Olivier Cussenot

Préface Denis Kessler



Medical Decision-Making in the age of Artificial Intelligence



Webinar 2 March 2023

···· OUTLINE ·····

1-Introduction to decision-making in medicine
2-Human intelligence & decision-making
3-Algorithmic medicine & decision-making
4-Conclusion

1-Introduction to decision-making in medicine

Purposes & Context

Actors & audiences







Predictive & Decision-making in Medicine



Key issues according to J. Pearl



* <u>Prediction</u>:

Announce an event in advance by calculation or by reasoning

Decision:

Choosing between several solutions that may solve a problem

*<u>Regret:</u>

A feeling of sadness about a mistake that you have made, and a wish that it could have been different and better





2-Human intelligence & decision-making

Evidence based Medicine Overview & current drift

Limitations and errors Cognitive illusion assessment Deciphering errors in prostate cancer policies

Thinking and Decision-Making



Bat + ball = 10\$ Bat costs 9\$ more than the ball



How much does the ball cost ?

Bat + ball = 10\$ Bat costs 9\$ more than the ball



How much does the ball cost ?



Typology of Reasoning



Logical reasoning: Asymmetric (Causal)

Aristotle (384-322 BC JC)



Syllogism

Theory Cancer \rightarrow PSA

Men

mortal \rightarrow X is a man

Observation

Cancer

 \rightarrow X is mortal

Conclusion

If & If \rightarrow Then

 $Obs \rightarrow PSA$

Deduction

Cancer $\rightarrow PSA\uparrow$ \rightarrow X has a Cancer \rightarrow X has PSA \uparrow



Evidence Based Medicine

The aim of EBM is to integrate the best available scientific information to guide decision-making about clinical management

Statistics



The foundations of the Evidence Based Medicine

Translational & clinical research

Discovery

Inference

by Induction



Decision problem

Evidence Based Medicine drift

Scientific Theory Mechanistic

Statistics Clinical Data / p value

Clinical research



Randomized trial ≠ scientific relevance

Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials

Gordon C S Smith, Jill P Pell

BMJ december 2003 Vol 327

Abstract

Objectives To determine whether parachutes are effective in preventing major trauma related to gravitational challenge.

Design Systematic review of randomised controlled trials.

Data sources: Medline, Web of Science, Embase, and the Cochrane Library databases; appropriate internet sites and citation lists.

Study selection: Studies showing the effects of using a parachute during free fall.

Main outcome measure Death or major trauma, defined as an injury severity score > 15.

Results We were unable to identify any randomised controlled trials of parachute intervention.

Conclusions As with many interventions intended to prevent ill health, the effectiveness of parachutes has not been subjected to rigorous evaluation by using randomised controlled trials. Advocates of evidence based medicine have criticised the adoption of interventions evaluated by using only observational data. We think that everyone might benefit if the most radical protagonists of evidence based medicine accepted intervention was a fabric device, secured by strings to a harness worn by the participant and released (either automatically or manually) during free fall with the purpose of limiting the rate of descent. We excluded studies that had no control group.

Definition of outcomes

The major outcomes studied were death or major trauma, defined as an injury severity score greater than 15.°

Meta-analysis

Our statistical apprach was to assess outcomes in parachute and control groups by odds ratios and quantified the precision of estimates by 95% confidence intervals. We chose the Mantel-Haenszel test to assess heterogeneity, and sensitivity and subgroup analyses and fixed effects weighted regression techniques to explore causes of heterogeneity. We selected a funnel plot to assess publication bias visually and Egger's and Begg's tests to test it quantitatively. Stata software, version 7.0, was the tool for all statistical analyses.

Results



p value <5% ≠ Scientific Reality

Empirical Article

Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results

Advances in Methods and Practices in Psychological Science 2018, Vol. 1(3) 337–356 © The Author(s) 2018



Article reuse guidelines: sagepub.com/journals-permissions DOI: 10.1177/2515245917747646

Same Data, Different Conclusions

Twenty-nine research teams were given the same set of soccer data and asked to determine if referees are more likely to give red cards to dark-skinned players. Each team used a different statistical method, and each found a different relationship between skin color and red cards.







Clinical research

Geocentrism

Heliocentrism



CLINICAL INTUITION

Learning

Inference

by Abduction

Clinical reasoning

General law / theory → individual case

Establishing a diagnosis

Theory

Cancer \rightarrow PSA

Observation

Cancer Abduction

Conclusion

Obs → PSA

Charles S. Peirce (1839-1914) Semeiology

Cognitive illusions: the limits of Human Intelligence

Do you think Be the best?



Cognitive illusions & Unconscious bias Perception

Perception Deduction Calculation



Inference Observational

Just the reality

Check all information without a subjective lecture

you have 10 seconds to count the number of «F»

FINISHED FILES ARE THE RESULT OF YEARS OF SCIENTIFIC STUDY COMBINED WITH THE EXPERIENCE OF YEARS

 Rechercher		?	\times
Rechercher : F Respecter la casse		<u>S</u> uivant	
		Fermer	

The number of «F»





NIH Public Access Author Manuscript

Psychol Sci. Author manuscript; available in PMC 2014 September 01.

Published in final edited form as: *Psychol Sci.* 2013 September ; 24(9): 1848–1853. doi:10.1177/0956797613479386.

"The invisible gorilla strikes again: Sustained inattentional blindness in expert observers"

Trafton Drew, Melissa L. H. Vo, and Jeremy M. Wolfe

Trafton Drew: TraftonDrew@gmail.com

Abstract

We like to think that we would notice the occurrence of an unexpected yet salient event in our world. However, we know that people often miss such events if they are engaged in a different task, a phenomenon known as "inattentional blindness." Still, these demonstrations typically involve naïve observers engaged in an unfamiliar task. What about expert searchers who have spent years honing their ability to detect small abnormalities in specific types of image? We asked 24 radiologists to perform a familiar lung nodule detection task. A gorilla, 48 times larger than the average nodule, was inserted in the last case. 83% of radiologists did not see the gorilla. Eye-tracking revealed that the majority of the those who missed the gorilla looked directly at the location of the gorilla. Even expert searchers, operating in their domain of expertise, are vulnerable to inattentional blindness.







In an up-to \$200M Acquisition by Nanox, Zebra Medical Vision Brings Its AI to Reimagine Radiology Globally

..............

OX

Comparison of physician and artificial intelligence-based symptom checker diagnostic accuracy

Markus Gräf^{1,2} · Johannes Knitza^{1,2,3} · Jan Leipe⁴ · Martin Krusche⁵ · Martin Welcker⁶ · Sebastian Kuhn⁷ · Johanna Mucke⁸ · Axel J. Hueber^{1,9} · Johannes Hornig¹⁰ · Philipp Klemm¹¹ · Stefan Kleinert¹² · Peer Aries¹³ · Nicolas Vuillerme^{3,14,15} · David Simon^{1,2} · Arnd Kleyer^{1,2} · Georg Schett^{1,2} · Johanna Callhoff^{16,17}

Received: 19 July 2022 / Accepted: 29 August 2022 / Published online: 10 September 2022 © The Author(s) 2022

Abstract

Symptom checkers are increasingly used to assess new symptoms and navigate the health care system. The aim of this study was to compare the accuracy of an artificial intelligence (AI)-based symptom checker (Ada) and physicians regarding the presence/absence of an inflammatory rheumatic disease (IRD). In this survey study, German-speaking physicians with prior rheumatology working experience were asked to determine IRD presence/absence and suggest diagnoses for 20 different real-world patient vignettes, which included only basic health and symptom-related medical history. IRD detection rate and suggested diagnoses of participants and Ada were compared to the gold standard, the final rheumatologists' diagnosis, reported on the discharge summary report. A total of 132 vignettes were completed by 33 physicians (mean rheumatology working experience 8.8 (SD 7.1) years). Ada's diagnostic accuracy (IRD) was significantly higher compared to physicians (70 vs 54%, p = 0.002) according to top diagnosis. Ada listed the correct diagnosis more often compared to physicians (54 vs 32%, p < 0.001) as top diagnosis as well as among the top 3 diagnoses (59 vs 42%, p < 0.001). Work experience was not related to suggesting the correct diagnosis or IRD status. Confined to basic health and symptom-related medical history, the diagnostic accuracy of physicians was lower compared to an AI-based symptom checker. These results highlight the potential of using symptom checkers early during the patient journey and importance of access to complete and sufficient patient information to establish a correct diagnosis.

Deductive inference

Peter Wason 1966

TESTING COGNITIVE BIASES BY PETER WASON



Which 2 cards should you turn over to determine the veracity of the rule

□ Card 1 □ Card 2 □ Card 3



The most common mistake: flip red-card and forget green-card

We trend to look more for a verification than a refutation of a rule (here, forget the green-card)
Testing for the Presence of Positive-Outcome Bias in Peer Review

A Randomized Controlled Trial

Gwendolyn B. Emerson, MD; Winston J. Warme, MD; Fredric M. Wolf, PhD; James D. Heckman, MD; Richard A. Brand, MD; Seth S. Leopold, MD

Confirmation

Background: If positive-outcome bias exists, it threa ens the integrity of evidence-based medicine.

modus ponens

Methods: We sought to determine whether positive outcome bias is present during peer review by testing whether peer reviewers would (1) recommend publica tion of a "positive" version of a fabricated manuscript over an otherwise identical "no-difference" version, (2) identify more purposefully placed errors in the no-difference version, and (3) rate the "Methods" section in the positive version more highly than the identical "Methods" section in the no-difference version. Every of a well-designed randomized controlled triar that differed only in the direction of the finding of the principal study end point were submitted for peer review to 2 journals in 2008-2009. Of 238 reviewers for The Journal of Bone and Joint Surgery and Clinical Orthopaedics and Related Research randomly allocated to review either a posi-

Refutation

tive or a no-difference version of the manuscript, 210 returned reviews.

modus tollens, remore likely to recommend the positive version of the test manuscript for publication than the no-difference version (97.3% vs 80.0%, *P* < .001). Reviewers detected more errors in the no-difference version than in the positive version (0.85 vs 0.41, P < .001). Reviewers awarded higher methods scores to the positive manuscript than to the no-difference manuscript (8.24 vs 7.53, P=.005), although the "Methods" sections in the 2 versions were identical.

the rule ositive-outcome bias was present during peer review. A fabricated manuscript with a positive outcome was more likely to be recommended for publication than was an otherwise identical no-difference manuscript.

Arch Intern Med. 2010:170(21):1934-1939

Bayesian inference

$$P(A|B) = rac{P(B|A)P(A)}{P(B)}$$

General population : (Prevalence) % sick people →Cancer 10% →Other 60%

Diagnosis' Test accuracy:	Cancer	Other
Sensitivity	90%	50%
Specificity	90%	50%

Odds (PPV) to be sick if the diagnosis test is positive

PPV	Cancer	Other
90%		
80%		
70%		
60%	ē	
50%		



	sensitivity x prevalence
	[sensitivity x prevalence] + [(1–specificity) x (1–prevalence)]
	NPV = specificity x (1–prevalence)
ľ	[(1-sensitivity) x prevalence] + [specificity x (1-prevalence)]

General population : (Prevalence) % sick people → Cancer 10% → Other 60%

Diagnosis' Test accuracy:	Cancer	Other
Sensitivity	90%	50%
Specificity	90%	50%

Odds (PPV) to be sick if the diagnosis test is positive

PPV	Cancer	Other
90%		
80%		
70%		
60%		60%
50%	50%	

JCI The Journal of Clinical Investigation

Published by The American Society for Clinical Investigation | Founded 1908

There's a goat behind door number 3: from Monty Hall to medicine

David J. Friedman,¹ Laurence A. Turka,^{1,2} and Simon C. Robson¹



Cognitive illusions could be just a game

But, unfortunately

They distort our health policies

ddre

.....

Expert reports and Health autority guidelines overide causal reasoning

Prostate cancer is not recommended to screen prostate cancer in asymptomatic men but should be use in a symptomatic man

PSA assay is not recommended to screen prostate cancer in the general population or in populations of men considered more at risk

The use of 5α reductase inhibitors increases the risk of prostate cancer worse outcome





Berkson Joseph 1899 -1982

Survival of the Berkson' selection bias (1946)

	E^+		E	,—
	D^+	D^{-}	D^+	D^{-}
H^+	800	600	400	200
H^{-}	200	400	600	800



- The prevalence of the disease (D) is 50% among exposed (E) and unexposed.
- 70% are hospitalized (*H*) among exposed patients (30% among non exposed)
- 60% are hospitalized among diseased patients (40% among non diseased).
- Within those hospitalized, the prevalence of the disease is 57% among exposed and 66% among unexposed patients.



Collider bias undermines our understanding of COVID-19 disease risk and severity

Gareth J. Griffith ^{1,2,4}, Tim T. Morris ^{1,2,4}, Matthew J. Tudball ^{1,2,4}, Annie Herbert^{1,2,4}, Giulia Mancano^{1,2,4}, Lindsey Pike^{1,2}, Gemma C. Sharp ^{1,2}, Jonathan Sterne², Tom M. Palmer ^{1,2}, George Davey Smith ^{1,2}, Kate Tilling ^{1,2}, Luisa Zuccolo^{1,2}, Neil M. Davies ^{1,2,3} & Gibran Hemani ^{1,2,4 ×}

Screening for prostate cancer in populations of men with risk factors



The High Authority for Health recalls that current knowledge does not allow to recommend screening for prostate
cancer by PSA assay systematically in the general population or in populations of men considered more at risk.



Screening for prostate cancer in populations of men with risk factors



Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO)

5000



Prostate cancer prevention using 5 αReductase inhibitors (5ARi)



Relative and Absolute Risk of Prostate Cancer According to Modified Gleason Score (mGS), PCPT and REDUCE Trial.

Health authorities

Increase aggressiveness of cancer

Effect of Finasteride on the Sensitivity of PSA for But..... Detecting Prostate Cancer

Ian M. Thompson, Chen Chi, Donna Pauler Ankerst, Phyllis J. Goodman, Catherine M. Tangen, Scott M. Lippman, M. Scott Lucia, Howard L. Parnes, Charles A. Coltman, Jr.

Journal of the National Cancer Institute, Vol. 98, No. 16, August 16, 2006

Background: In the Prostate Cancer Prevention Trial (PCPT), men receiving finasteride had a 24.8% lower risk of prostate cancer than men receiving placebo but a higher risk of highgrade cancer. We examined the impact of finasteride on the sensitivity and area under the receiver operating characteristic curve (AUC) of prostate-specific antigen (PSA) for detecting prostate cancer. *Methods:* We studied men in the placebo PSA cutoffs matched by specificity. *Conclusions:* PSA had statistically significantly better sensitivity and AUC for detecting prostate cancer in the finasteride arm of the PCPT than in the placebo arm. This bias would be expected to contribute to greater detection of all grades of prostate cancer with finasteride. [J Natl Cancer Inst 2006;98:1128–33]





Fig. 1. Receiver operating characteristic (ROC) curves for prostate-specific antigen detection of all prostate cancer and high-grade prostate cancer. Left, ROC curves for all prostate cancer; middle, ROC curves for Gleason grade 7 or higher prostate cancer; right, ROC curves for Gleason grade 8 or higher prostate cancer. Solid line = placebo group; dashed line = finasteride group. *P* values for difference between placebo and finasteride groups [from test of DeLong et al. (3)] are <.001 for all prostate cancer, .003 for Gleason grade 7 or higher prostate cancer, and .071 for Gleason grade 8 or higher prostate cancer.

=5ARi drug

The Will Rogers Phenomenon — Stage Migration and New Diagnostic Techniques as a Source of Misleading Statistics for Survival in Cancer

Alvan R. Feinstein, M.D., Daniel M. Sosin, M.D., and Carolyn K. Wells, M.P.H.







BACK TO ARTICLE

January 24, 2019 N Engl J Med 2019; 380:393-394

This content requires an NEJM.org account. Create a free account now. Already have an account? Sign in.

CORRESPONDENCE FREE PREVIEW

Long-Term Effects of Finasteride on Prostate Cancer Mortality



Decision-making engine architecture The « Integratome » experience in preventive medicine

Interactive computer-based decision-making systems Bringing psychology closer to algorithms

3-Algorithmic medicine and decision-making





+

+

4

Positive

2

Yes

Negative or Untested

Please click on the info button

3

1

No

Reset

ł

E

H

E

E

Age at

diagnosis

PSA (ng/ml)

Hospital

BRCA

admission in

last 2 years?

Ethnic Origin

Clinical T stage

We recommend that patients use this tool in consultation with their doctor.

This tool is only for use in men without metastatic disease where conservative management and radical treatment are both options being considered.



The Best decision minimises regret



How cognitive illusions can be overcome.

System 1 operates automatically and cannot be turned off at will.

System 2 is too slow to serve as a permanent substitute for System 1

Can Artificial intelligence help?



The kind of causal inference seen in natural human thought can be "algorithmitized" to help produce human-level machine intelligence.

BY JUDEA PEARL

The Seven Tools of Causal Inference, with Reflections on Machine Learning

THE DRAMATIC SUCCESS IN machine learning has led to an explosion of artificial intelligence (AI) applications and increasing expectations for autonomous systems that exhibit human-level intelligence. These expectations have, however, met with fundamental obstacles that cut across many application areas. One such obstacle Intensive theoretical and experimental efforts toward "transfer learning," "domain adaptation," and "lifelong learning"⁴ are reflective of this obstacle.

Another obstacle is "explainability," or that "machine learning models remain mostly black boxes"26 unable to explain the reasons behind their predictions or recommendations, thus eroding users' trust and impeding diagnosis and repair; see Hutson^a and Marcus.11 A third obstacle concerns the lack of understanding of cause-effect connections. This hallmark of human cognition1023 is, in my view, a necessary (though not sufficient) ingredient for achieving human-level intelligence. This ingredient should allow computer systems to choreograph a parsimonious and modular representation of their environment, interrogate that representation, distort it through acts of imagination, and finally answer "What if?" kinds of questions. Examples include interventional questions: "What if I make it happen?" and retrospective or explanatory questions: "What if I had acted differently?" or "What if my flight had not been late?" Such guestions cannot be articulated, let alone an swered by systems that operate in purely statistical mode, as do most learning machines today. In this article, I show that all three obstacles can be overcome using causal modeling tools, in particular, causal diagrams and their associated logic. Central to the development of these tools are advances in graphical and structural models that have made counterfactuals computationally manageable and thus rendered causal reasoning aviable com-

key insights

 Data science is a two-body problem, connecting data and reality, including the forces behind the data.

Data science is the art of interpreting



	Mathematical	Reporting sys	stems Psy	chological
Questions	According to J. Pearl		According to D. Kahne	eman
What can I deduce from this by observing this? (ex: What does this sign tell me about this disease?)	Assoc calculated with n syste	iative nachine learning ems SEEING »	Automatio	
What happens if I do this? (e.g. will this treatment be effective?)	Interve estimated by ran causal Bayesi	ntional domized trials or an networks DOING »	Logic	
What would have happened if I had done that? (e.g. is the therapeutic hazard observed due to my treatment?)	Counter calculated with fur structural « I	rfactual nctional models or equations MAGINING »	Executive	



Machine-learning Deep-learning + Symbolic AI

Involves the explicit embedding of human knowledge & behavior rules into computer





Cross-KIC EIT Health : un projet européen récompensé







Exploring with XAI the factors that influence the onset of the disease to provide insight into the mechanism of the disease and its points of attack to prevent it.



Complex Networks for Precision Medicine

Most human diseases are not independent of each other, although they are often treated separately. diseases are associated with the breakdown of functional modules of relevant genetic, metabolic, and environnemental interactions. interrelationships among human diseases allow to construct a network in which two diseases are connected if they have a common genetic or environmental or functional links



Modelling preventive medicine

Holistic Medicine



			-Pro
Inte	ear	atome	Event Series
	9.		Te
		Holistic Medicine	Pe
		Nervous	Lo
Endocrine	0		— Kic
			— Ur
		Respiratory	— Ur
	14		— Os
Cardio-vascular			— Sa
		Digostivo glando	— Ga
	9	Digestive giarius	— Ga
Digestive tract	m		Co
Digestite duce	Y		Co
	6.9	Urinary system	BII
Prostate	Y		-Pa
Prostate	(Ø	Female genital	
Male genital	X		Th
		Metabolism nutrition	
DI 1/1	K		
Blood / Immune			-DI/
	3	Locomotor	M
	A	Locomotor	-Blo
Intolerance	0		— Во
intolerance	Ø		—— Ble
	3	Discomfort	

Prostate cancerRisk
Evolutive Benign Prostate Hyperplasia

Severe Erectile dysfunction

Testis cancerRisk

Pelvis inflammation

Low Urinary Tract Disorders

Kidney cancerRisk

Kidney insufficiency

Urinary StoneRisk

Urinary Tract TumorsRisk

OsteopeniaRisk

Sarcopenia

Gastric cancerRisk

Gastritis

Colon cancerRisk

Colitis/Diverticulosis

Biliary stones/Pancreatitis

Pancreas cancerRisk

Androgen decline

Thyroid cancerRisk

Thyroiditis/ insufficiency

LIPIDS DISORDERS

DIABETE Mellitus

Metabolic syndrome

Blood deficiency / Iron metabolism

Body composition imbalance

Bleeding disorders



Recommendations Based on your informations



4-Conclusion

•





.









•

. .

. .











Forbes

The Future Of Technology Support



Andrew Barlow Forbes Councils Member Forbes Technology Council COUNCIL POST | Membership (Fee-Based)

Sep 3, 2020, 08:50am EDT



Technology Is On The Rise, While IQ Is On The Decline



Will Conaway Forbes Councils Member Forbes Technology Council COUNCIL POST | Membership (Fee-Based)

POST WRITTEN BY

Will Conaway
Olivier Cussenot

o.cussenot@cerepp.org olivier.cussenot@nds.ox.ac.uk



Twitter @CussenotO linkedin.com/in/proliviercussenot



