VVSOR
BLOG

Why statisticians prefer scoring good on average over exceptional only once

Judith ter Schure   📅 3 weeks ago   💬 Add comment   👁 75 views   🕐 6 min read

How scores turn out can be surprising, whether they come from sports or medicine. In sports we like the excitement: even the best-known favourites have only one opportunity to win at the Olympic Games or in a World Championship final. But what happens when a clinical study relies only on a single measurement? In such a case we should be careful with the promises of personalized medicine and precision medicine. Medical science cannot do without repetition and averages.

Games? And who will win the Netherlands-Germany football match? A skillful prediction for the Vancouver Winter Olympics in 2010 would be: Sven Kramer, for Rio de Janeiro 2016: Annemiek van Vleuten, and for the 1974 World Cup: The Netherlands. Yet the results were: Lee Seung-Hoon, Anna van der Breggen, and Germany.

was written for the Dutch Journal of Medicine (Nederlands Tijdschrift voor Geneeskunde) and will appear digitally on April 8 at www.ntvg.nl and on April 15 in print.

In sports it is not always the best who wins. Every result contains an element of bad or good luck, arbitrariness, and coincidence. This is what makes it exciting, of course, but sports watchers do not let themselves be carried away by the outcome of a single race[1]. The more you have seen of an athlete or a club, the better you can estimate what their ranking should be. This is not an easy task, however. As PSV football coach Roger Smidt once said: "In a season you can end up much lower than what you deserve, even though you were better in many matches."[2]

A single result is often not accurate

Similar to sports, in medicine a single measurement alone often cannot provide very useful information. I have once experienced this myself as an athlete, when I took part in a small experiment where one single measurement was made. In 2013 I was on the skating track when I became short-breathed, and I couldn't stop coughing. My family doctor asked many questions, did quite some tests, and also suggested a small experiment. After running to the lab I would take some medication and we would then measure if my lungs would respond to it. Unfortunately, the two measurements, one with and one without the asthma medication, on that single winter day in 2013, differed only marginally. Was I a non-responder?

Medical scores are not always accurate. Variations appear not only between different people. They appear also in the same person, and also in the same measuring instrument. If you measure your body weight, blood pressure, or body temperature at different moments, you get different results. Now I know: this was also the case with my lung measurement. My lung's reaction to medication seemed too small, but it

averages out and the real effects become visible.

Statistics in medical research

In statistics, it is essential to take noise in measurements into consideration, especially in randomized controlled trials (RCTs). In such cases, the trial statistician is similar to a sports coach who will determine the rules in advance in order to score a new medicine or treatment. Not with the fluctuations of a World Championship match, but more as a try-out that has to provide just enough information in order to make a decision. Is it a good idea to recommend this treatment on a larger scale? We don't reject a talent after one disappointing performance, but we also do not get carried away by one positive outlier. It is then very important to know how uncertain our measurements are. Is it more like football – where luck plays quite an important role – or more like a chess championship? [3]

Measurements of lung function are very noisy, hence many RCTs have already been performed for asthma inhalers. And they bring good news. In large groups of athletes with exercise-induced asthma, the results after using an inhaler are, on average, better than when an inhaler is not used [4]. The effect is present, even if everyone had a noisy result individually. My teammates were telling me not to worry: biker Chris Froome and race skater Bart Veldkamp can sport on the top level, and they do this using such an inhalator.

Personalized/Precision medicine

We ought to be careful with classifying individuals as responders and non-responders if we have just one measurement from each individual. The outcomes of these single measurements could just happen to be too large or too small due to chance or noise. Statisticians like Stephen Senn regularly warn about such practices: "We have moved from finding highly effective treatments for most patients to trying to find expensive ones for almost nobody at all." [5]

A typical example of such a case is again the lung volume test. Stephen Senn has found a dataset that illustrates this perfectly, with data that compares two almost-identical, instead of two different asthma drugs. [5] Two times almost the same drug for the same patient: you would expect almost the same effect. When you examine the

VVSOR
BLOG

other way around, where the first result is disappointing while the second one shows an improvement. The data of the individual patients can thus create the illusion that they concern two groups of patients, non-responders for the one inhalator, and non-responders for the other inhalator. And it's easy to mistake ourselves on that: these two inhalers were almost the same after all!

Better good on average

In 2013 my family doctor listened to my story from ice skating, about the shortness of breath and the coughing. He took the result of that single measurement with a grain of salt, and made the diagnosis of exercise-induced asthma. Consequently, he prescribed based on the clinical guidelines: the RCTs had shown that exercise-induced asthma improves, on average, when an inhaler is used. I am happy that my doctor didn't think I was exceptional.

Special thanks go to Michiel de Hoog, Mark van de Wiel, Dirk van der Hoeven and Gerard Sierksma for their input on good sports examples, and to Anne Top for her great help with the article and her input on asthma diagnoses. Any sports or medicine mistakes still in this blog post are Judith's.

Notes

[1] Van de Wiel, M. Eerlijke sport is vaak minder leuk. Blog Vereniging voor Statistiek en Operations Research 4 november 2021. https://blog.vvsor.nl/2021/11/eerlijke-sport-is-vaak-minder-leuk/

[2] De Hoog, M. Succesvolle mensen geven zelden toe hoeveel geluk ze hebben gehad. Deze trainer doet dat wel. De Correspondent 4 juli 2020. https://decorrespondent.nl /11221/succesvolle-mensen-geven-zelden-toe-hoeveel-geluk-ze-hebben-gehad-deze-trainer-doet-dat-wel/475882172381-7419c3f3

[3] Fong, J. Why it's so much harder to predict winners in hockey than basketball. A statistical look at luck and skill in sports. Vox 5 juni 2017. www.vox.com/videos /2017/6/5/15740632/luck-skill-sports

[4] Bonini M, Di Mambro C, Calderon MA, Compalati E, Schünemann H, Durham S,

[5] Senn, S. Personalised medicine a sceptical view. Slideshare. 2019. https://www.slideshare.net/StephenSenn1/personalised-medicine-a-sceptical-view.

fouten   gezondheid   kans   sport

1 comment on PubPeer (by: Harri Hemilä)

Judith ter Schure

Wat krijg je als een bergsporter besluit statisticus te worden? Micromorts, survival analyse en een hoop gemijmer over de rol van risico, pech en geluk in het leven. En zoals bij wel meer statistici: een grote liefde voor knap ontworpen experimenten, strak bijgehouden tijdsreeksen en open data met al hun beperkingen.

VIEW ALL POSTS

Add comment

Comment

VVSOR
BLOG

Name *

Email *

Website

☐ Save my name, email, and website in this browser for the next time I comment.

SUBMIT COMMENT

You may also like

ENGLISH

When statistics lie, Statistics can unravel the lie

Were the results of Mendel's famous experiments on pea hereditary too good to be true? Is the efficacy of the Sputnik vaccine really thát homogeneous across age groups? When we don't trust the numbers - i.e. statistics ...

Mark van de Wiel   📅 2 weeks ago

An Italian CSI drama

This is the case of Daniela Poggiali, a nurse suspected of having killed around 90 patients in a two-year killing spree. Thrilling longread about social media, a broken legal system, and Micky Mouse statistics.

ENGLISH

Ski touring vs the daily commute: how risky is the backcountry?

What is the risk of unassisted backcountry ski touring? I decided to clarify the concept by comparing backcountry ski trips to daily commutes.

Judith ter Schure   📅 4 months ago

Recent posts   ‹ ›

VVSOR
BLOG

Q ≡

Mark van de Wiel

📅 2 weeks ago

Kansenongelijkheid in het onderwijs voor de niet-gemiddelde leerling

Carolien de Kovel

📅 1 month ago

Moet je altijd bij het begin beginnen?

Sanne Willems

📅 2 months ago

Type here to search...

SEARCH

Follow us

🐦