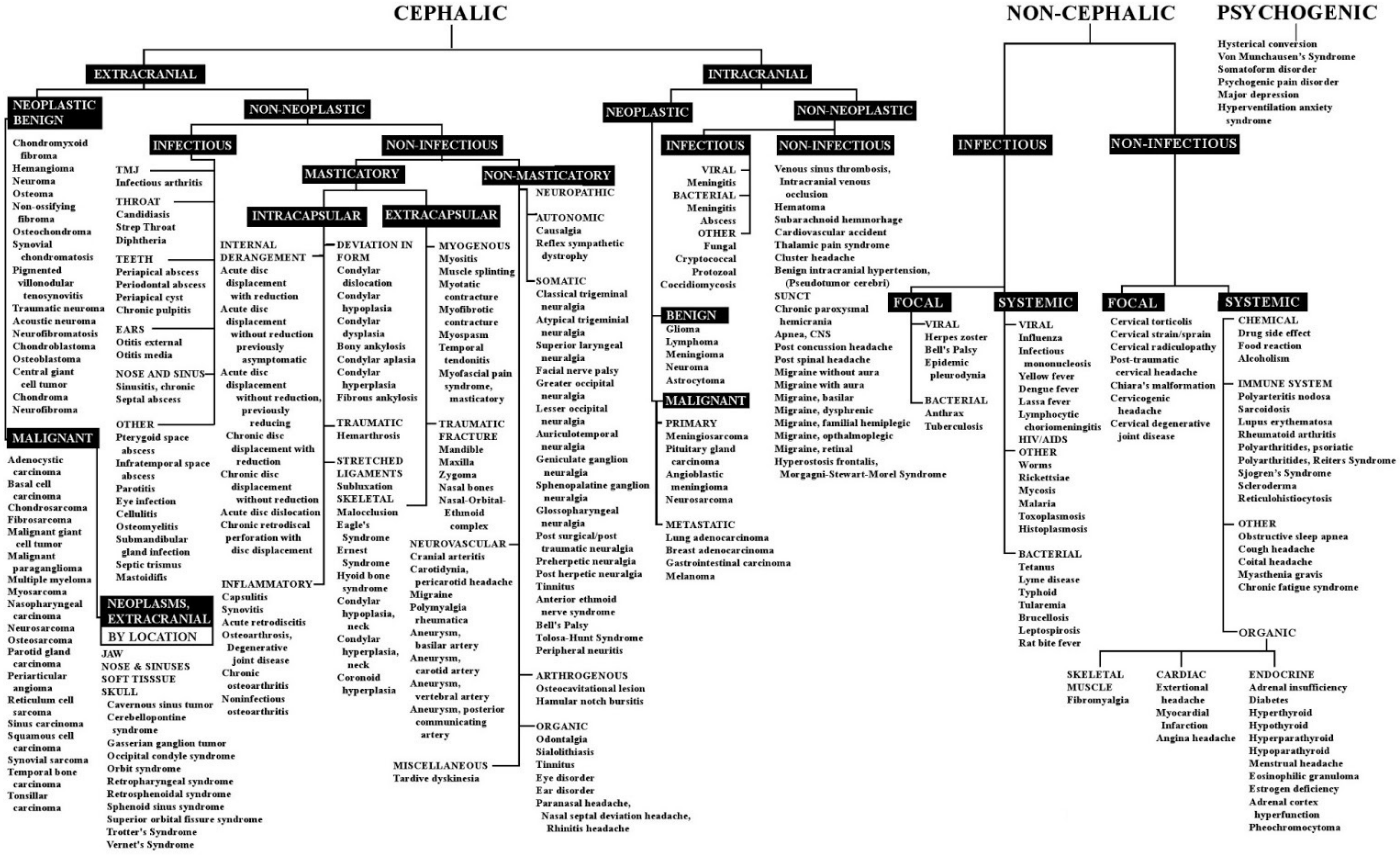
Flowcharts

One of the earliest methods for differential diagnosis, and still in use at hospital triage:

- Involves asking a standard set of questions
- Follows the path on the flow chart based on the answers given by the person and then come to conclusion about the diagnosis.
- Generally not a good method because it is fragile, very specific and does not take unusual cases into consideration.

Flowcharts

DIAGNOSTIC HIERARCHY FOR CHRONIC HEADACHE AND OROFACIAL PAIN



Models for Diagnostic Reasoning

- Flowcharts
- Associations between diseases and {signs, symptoms}
- Hypothetico - deductive
- Pattern matching
- Rule based
- Probabilistic

Disease: signs and symptoms

- Most models of diagnostic reasoning are based on associations between diseases and signs or symptoms

- **Sign**:
  – Any objective evidence of disease
  – Example: fever, rash, blood pressure, …

- **Symptom**:
  – By nature subjective, and only perceived by the patient
  – Example: pain (headache, abdominal/chest/…), fatigue, distress…

Models for Diagnostic Reasoning

- Flowcharts
- Associations between diseases and {signs, symptoms}
- Hypothetico - deductive
- Pattern matching
- Rule based
- Probabilistic

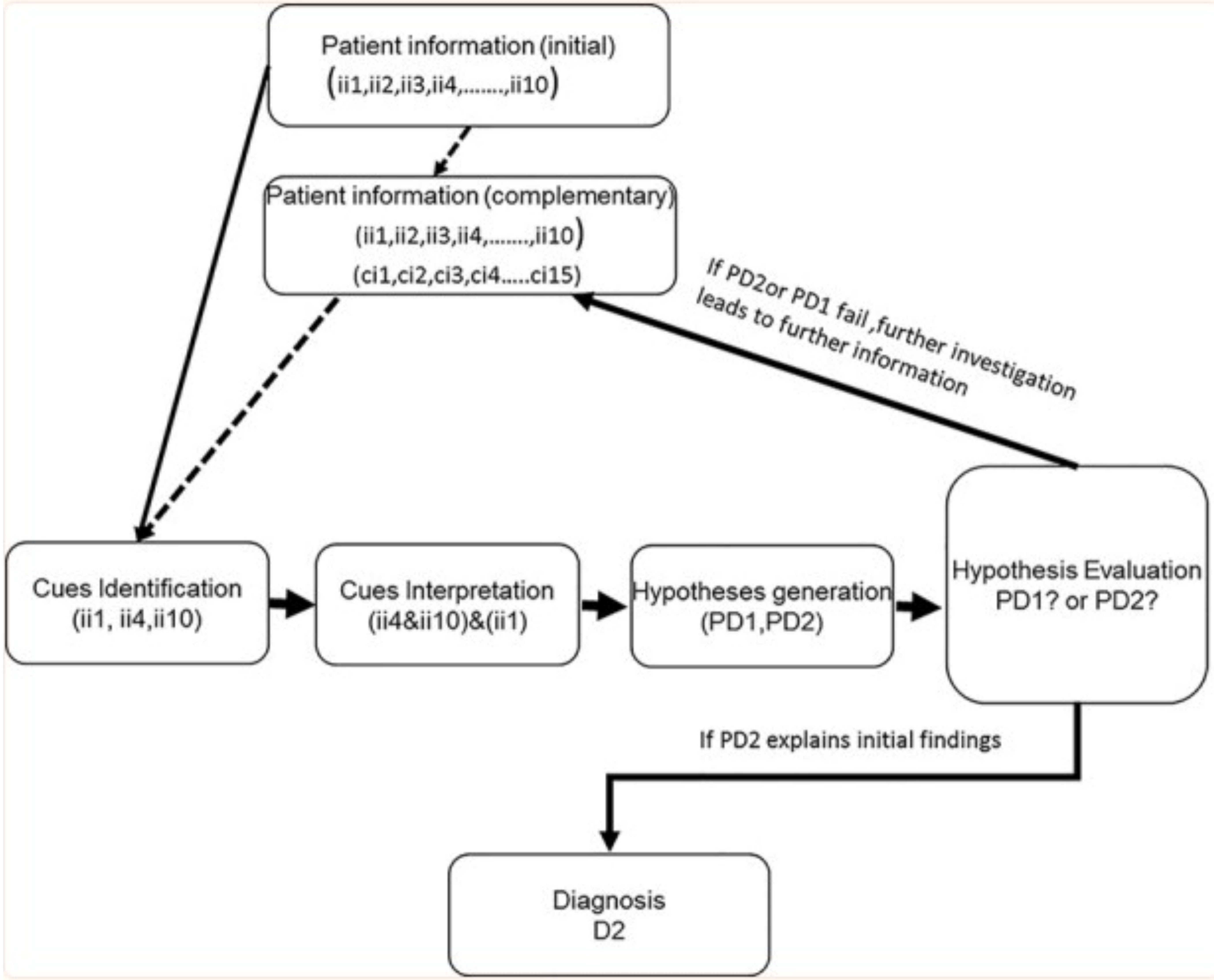
Hypothetico-deductive

Most common approach taken by clinicians:

- At first patient physician visit, few initial information is collected
- The practitioner identifies some information as important cues.
- They interpret the cues to form hypotheses
- If in evaluation one hypothesis explains the findings then the diagnosis is confirmed.
- If all hypotheses fail to explain the findings, further investigation (from more detailed history to lab tests) generates complementary information and the process is repeated

Source: J Adv Med Educ Prof. 2017 Oct; 5(4): 177–184.

Hypothetico-deductive



Source: [J Adv Med Educ Prof. 2017 Oct; 5(4): 177–184.]

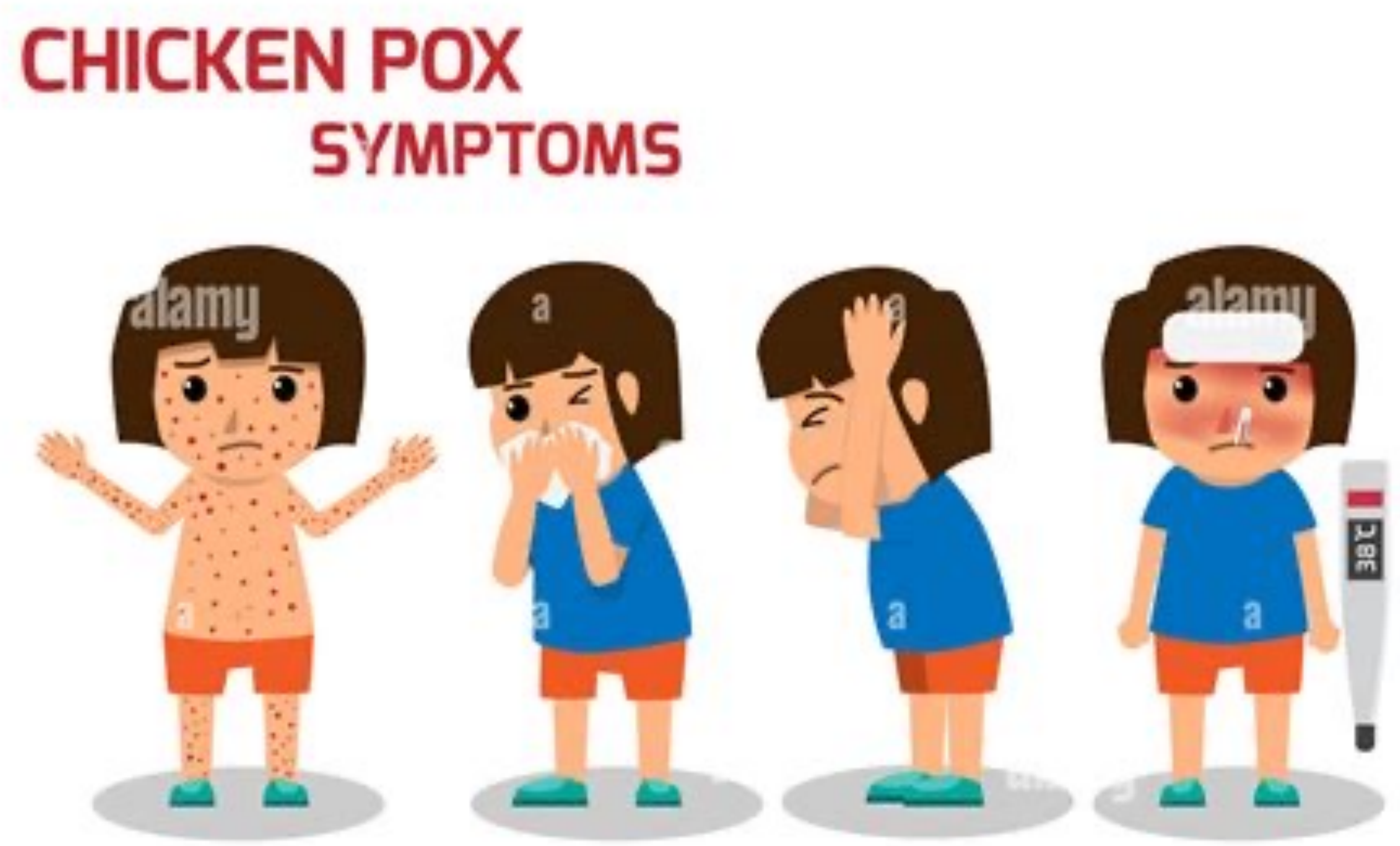
Models for Diagnostic Reasoning

- Flowcharts
- Associations between diseases and {signs, symptoms}
- Hypothetico - deductive
- Pattern matching
- Rule based
- Probabilistic

Pattern matching

Primary method used in cases where diseases are “obvious”

- *“The diagnosis occurs through rapid non-analytical matching of clinical presentation with a pattern previously formed of constructs of clinical signs and symptoms (or pattern) in memory.*
- *The retrieval of these patterns is triggered by recognition of key features within the case.”**



*Source: J Adv Med Educ Prof. 2017 Oct; 5(4): 177–184.

Pattern matching

- Mainly based on certain symptoms or signs being associated with certain diseases or conditions, but not necessarily involving the more cognitive processing involved in a differential diagnosis.
- In principle, a certain pattern of signs or symptoms can be directly associated with a certain therapy, even without a definite decision regarding what is the actual disease.
- But this carries a substantial risk of missing a diagnosis which actually would need to receive a different therapy

*Source: https://www.primidi.com/medical_diagnosis/specific_methods/pattern_recognition

Models for Diagnostic Reasoning

- Flowcharts
- Associations between diseases and {signs, symptoms}
- Hypothetico - deductive
- Pattern matching
- Rule based (mostly categorical/deterministic)
- Probabilistic

Rule-based

- Many of the *earliest** rule-based expert systems were developed for medical diagnosis, aimed to provide diagnostic decision support, via:

- A **knowledge base** in form of a collection of facts and rules
  - Requires **significant human effort**
- An **inference engine**, which acts like a search engine
  - Examines the **information** knowledge base matching the user's query
- A **user interface**, allowing non-expert users to query the system
  - Much easier today than in the 1970s!

**1970s onwards!*

Rule-based

- Expert systems are still around but suffer from several limitations:
  - Complexity of **acquisition of knowledge** for building the **knowledge base**
  - Expert **domain knowledge** needed
  - **Updating of knowledge base** requires ignificant human effort
  - Inference engine often of "**if-then-else**" form, using **heuristic expert knowledge** without any uncertainty, but also allowing for Bayesian reasoning and data-driven approaches
  - Usually designed to support **users with an expert level of medical knowledge**

- Recent examples include diagnostic systems for COVID-19 using medical websites, e.g. medRxiv preprint server, for knowledge base

*1970s onwards!

Rule-based

- Nonetheless, the limitations of expert systems led to advances in **rule learning** and **classification** for differential diagnosis

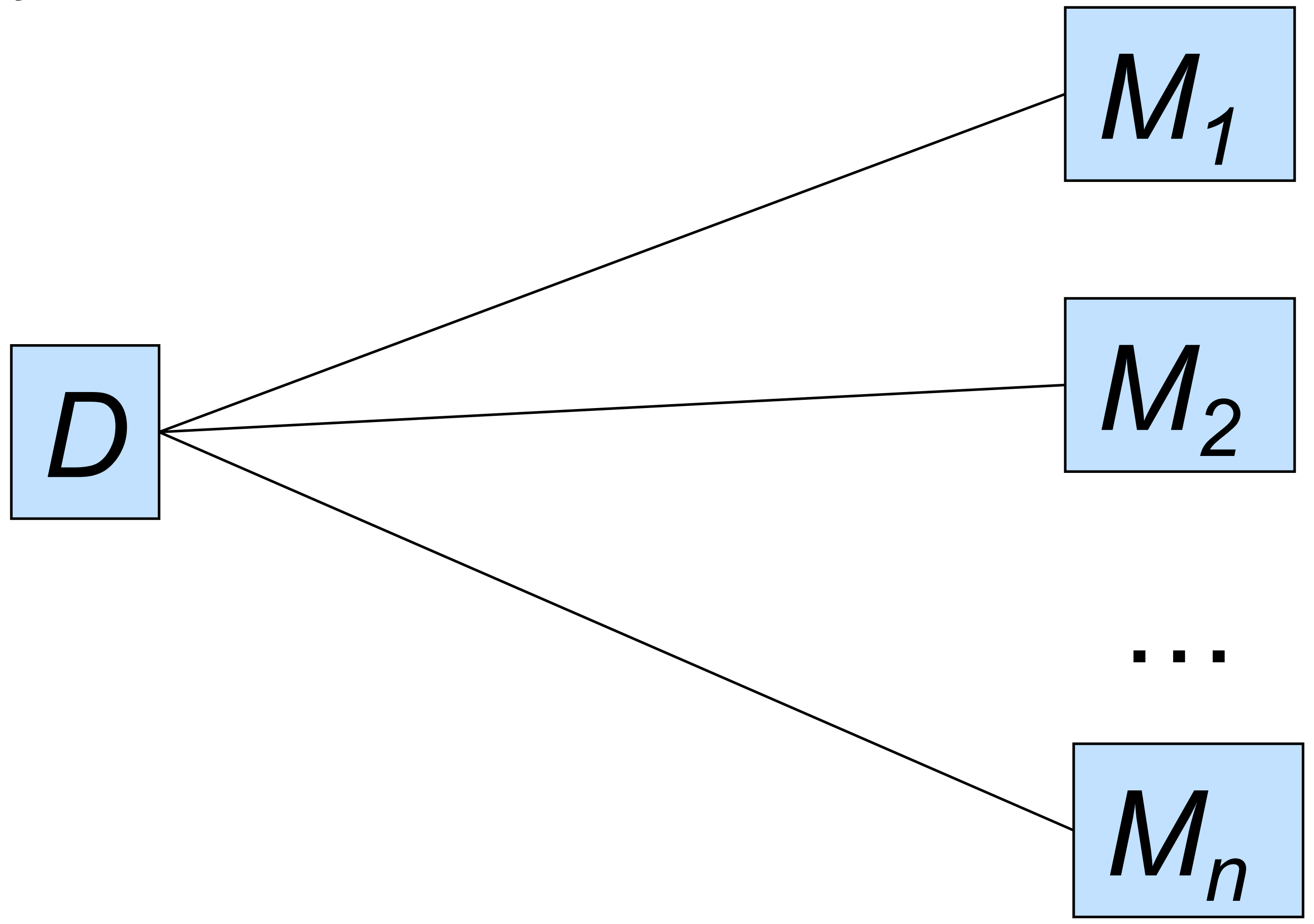
- Knowledge base is now largely replaced by **expert data annotations** on which models are trained
  - Implicit use of expert domain knowledge
  - Expert annotations also require significant human effort and resources

Models for Diagnostic Reasoning

- Flowcharts
- Associations between diseases and {signs, symptoms}
- Hypothetico - deductive
- Pattern matching
- Rule based
- Probabilistic

Probabilistic: Naïve Bayes

- Probabilistic classifier based on applying Bayes' theorem with strong (naïve) independence assumptions between the features
  - Exhaustive and mutually exclusive disease hypotheses (one and only one)
  - Conditionally independent observables (manifestations)
  - $p(D_i)$, $p(M_{ij} \mid D_i)$



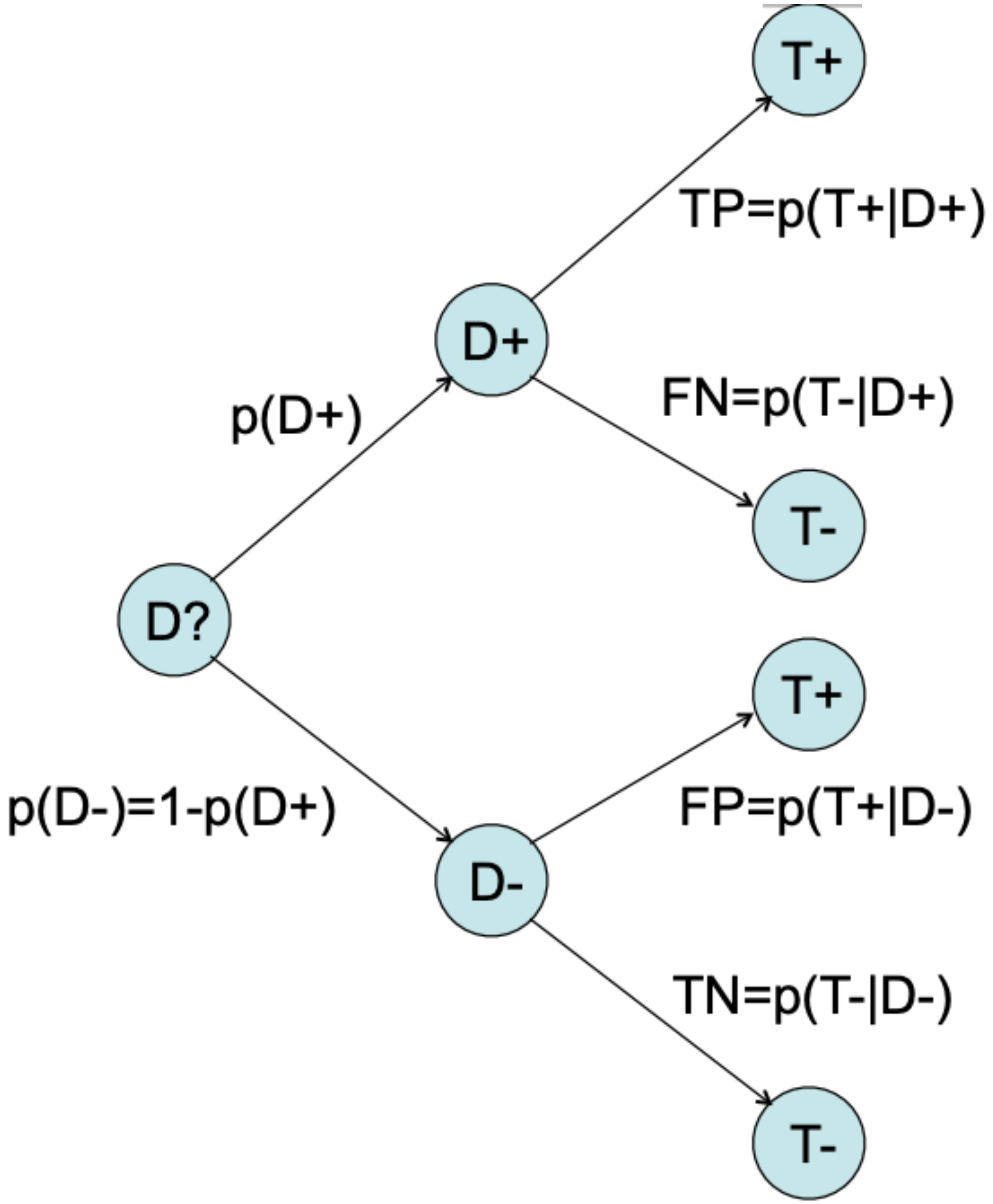
Thomas Bayes (1701 – 7 April 1761) was an English statistician, philosopher and Presbyterian minister .

Bayes never published what would become his most famous accomplishment; his notes were edited and published posthumously by Richard Price. – so Bayes rule is is also called the **Bayes-Price theorem**.

Probabilistic: Naïve Bayes

- Probability of having a disease: p(D+)
  - E.g. disease prevalence in the population
  - …or of not having the disease, p(D-) = 1-p(D+)
- Probability of testing positive for a disease: p(T+) or negative: p(T-)
  - The test could be correct or not, ie we can have true postives (TP) or false positives (FP);
  - Similarly, we can have true negatives (TN) or false negatives (FN)
- Bayes' rule:

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)}$$



Using AI for Differential Diagnosis is not a new concept!

- Over 50 years in action
  - Paper from 1967 already understood the potential of **Bayesian inference** in medical diagnosis:

Medical Diagnosis Using Bayes Theorem

by Thomas L. Lincoln and Rodger D. Parker

A computerized study of the applicability of Bayes theorem to the differential diagnosis of liver disease has been made. Statistical independence of symptoms is not presumed. The semantic obstacle involved in precise definition of the symptom and disease categories is discussed. Input for the study was obtained from patient records, and diagnosis supported by tissue examination, either at autopsy or by biopsy. Correct diagnosis rate is considered sufficiently high to warrant further investigation.

Introduction and Objective

The purpose of this study is to investigate the efficacy of Bayes theorem in the diagnosis of disease states. If a group of patients with known symptoms and a single diagnosis define specified disease, it is possible to calculate the probabilities of particular sets of symptoms, given the specified disease. If, in addition, the prevalence of the disease is known, Bayes theorem may be used to calculate the probability of incidence of specified diseases given a particular set of symptoms even when the diagnosis is not known.

These probabilities may be used in both a quantitative and a qualitative way. In a differential diagnosis, the values of the probabilities assign a rank order to the alternatives, which represents a weaker statement of their relative importance than the values themselves. In the present study, rank order proved to be of particular significance.

Methodology

Probabilities are defined in this study in terms of diseases and sets of symptoms. Let the symbol d_i denote the i^{th} disease of a set of diseases. Let S stand for a set of symptoms exhibited by an individual patient. We shall denote the probability that this patient has disease d_i by $P(d_i \mid S)$. By Bayes theorem this is given as

$$P(d_i \mid S) = C\ P(S \mid d_i)\ P(d_i).$$

$P(d_i)$ is the *a priori* probability of having disease i for the population of which the patient is a member. In other words, $P(d_i)$ is a statement of the prevalence of disease i expressed as its frequency in the population from which the patient is considered to be a sample. (This might be a hospital population, or the population at large, or some other population.)

…not new but still poor understanding in the general public!

Opinion
Mathematics

This article is more than **10 months old**

The Guardian view on the god of science: a divine but rational disagreement
Editorial

Sun 6 Jun 2021 18.25 BST

A key maths tool during the pandemic came about because of an 18th-century debate about Christianity. The lesson we can draw today is that moral philosophy matters

'The work of Thomas Bayes, an 18th-century clergyman and mathematician, has become central to understanding the pandemic.'

The Observer
Coronavirus

The obscure maths theorem that governs the reliability of Covid testing

Tom Chivers

Sun 18 Apr 2021 07.00 BST

613

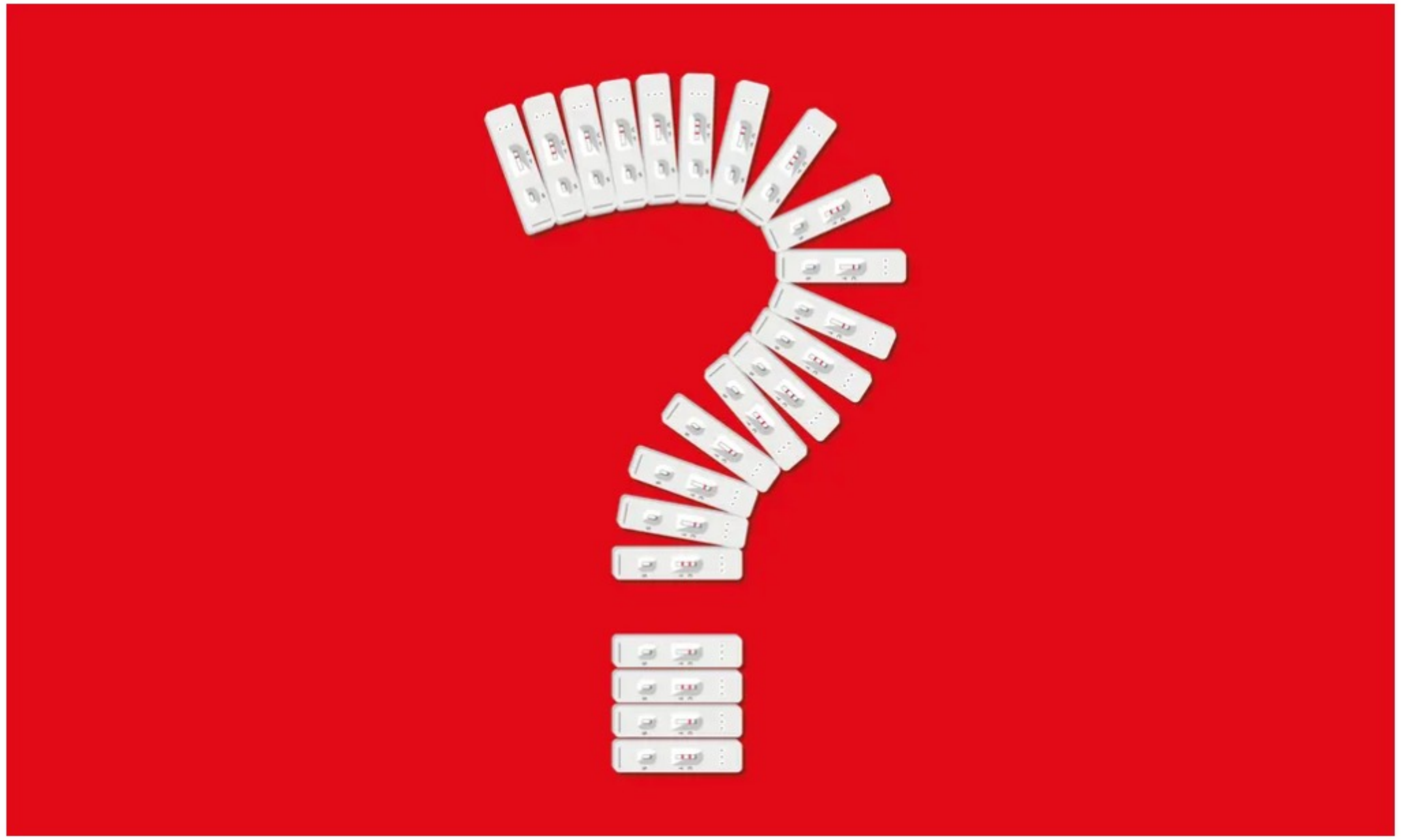


Illustration: Paul Tansley/The Observer

There's been much debate about lateral flow tests - their accuracy depends on context and the theories of a 18th-century cleric

Covid, false positives and conditional probabilities*

- *“Take lateral flow tests. The conditional probability of getting an (incorrect) positive result (call this A), given you are not infected (B), is less than one in 1,000. That rate is very low.*
- ***Bayes’ theorem** shows how to calculate what we really want: the conditional probability you are not infected (B), given you have a positive test (A). In that case, you would be isolating with no benefit.*
  - *Unintuitively, when the virus is rare and there are very few “true positives”, this probability can be high.*
  - *Currently in secondary schools, around 3 in 10 positive lateral flow tests turn out to be false.”*

*Source: The Guardian, 25 April 2021, UK

Bayesian Inference for Differential Diagnosis

- Bayes' theorem allows to revise a probability given additional information:

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)}$$

- **Example #1**: finding out a patient's probability of having liver disease if they are an alcoholic.
- *A* : event that "Patient has liver disease."
  - Past data says that 10% of patients entering your clinic have liver disease: *p(A) = 0.10*
- *B* : test that a "Patient is an alcoholic."
  - 5% of the clinic's patients are alcoholics*: p(B) = 0.05*
- Among the patients diagnosed with liver disease, 7% are alcoholics:
  - *p(B|A) = 0.07* (probability that a patient is alcoholic, given that they have liver disease)

Bayesian Inference for Differential Diagnosis

- Bayes' theorem allows to revise a probability given additional information:

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)}$$

- **Example #1**: Rearranging Bayes' theorem also tells you:

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}$$

- We have already found: *p(A)* = 0.10, *p(B)* = 0.05, *p(B*/*A)* = 0.07 , hence:

$$p(A|B) = \frac{0.07 \ * 0.1}{0.05} = 0.14$$

  - In other words, if the patient is an alcoholic, their chances of having liver disease is 0.14 (14%). This is a large increase from the 10% suggested by past data, *p(A).*
  - But it's still unlikely that any particular patient has liver disease.

Bayesian Inference for Differential Diagnosis

- Bayes' theorem allows to revise a probability given additional information:

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)}$$

- **Example #2:** In a particular pain clinic, 10% of patients are prescribed narcotic pain killers. Overall, five percent of the clinic's patients are addicted to narcotics (including pain killers and illegal substances). Out of all the people prescribed pain pills, 8% are addicts.
- *If a patient is an addict, what is the probability that they will be prescribed pain pills?*

Bayesian Inference for Differential Diagnosis

- Bayes' theorem allows to revise a probability given additional information:

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)}$$

- **Example #2:** Solve in 4 steps:
  1. Figure out what your event "A" is from the question.
  2. Figure out what your event "B" is from the question.
  3. Figure out what the probability of event B (Step 2) given event A (Step 1), ie p(BIA)
  4. Insert your answers from Steps 1, 2 and 3 into the formula and solve.

- You can have a go in the tutorial!

Bayesian Inference for Differential Diagnosis

- Bayes' theorem allows to revise a probability given additional information:

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)}$$

- **Example #3**: Given the following statistics, what is the probability that a woman has cancer if she has a positive mammogram result?
  - One percent of women over 50 have breast cancer.
  - Ninety percent of women who have breast cancer test positive on mammograms.
  - Eight percent of women will have false positives.

- We will again leave this to the tutorial!

Bayesian Inference for Differential Diagnosis

- Bayes' theorem allows to revise a probability given additional information:

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)}$$

- There are **several forms of Bayes' Theorem** out there, and they are all equivalent (they are just written in slightly different ways).
- In this equation, "X" is used in place of "B", with some changes in the denominator:

$$\Pr(\mathrm{A}|\mathrm{X}) = \frac{\Pr(\mathrm{X}|\mathrm{A})\Pr(\mathrm{A})}{\Pr(\mathrm{X}|\mathrm{A})\Pr(A) + \Pr(\mathrm{X}|\sim \mathrm{A})\Pr(\sim A)}$$

The medical test paradox

- To evaluate the performance of diagnostic tests, we can also use a **confusion matrix**:
  - TP: If a patient is predicted positive and is truly affected by the disease
  - TN: If a patient is predicted negative and is healthy
  - FP: People considered affected by disease but are healthy
  - FN: People considered healthy but are affected by disease

	Predicted Positive	Predicted Negative
Affected	TP	FN
Not affected	FP	TN

The medical test paradox

- Example: Medical test, reported to have an *accuracy* of 99%, for a disease that *affects* 0.1% of the population, if performed on 100,000 people.

- We now wish to know: *What is the probability of the test to result $P(+)$ and, to be affected by the disease $P(D \mid +)$?*

	Predicted positive	Predicted negative
Affected	97	3
Not affected	999	98901

The medical test paradox

- Example: Medical test, reported to have an *accuracy* of 99%, for a disease that *affects* 0.1% of the population, if performed on 100,000 people.

	Predicted positive	Predicted negative
Affected	97	3
Not affected	999	98901

- Given the accuracy of the test of 99%, we might think that there is a 99% chance that the test is accurate.
- **But the confusion matrix and Bayes tell us something else!**

Accuracy:

$$ACC = \frac{TP+TN}{P+N},\ P = TP + FN,\ N = TN + FP$$

For the example above:

$$ACC = \frac{97+98901}{(97+3)+(98901+999)} = \frac{98998}{100000} = \sim 99\%$$

The medical test paradox

- Example: Medical test, reported to have an *accuracy* of 99%, for a disease that *affects* 0.1% of the population, if performed on 100,000 people.

	Predicted positive	Predicted negative
Affected	97	3
Not affected	999	98901

- The probability to be affected by the disease (*D*), without testing is equal to the spread of it in the population. This is the *a priori* assumption:

  $p(D) = 0.1\%$

- The opposite, to not have the disease:

  $p(\neg D) = 1\text{-}p(D) = 99.9\%$

- The probability to test positive is the sum of all positives divided by the total population tested:

  $p(+) = (97+999)/100{,}000 = 1.096\%$

- The opposite of that is

  $p(-) = 1\text{-p}(+) = 98.904\%$

The medical test paradox

- Using the confusion matrix, we can also find sensitivity and specificity:
  - TPR (*Sensitivity)* :

    $$p(+ \mid D) = \frac{97}{97+3} = 97\%$$

  - TNR (*Specificity)* :

    $$p(- \mid \neg D) = \frac{98901}{98901+999} = 99\%$$

  - FNR (*Missing Rate):*

    $$p(- \mid D) = \frac{3}{3+97} = 3\%$$

  - FPR (*False Alarm):*

    $$p(+ \mid \neg D) = \frac{999}{999+98901} = 1\%$$

	Predicted positive	Predicted negative
Affected	97	3
Not affected	999	98901

	Predicted Positive	Predicted Negative
Affected	True Positive Rate TP/TP+FN	False Negative Rate FN/FN+TP
Not affected	False Positive Rate FP/FP+TN	True Negative Rate TN/TN+FP

The medical test paradox

- We can now rewrite all this using Bayes' theorem:

$$p(D|+) = \frac{p(+|D)p(D)}{p(+)} = \frac{97\%\ 0.1\%}{1.096\%} = {\sim}8.9\%$$

$$p(\neg D|+) = \frac{p(+|\neg D)p(\neg D)}{p(+)} = \frac{1\%\ 99.9\%}{1.096\%} = {\sim}91.1\%$$

$$p(D|-) = \frac{p(-|D)p(D)}{p(-)} = \frac{3\%\ 0.1\%}{98.904\%} = {\sim}0\%$$

$$p(\neg D|-) = \frac{p(-|\neg D)p(\neg D)}{p(-)} = \frac{99\%\ 99.9\%}{98.904\%} = {\sim}100\ \%$$

The medical test paradox

- In answer to our earlier question, what the probability is of being affected with the disease while testing positive:

- *We can see that if we would test positive with an **accuracy** of 99%, we would still have*
  $p(\neg D \mid +) = 91.1\%$, of probability to be healthy and only
  $p(D \mid +) = 8.9\%$, of probability to be sick.

  – *That's way less scary than 99%.*

AI for differential diagnosis

- "*Artificial Intelligence System Approaching Neuroradiologist-level Differential Diagnosis Accuracy at Brain MRI*"*
- Probabilistic diagnosis in patients with 19 common and rare diagnoses at brain MRI
- Combines data-driven and domain-expertise methodologies, including deep learning and Bayesian networks
- Steps:
  1. Lesion detection using deep learning
  2. Extraction of 18 quantitative imaging features using atlas-based co-registration and segmentation
  3. Image features from 2. were combined with five clinical features by using Bayesian inference to develop probability-ranked differential diagnoses

**Rauschecker, Rudie et al. Radiology (2020) Vol. 295, No. 3 https://doi.org/10.1148/radiol.2020190283*

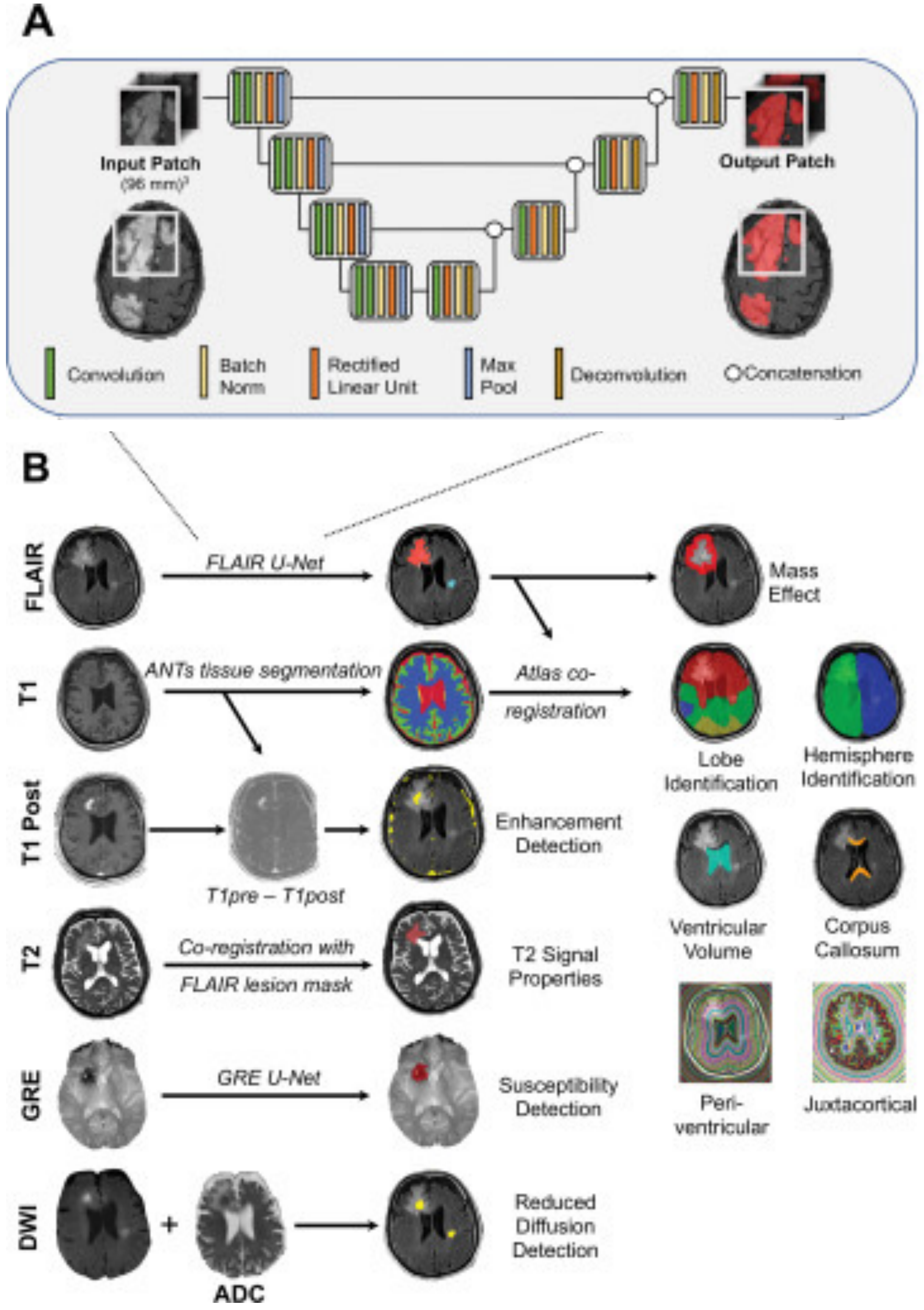
AI for differential diagnosis

A. U-net for lesion segmentation:

- training (n = 86) and test (n = 92) sets
- 20 different physical MRI scanners (16 scanner models across multiple locations)

B. MR sequences:

- T1-weighted, T1 postcontrast, T2-weighted, FLAIR, diffusion-weighted, apparent diffusion coefficient, and gradient-echo or susceptibility-weighted imaging
- Co-registration of sequences to propagate extracted lesion masks and brain segmentations



*Rauschecker, Rudie et al. Radiology (2020) Vol. 295, No. 3 https://doi.org/10.1148/radiol.2020190283

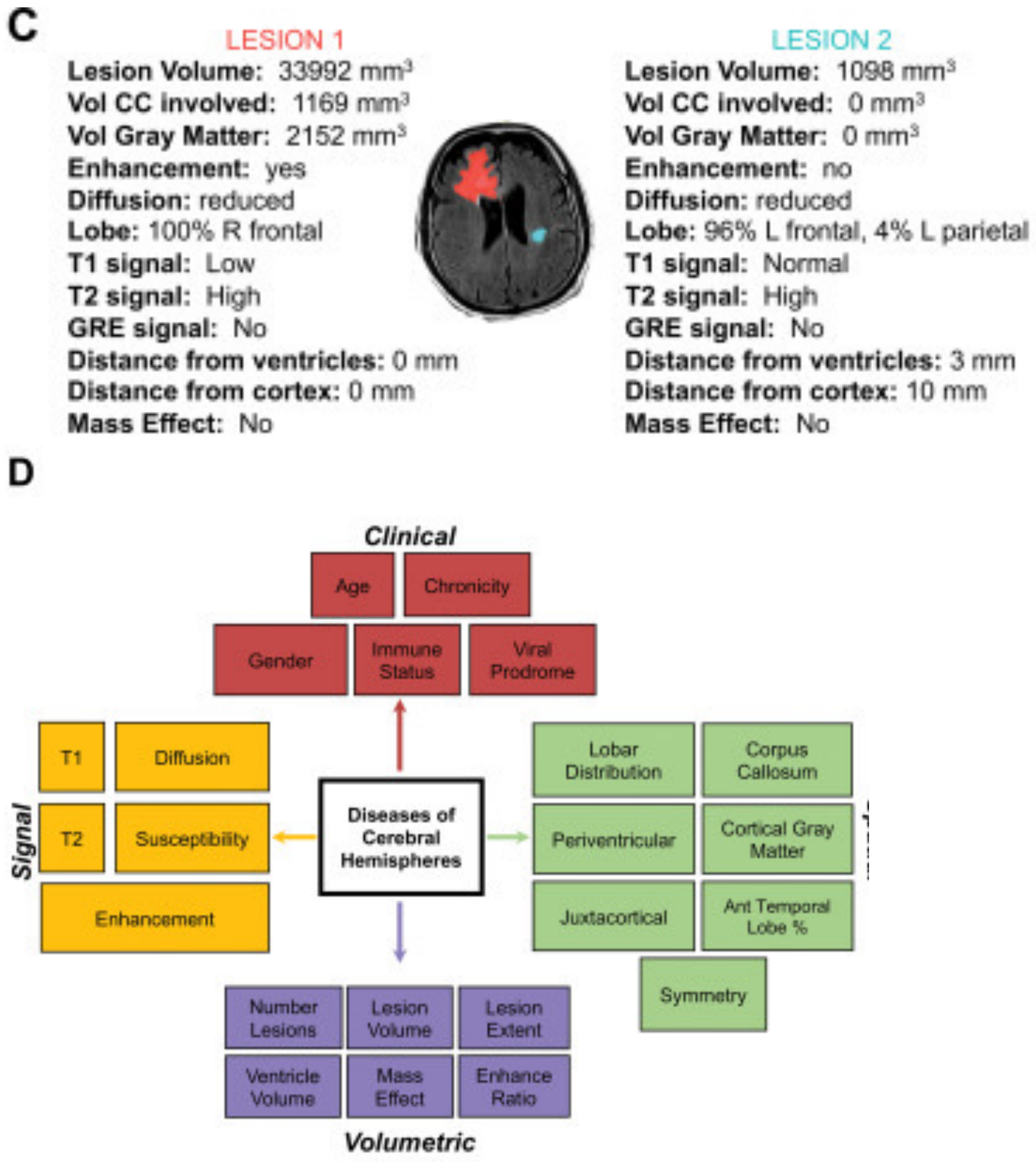
AI for differential diagnosis

C. Lesion quantification:

– 18 extracted signal (n = 5), volumetric (n = 6), spatial (n = 7) quantitative features, plus five clinical features

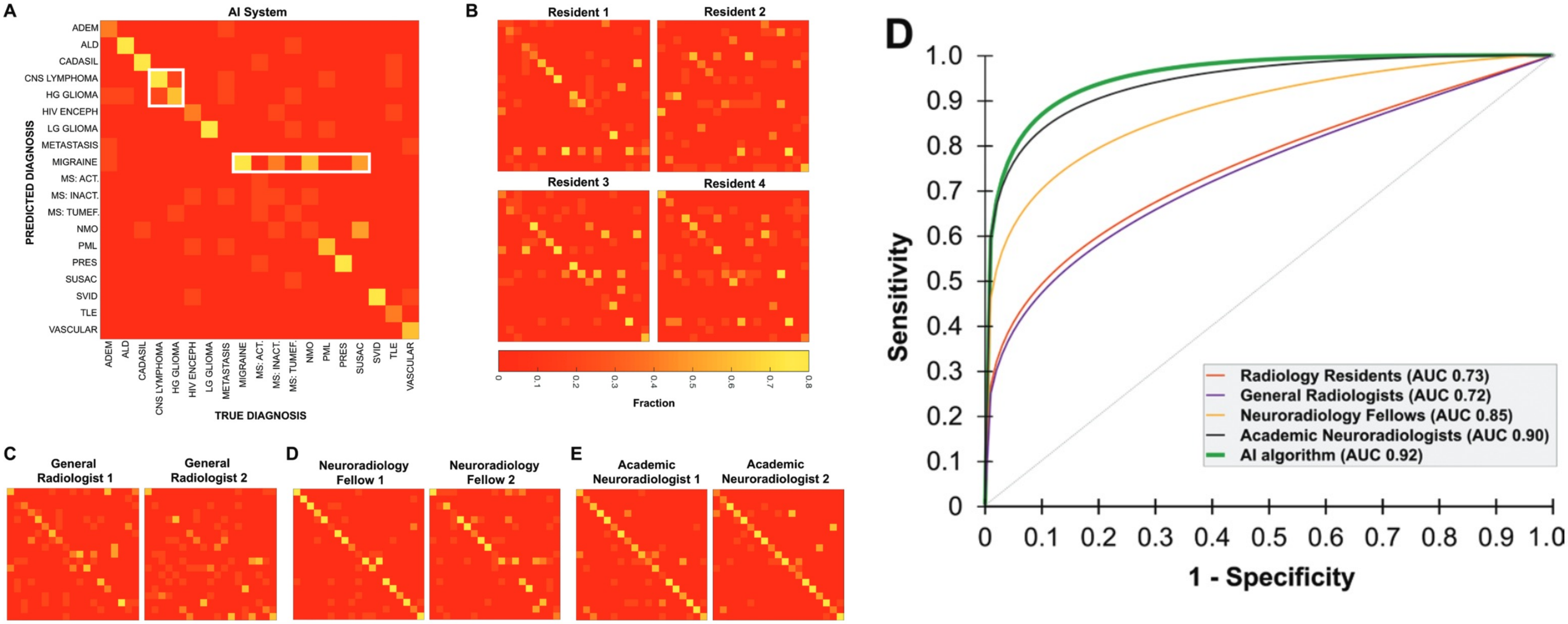
D. Analysis using naïve Bayesian inference to calculate a probability for each possible diagnosis

– https://github.com/rauscheck/radai



*Rauschecker, Rudie et al. Radiology (2020) Vol. 295, No. 3 https://doi.org/10.1148/radiol.2020190283

AI for differential diagnosis



Limitations of (most) current AI medical diagnosis systems

- Current systems, including Bayesian model-based and deep learning approaches, rely on so-called associative inference:
  - they identify diseases based on how **correlated** they are with a patients symptoms and medical history
  - This is in contrast to how doctors perform diagnosis, selecting the diseases which offer the best causal explanations for the patients' symptoms
- There is a need to disentangle correlation from causation
- One approach is to design a counterfactual diagnostic algorithm*

*Richens, J.G., Lee, C.M. & Johri, S. Improving the accuracy of medical diagnosis with causal machine learning. *Nat Commun* **11,** 3923 (2020). https://doi.org/10.1038/s41467-020-17419-7

The Three Layer Causal Hierarchy*

- “Current machine learning systems operate, almost exclusively, in a statistical, or model-free mode, which entails severe theoretical limits on their power and performance”

The Three Layer Causal Hierarchy

Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y\|x)$	Seeing	What is? How would seeing X change my belief in Y?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y\|do(x), z)$	Doing Intervening	What if? What if I do X?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x\|x', y')$	Imagining, Retrospection	Why? Was it X that caused Y? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past 2 years?

Figure 1: The Causal Hierarchy. Questions at level i can only be answered if information from level i or higher is available.

*Judea Pearl, “Theoretical Impediments to Machine Learning With Seven Sparks from the Causal Revolution”, 2018, https://doi.org/10.48550/arXiv.1801.04016

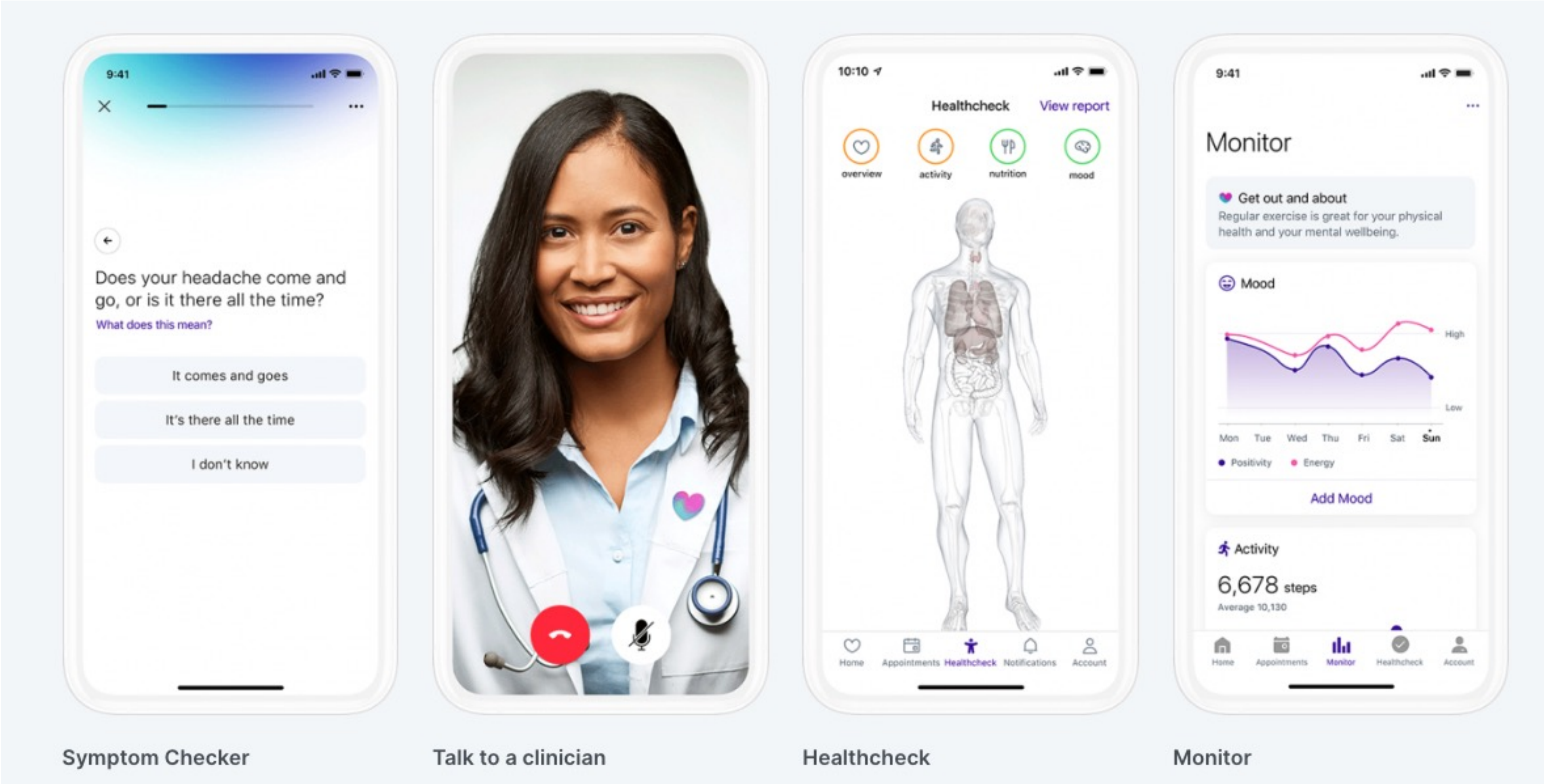
The Three Layer Causal Hierarchy*

- Association:
  - invokes purely statistical relationships, defined by the naked data
  - “Observing a customer who buys toothpaste makes it more likely that they buy floss; such association can be inferred directly from the observed data using conditional expectation”
- Intervention:
  - ranks higher than Association as it involves not just observation, but changes what we see
  - "What happens if I take aspirin?”
- Counterfactuals
  - Top of the hierarchy
  - A typical question is “What if I had acted differently?”

*Judea Pearl, “Theoretical Impediments to Machine Learning With Seven Sparks from the Causal Revolution”, 2018, https://doi.org/10.48550/arXiv.1801.04016

Commercial solutions

- Babylon Health



Richens, J.G., Lee, C.M. & Johri, S. Improving the accuracy of medical diagnosis with causal machine learning. *Nat Commun* **11,** 3923 (2020). https://doi.org/10.1038/s41467-020-17419-7

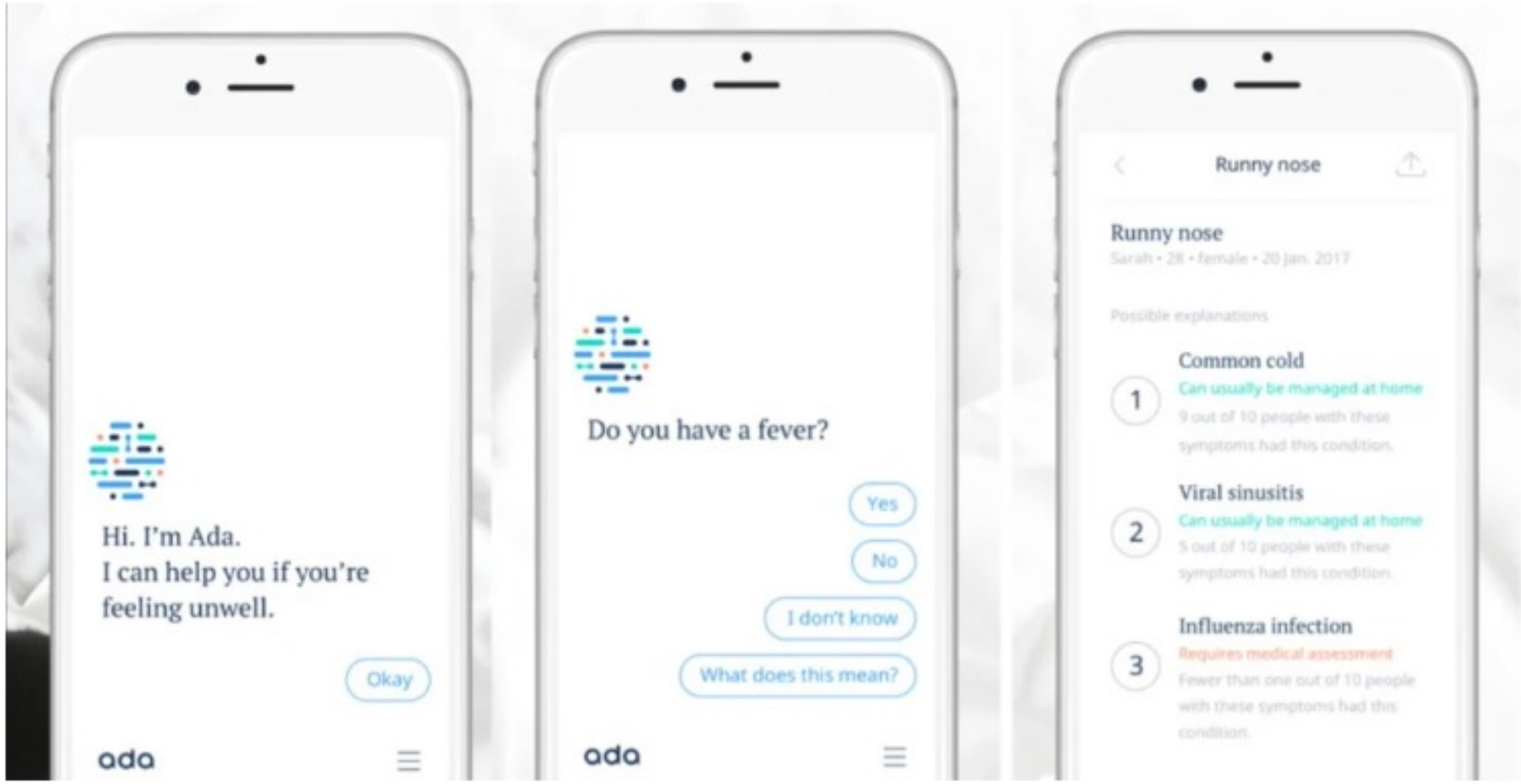
Commercial solutions

- Babylon Health
  - Uses machine learning to diagnose disease, overcoming limitations of earlier versions by using causal reasoning in its machine learning*
  - Previously, diagnoses were based solely on correlations between symptoms and the most likely cause.
  - Reformulate diagnosis as a counterfactual inference task and derive counterfactual diagnostic algorithms

Richens, J.G., Lee, C.M. & Johri, S. Improving the accuracy of medical diagnosis with causal machine learning. *Nat Commun* **11,** 3923 (2020). https://doi.org/10.1038/s41467-020-17419-7

Commercial solutions

- Ada Health



Ronicke, S., Hirsch, M., Türk, E. *et al.* Can a decision support system accelerate rare disease diagnosis? Evaluating the potential impact of Ada DX in a retrospective study. *Orphanet J Rare Dis* **14,** 69 (2019). https://doi.org/10.1186/s13023-019-1040-6

Commercial solutions

- Ada Health

ada

- Based on the input of patient symptoms and supported by the Ada knowledge base and reasoning engine, Ada DX generates two ranked lists of differential diagnoses including both common and rare diseases:
- ‘probability’ list is ranked by the disease probability estimations
- ‘fit’ (rare) list is ranked by the estimated fit of the symptom constellation to the disease - **regardless of its prior probability.**

Ronicke, S., Hirsch, M., Türk, E. *et al.* Can a decision support system accelerate rare disease diagnosis? Evaluating the potential impact of Ada DX in a retrospective study. *Orphanet J Rare Dis* **14,** 69 (2019). https://doi.org/10.1186/s13023-019-1040-6

From diagnostics to real-time prediction: Zoe app

Zoe Covid symptom tracker app:

- 4.5m global contributors
- Discovered 20 new Covid symptoms
- Led to change in UK health policy
- Predicts incidence of new cases daily

- Top five symptoms of Omicron*:
  - **Runny nose**
  - **Headache**
  - **Fatigue (mild or severe)**
  - **Sneezing**
  - **Sore throat**

**Only half of the people now experience the classic three symptoms of fever, cough, or loss of sense of smell or taste common with the earlier variants – in fact, only 1 in 5 experience loss of sense of smell or taste with Omicron.*

Summary

- AI takes an increasingly important role in differential diagnosis tasks in healthcare:
  - Clinicians approach diagnosis in an intuitive and deductive manner, whereas AI takes a more analytical and inductive approach
  - Clinicians' reasoning may be affected by fatigue, interruptions, cognitive overload and biases
  - AI in this context generally takes on a supporting, complementary role, but it is more objective – however it is still subject to biases!
  - AI can enable clinicians to favour one hypothesis over another, and even generate hypotheses not previously considered

- We will revisit some of these concepts in our upcoming lectures on e.g. **risk scores**, **trustworthy AI**, **causal inference**, and **reinforcement learning**

References

- Ian Scott, *"Errors in clinical reasoning: causes and remedial strategies",* BMJ 2009;338:b1860 (2009) https://doi.org/10.1136/bmj.b1860
- David Spiegelhalter. *"The Art of Statistics (How to Learn from Data)"* (Pelican Books, 2019)
- David Spiegelhalter and Anthony Masters. *"Covid By Numbers: Making Sense of the Pandemic with Data"* (Pelican Books, 2021)
- Daniel Kahnemann, Olivier Sibony, Cass R Sunnstein. *"Noise: A Flaw in Human Judgment"* (2021)

References

- Rauschecker, Rudie et al. *"Artificial Intelligence System Approaching Neuroradiologist-level Differential Diagnosis Accuracy at Brain MRI"*, Radiology (2020) 295(3) https://doi.org/10.1148/radiol.2020190283
  - See also https://pubs.rsna.org/doi/full/10.1148/radiol.2020200819
- Pearl, Judea *"Theoretical Impediments to Machine Learning With Seven Sparks from the Causal Revolution"*, 2018, https://doi.org/10.48550/arXiv.1801.04016
- Richens, J.G. et al. *"Improving the accuracy of medical diagnosis with causal machine learning*". *Nat Commun* **11,** 3923 (2020). https://doi.org/10.1038/s41467-020-17419-7
- Ronicke, S.,. *et al.* *"Can a decision support system accelerate rare disease diagnosis? Evaluating the potential impact of Ada DX in a retrospective study*". *Orphanet J Rare Dis* **14,** 69 (2019). https://doi.org/10.1186/s13023-019-1040-6

Happy holidays and a great start into 2023!