#### **Personalization in Statistics**

Giorgio Bedogni

www.giorgiobedogni.it

#### Question

• Where is the person in my (or others') statistical model?

Something I continuously ask myself as a physician (internist) who serendipitously became an applied statistician

# Outline

- Is mathematics the language of nature?
- What is a person?
- What is a model?
- Where is the person in the model?
- What should I take home?

# Outline

- Is mathematics the language of nature?
- What is a person?
- What is a model?
- Where is the person in the model?
- What should I take home?

#### Is mathematics the language of nature?

 "It was Galileo who came up with the saying that mathematics is the language of nature. <u>That is a lovely analogy, but nothing</u> <u>more</u>. "Language" in that sense is not a means of communication as we understand the word in the context of interpersonal expression."

Artstein Z. Mathematics and the real world. Prometheus Books; 2014.

On mathematics as a <u>social enterprise</u> see also: Hersh R. What is mathematics, really. Oxford University Press; 2009.

# Outline

- Is mathematics the language of nature?
- What is a person?
- What is a model?
- Where is the person in the model?
- What should I take home?

#### What is a person?

- Person: "a human being regarded as an individual"
- Individual: "a single human being as distinct from a group"

https://en.oxforddictionaries.com

#### What is a person?

- A "person" has a face and can speak with you <u>https://it.wikipedia.org/wiki/Persona\_(filosofia)</u>
- The "individual" mentioned in the guidelines does not have a face and cannot speak with you

#### Is evidence-based Medicine person-centered? Yes, it is!



Haynes RB et al. ACP Journal Club 2002;136:A11.

More on the person-centerdness of EBM <u>https://bit.ly/2J5zyDQ</u> (Giorgio's 2017 course) See also <u>https://bit.ly/2kodOVP</u> (Stephen Senn and Deborah Mayo)

www.giorgiobedogni.it

# Outline

- Is mathematics the language of nature?
- What is a person?
- What is a model?
- Where is the person in the model?
- What should I take home?

"Every model is ultimately the expression of one thing we hope to understand in terms of another thing we do understand"

Weinberg G. An Introduction to General Systems Thinking; 2015 <u>https://leanpub.com/u/jerryweinberg</u>.

 "The value of a model is that often it suggests a simple <u>summary</u> of the data in terms of the major systematic effects together with a <u>summary</u> of the nature or magnitude of random variations"

McCullagh P, Nelder JA. Generalized linear models. London: Chapman and Hall; 1989.

 "Analysis interpretation depends on <u>contextual judgements</u> about how reality is mapped onto the model and how the formal analysis results are mapped back to reality"

Figure 2017;32:3-20.

• "All models are wrong; some models are useful"

Box G *et al.* Statistics for experimenters: design, innovation, and discovery. Wiley-Interscience; 2005.

Box G. An accidental statistician [autobiography]. Wiley; 2013.

• "<u>You can never prove</u> that your model is the actual one or the only possible one"

De Bono E. Teach yourself to think. Penguin; 2017.

(I started systematically reading De Bono because he was highly estimeed by George Box. Box was right...)

# Outline

- Is mathematics the language of nature?
- What is a person?
- What is a model?
- Where is the person in the model?
- What should I take home?

# Where is the person in the model? Jakob Bernoulli on "personalized prediction"\*

 "Thus, if it is asked, <u>in general</u>, how much more probable is it for a 20-year-old youth to outlive an aged man of 60 rather than the other way round, we have nothing to take into consideration other than the distinction between the generations and ages"

Bernoulli J, Ars Conjectandi (1713) quoted by Raper S. Significance. 2017;14:12-17.

\*Much before the ideas of probability, average, prediction, etc. were invented.

# Where is the person in the model? Jakob Bernoulli on "personalized prediction"

 "But if the question concerns two definite persons, the youth <u>Petrus and the old man Paulus</u>, we also ought to pay attention to their complexion, and to the care that each of them takes over his health. Because if Petrus is in poor health, indulges in passion, and lives intemperately, Paulus, although much older, may still hope, with every reason, to live longer."

Bernoulli J, Ars Conjectandi (1713) quoted by Raper S. Significance. 2017;14:12-17.

#### Where is the person in the model? The difficult birth of the average

 "And if combining observations on a single object was considered radical, doing the same for measurements made <u>on many different objects was almost unthinkable</u>"

Raper S. Significance. 2017;14:12-17.

See also: Stigler SM. The Seven Pillars of Statistical Wisdom. Harvard University Press; 2016. (The "first pillar" is aggregation).

 Carl Gauss (1777-1855) - <u>Provided that there is no</u> <u>measurement error</u>, an average measurement is as close to a measurement's true value as one could ever hope to get.

If you think that errors have to be normally distributed, please read: Bailey DC. Royal Society Open Science. 2017;4:160600.

If you think that anything can be normally distributed, please read: Geary R. Testing for Normality. Biometrika. 1947;34:209-242.

- Adolphe Quetelet (1796-1874) The average person represents the <u>true human being</u> while the single individual is synonymous with error
- (It won't escape you that a value judgement is being used here to map Gauss' theorem onto the real word...)

- Francis Galton (1822-1911) You are "mediocre" if you are near the average, "eminent" if you are far above the average, and "imbecile" if you are far below the average.
- (Again a value judgement... Galton coined the term "eugenics" in 1883)

• Karl Pearson (1857-1936) – The only thing of interest is the distribution and its moments

More on Pearson's "revolutionary idea" here <u>https://bit.ly/2lzH9ee</u> [Giorgio's 2017 course].

For a partisan but very well written history of the average: see Rose T. The end of average: how we succeed in a world that values sameness. Penguin; 2017.

# Where is the person in the model? S.J. Gould – the median is not the message

- "When I learned about the 8 month median [survival time of the abdominal mesothelioma I was diagnosed], my first intellettual reaction was: fine, half the people will live longer"
- "Another technical point added even more solace... The distribution was right-skewed, with a long tail (however small) that extended for several years above the 8 month median"

https://bit.ly/2zayNob

# Where is the person in the model? S.J. Gould – the median is not the message

- Gould died <u>20 years after</u> the diagnosis of abdominal mesothelioma
- "Variation is the hard reality, not a set of imperfect measures of central tendency. Means and medians are the abstractions."

https://bit.ly/2zayNob

# Where is the person in the model? P. Kalanithi, MD - "When breath becomes air"

- "Those <u>apocryphal doctors</u> who gave specific numbers ("The doctor told me I had six months to live"): Who were they, I wondered, and who taught them statistics?"
- "Weren't the numbers just the number? Had we all given the "hope" that every patient was above average?"
- Paul Kalanithi was not as lucky as Gould but he had the time to see the birth of his child and to write a wonderful book that every physician should read

https://en.wikipedia.org/wiki/When\_Breath\_Becomes\_Air

www.giorgiobedogni.it

# Where is the person in the model? How can we put the person into the model?

- By using underused study designs
- By using new study designs

#### Intermezzo

# Under which inferential framework?

 "Both frequency (error) calibration and Bayesian (betting) coherency are unattainable ideals in health and social research. In practice, the data models used by <u>both</u> methodologies <u>fall far short</u> of capturing all sources of error and uncertainty. Thus frequentist and Bayesian methods are simply alternative ways of looking at models and data, and are <u>more complementary than competitive</u>"

Greenland S et al. Annu Rev Public Health. 2015;36:89-108. On "principled analyses" see Terence's Stuff <u>https://bit.ly/2GLpsmg</u> See also <u>https://errorstatistics.com/</u>

# Where is the person in the model? How can we put the person into the model?

- By using underused study designs
- By using new study designs

# Using underused designs (✓) RCT designs and variation components

Randomized controlled trial	Identifiable components of variation
Parallel group	Between Treatments
✓ Classical cross-over	Between Treatments Between Patients
✓ Repeated period cross-over	Between Treatments Between Patients Patient by treatment interaction
√ N-of-1	Within patient (n = 1!)

Senn S Stat Med. 2016;35:966-977 (repated cross-over trial).

Vohra S *et al. BMJ*. 2015;350:h1738 (CONSORT for n-of-1-trial); erratum in BMJ. 2016;355:i5381.

#### Using underused designs Problems

- "Absorbing" outcomes, e.g. death, cannot be studied.
- Suboptimal attrition rates for n-of-1 RCT

e.g. Ioannidis JPA at al. Hum Mol Genet. 2018 <u>but see</u> Sackett DL. Clin Trials. 2011;8:350-352 on how this can be avoided.

# Where is the person in the model? Using new study designs

- By using underused study designs
- By using new study designs

#### Using new study designs Generalized linear dynamics as example

MEASUREMENT, 2(4), 201–218 Copyright © 2004, Lawrence Erlbaum Associates, Inc.

#### FOCUS ARTICLE

A Manifesto on Psychology as Idiographic Science: Bringing the Person Back Into Scientific Psychology, This Time Forever

> Peter C. M. Molenaar Department of Psychology University of Amsterdam

www.giorgiobedogni.it

# Using new study designs Generalized linear dynamics as example

- "What is lacking is <u>the scientific study of the individual</u> in his or her structure of intra-individual variation for it own sake"
- "Standard factor analysis of inter-individual variation appears to be quite insensibile to the presence of substantial heterogeneity"
- (Repeated measures is the key here as it is for the underused designs discussed before)

Molenaar PCM. Measurement 2004;2:201-218.

# Where is the person in the model? Some notes about the "quantified self"

- <u>Repeated measurements</u> of the same of the same person are the core such purported revolution <u>https://en.wikipedia.org/wiki/Quantified\_self</u>
- Some personal comments
  - We need first of all "quality data" not "big data"
  - "Small data" will often suffice if they are of good quality
  - We need to ask good questions and use good designs even if we are interested just in associations (and not causations)

#### Where is the person in the model? A caveat worth repeating

 "Analysis interpretation depends on <u>contextual judgements</u> about how reality is mapped onto the model and how the formal analysis results are mapped back to reality"

Figure 2017;32:3-20.

#### Contextual judgements applied to modeling An example

- Ricco *et al.*, to be submitted will be presented at this Congress.
- Aim: to estimate the time-related changes of AFP and PIVKA2 in patients who developed and not developed HCC (multicenter retrospective case-control proof-of-concept study).
- (Please <u>don't be impressed by statistical jargon</u>, which I have tried to keep to a minimum, but focus on contextual judgements!)

#### Contextual judgements applied to modeling 1 figure may be worth 10<sup>3</sup> words but...



www.giorgiobedogni.it

# Contextual judgements applied to modeling ... how was reality mapped onto this model?

- $Y = f(time, time^2, hcc, hcc^*time, hcc^*time^2)$  where:
  - prespecified model
  - $Y = log_{10}AFP \text{ or } log_{10}PIVKA2$
  - time = continuous (3 unequally spaced time-points)
  - hcc = discrete (0 = no; 1 = yes)
  - $f(\cdot) =$  generalized least squares (GLS)
    - Bootstrap
    - Random effect (patient)

On (the power of) the functional notation [y = f(x)] see: Weinberg G. An Introduction to General Systems Thinking; 2015 <u>https://leanpub.com/u/jerryweinberg</u>.

# Contextual judgements applied to modeling How reality will constrain modeling choices

- $Y = f(time, time^2, hcc, hcc^*time, hcc^*time^2)$  where:
  - prespecified model
  - $Y = \log_{10}AFP \text{ or } \log_{10}PIVKA2$
  - time = continuous (<u>3 unequally spaced time-points</u>)
  - hcc = discrete (0 = no; 1 = yes)
  - $f(\cdot) =$  generalized least squares (GLS)
    - Bootstrap
    - <u>Random effect (patient)</u>

# Contextual judgements applied to modeling How reality will constrain modeling choices

- $Y = f(time, time^2, hcc, hcc^*time, hcc^*time^2)$  where:
  - prespecified model
  - $Y = \log_{10}AFP \text{ or } \log_{10}PIVKA2$
  - time = continuous (<u>3 unequally spaced time-points</u>)
  - hcc = discrete (0 = no; 1 = yes)
  - $f(\cdot) =$  generalized least squares (GLS)
    - Bootstrap
    - <u>Random effect (patient) added benefit that it can</u> <u>handle missing data if they are missing at random</u>.

#### Outline

- Is mathematics the language of nature?
- What is a person?
- What is a model?
- Where is the person in the model?
- What should I take home?

#### What should I take home?\*

- Remember that "No one is average", including your patients.
- Have clear what "degree" of individual variability can be explained by a given study design.
- Be aware of the limitations and assumptions of the old or new methods you use or intend to use.

\*Which should be more properly read as: what has Giorgio taken home during all these years working on the difficult task to put numbers at the service of <u>persons</u>?

#### Be prepared to use new methods

• Unless you agree with:

PO\*: "Experts are guardians of the past and people expect them do so"

De Bono E. Teach yourself to think. Penguin; 2017.

\*In the language of lateral thinking, PO means "a provocation will follow".

#### Thank you!

www.giorgiobedogni.it