Data are not enough to reenchant public health surveillance

Prof Arnaud Chiolero MD PhD[1,2,3]

1. Population Health Laboratory (#PopHealthLab) University of Fribourg, Switzerland; 2. Observatoire valaisan de la santé, Sion; 3. School of Populaion and Global health, McGill University, Montreal

arnaud.chiolero@unifr.ch
ETH, Zurich, June 2, 2021

#Pop Health Lab

Plan of the presentation

- What is public health surveillance

- Surveillance in the age of (big) data science

- Data are not enough
  - Surveillance bias
  - Infodemic

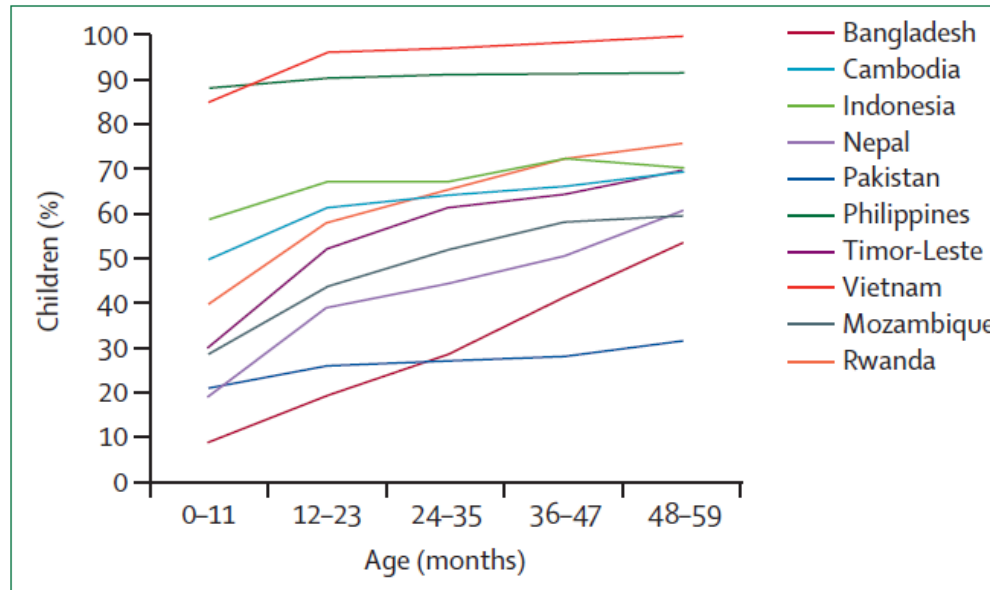- Toward evidence-based and data-informed public health

What is public health surveillance

To make people count, we first need to be able to count people

said the late WHO Director-General Lee Jong-wook in an address to WHO staff a decade ago.

Lancet 2013; 382:2040

Counting people

Figure: Percentage of children aged less than 5 years whose births are registered by age (months)
Data from the United Nations Children's Fund.[14]

AbouZahr C, et al . Lancet 2015

More data needed!
But is that enough?

What is public health surveillance

- Public health surveillance is the
    - **ongoing and systematic**
    - **collection, analysis, and interpretation** of data,
    - closely integrated with the timely **dissemination** of these data to those responsible for **preventing and controlling disease and injury** [Thacker and Berkelman 1988]

- To provide **information useful for decision and action in public health**

What is public health surveillance

- Surveillance ≠ research
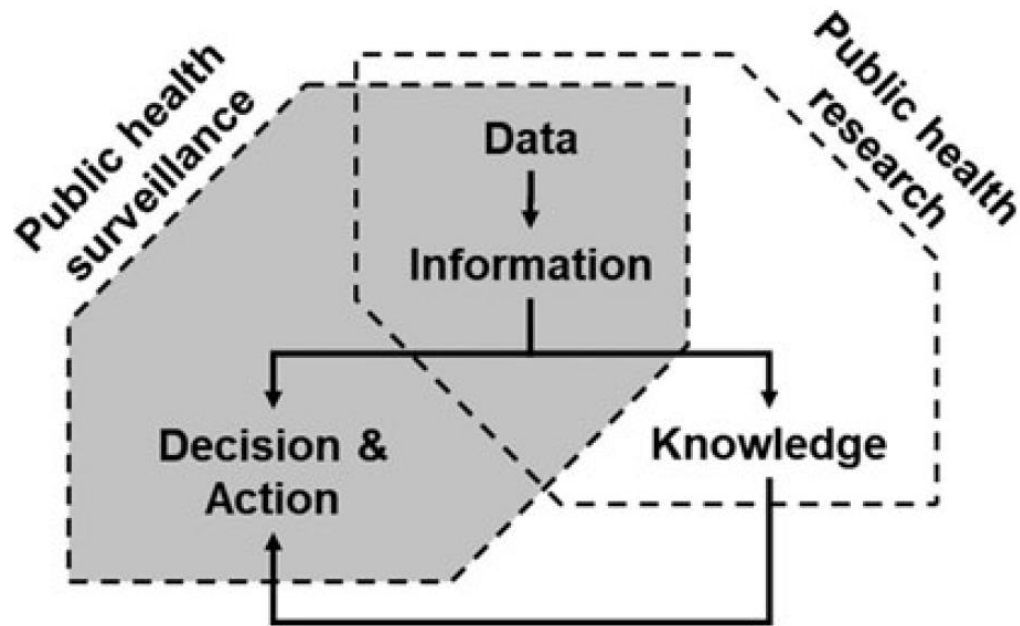
Chiolero & Buckeridge. Glossary for public health surveillance in the age of data science JECH 2020
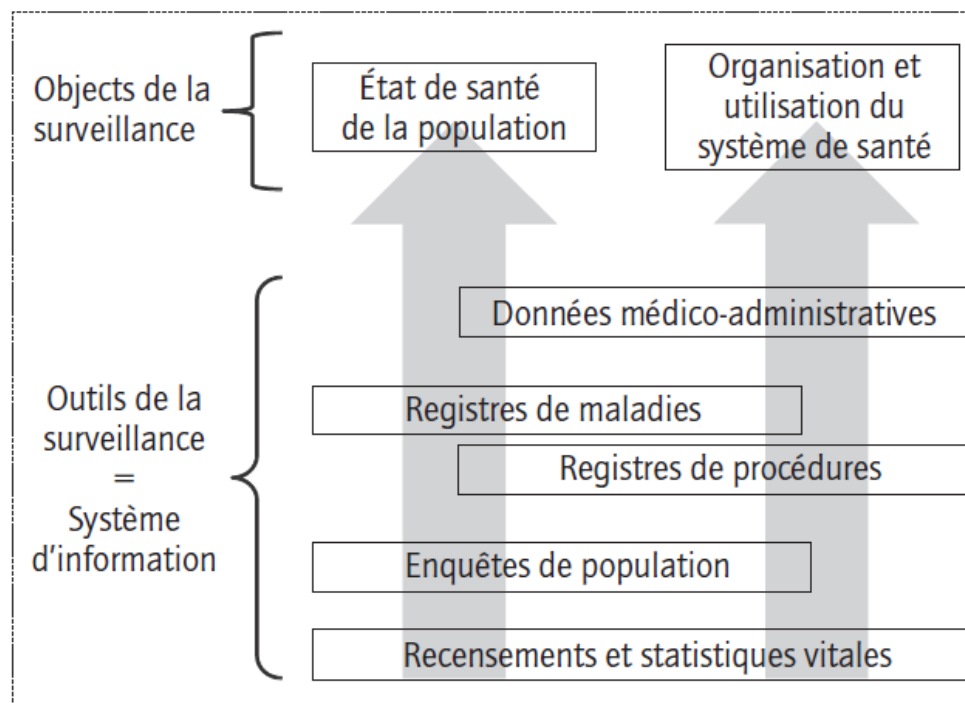
We need a system

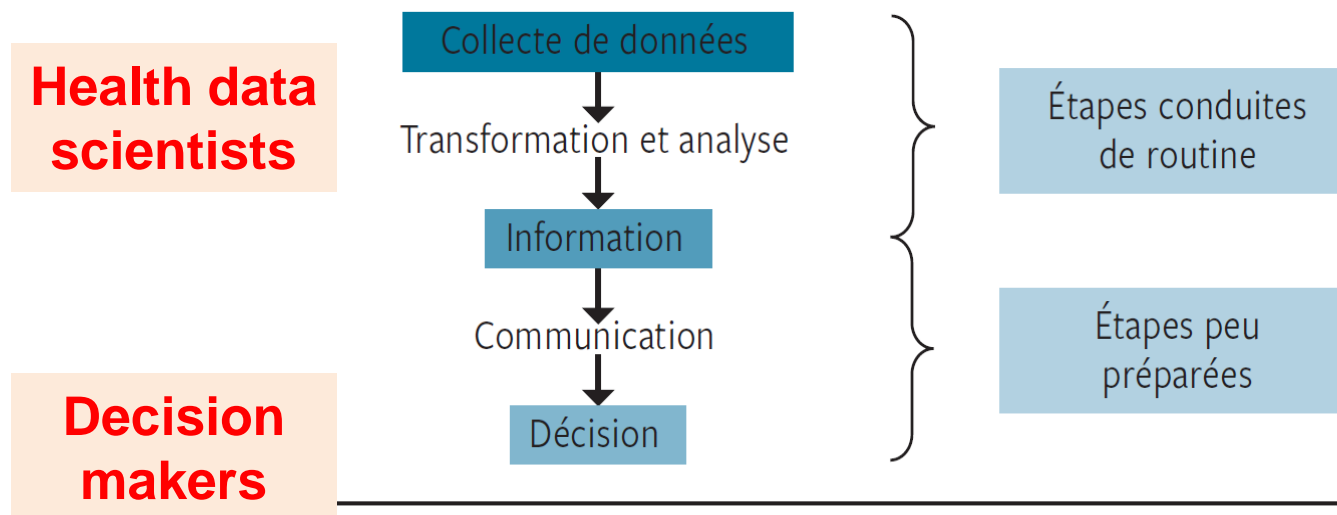Figure 1 : Objets et outils de la surveillance sanitaire

Chiolero A, Paccaud F, Fornerod L. Comment faire de la surveillance sanitaire ? L'exemple de l'Observatoire valaisan de la santé en Suisse. Santé Publique 2014.

We need a process

FIG 1 — Étapes dans le traitement des données pour la surveillance sanitaire

La figure illustre le processus de la collecte des données à la prise de décision.[2] Les étapes qui vont de la collecte à la production d'information sont souvent conduites de routine. Les étapes qui rendent les informations utilisables pour la prise de décision sont souvent insuffisamment préparées.

Health data scientists

Collecte de données → Transformation et analyse → Information → Communication → Décision

Étapes conduites de routine

Étapes peu préparées

Decision makers

Surveillance in the age of (big) data science

Surveillance in the age of data science

- rapid development of data science, e.g., big data and artificial intelligence (AI)
- exponential growth of accessible and highly heterogeneous health-related data
- Impact on surveillance & monitoring
  - widening scope of application, increasing depth and new methods
  - opportunities and challenges

Glossary

Glossary for public health surveillance in the age of data science

Arnaud Chiolero [1,2,3,4] David Buckeridge[4]

Chiolero A, Buckeridge D. *J Epidemiol Community Health* 2020;**0**:1–5. doi:10.1136/jech-2018-211654

Surveillance in the age of data science

- Opportunities
  - Access to new types of data, health- and non-health related, etc…
  - Health services research
  - Toward true data-informed public health

- Challenges
  - Ethics and privacy protection
  - Source populations?
  - Paradigm change in surveillance

Chiolero A, Buckeridge D. *J Epidemiol Community Health* 2020;**0**:1–5. doi:10.1136/jech-2018-211654

Let the data speak

SCIENCE : DISCOVERIES

The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

By Chris Anderson   06.23.08

Illustration: Marian Bantjes

THE PETABYTE AGE:
Sensors everywhere. Infinite storage. Clouds of processors. Our ability to capture, warehouse, and understand massive amounts of data is changing science, medicine, business, and technology. As our collection of facts and figures grows, so will the opportunity to find answers to fundamental questions. Because in the era of big data, more isn't just more. More is different.

"All models are wrong, but some are useful."

So proclaimed statistician George Box 30 years ago, and he was right. But what choice did we have? Only models, from cosmological equations to theories of human behavior, seemed to be able to consistently, if imperfectly, explain the world around us. Until now. Today companies like Google, which have grown up in an era of massively abundant data, don't have to settle for wrong models. Indeed, they don't have to settle for models at all.

www.wired.com/science/discoveries/magazine/16-07/pb_theory

Chris Hendeson, 23.8.2008

Paradigm change in surveillance (1)

- Classical process: identify the health problem → define and collect data (finite amount) → analyze data to address the problems
  - Pro: designed data, i.e., tailored for your problem [Keller 2012], information on their validity, reliability, and completeness (or its lack)
  - Cons: poor timeliness, limited scope, high cost

Paradigm change in surveillance (2)

■ **Big data/eHealth age:** all types of data collected from multiple sources without knowing exactly what you will do with these data → analyze data to identify problems and address problems

    ○ Pro: timeliness, representativeness, low cost

    ○ Cons: organic data, i.e., not tailored for your problem [Keller 2012], quality (poor, unknown), selectivity, management/storage, privacy/access [Chiolero 2020]

Data are not enough

Reimagining Public Health in the Aftermath of a Pandemic

The coronavirus disease 2019 (COVID-19) pandemic is an unprecedented challenge for society, affecting those already subject to unacceptable health inequalities and resulting in vast economic impacts. The pandemic reminds everyone of the value and necessity of public health.

Ross C. Brownson, PhD, Thomas A. Burke, PhD, Graham A. Colditz, MD, DrPH, and Jonathan M. Samet, MD, MS

AJPH REIMAGINING PUBLIC HEALTH

The worldwide pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a medical and public health emergency on an

Data Are Not Enough to Reimagine Public Health

See also Morabia, p. 1590, and the *AJPH* Reimagining Public Health section, pp. 1605–1623.

COVID-19 has revealed the weaknesses of health information systems worldwide. In many countries, included the United states, data were not missing. We have been rather overwhelmed with data, which have been highly accessible through open access repositories, like never before. From these data emerged many statistical analyses and predictions as well as comments in the media and social networks. Hence, on top of the viral epi- health surveillance and health data science. Public health surveillance is effective when it produces information useful for decision making and action.[2] It does not consist merely in gathering data, conducting analyses, and making them available. Key is to make this information useful for decision making. Of course, we need to improve the quality, scope, and completeness of data, but this would be insufficient. Many health data scientists believe that properly system and of public health, the needs of policymakers, and the population risk perception as well as the socioeconomic and political implications of their analyses.

We must have stronger public health surveillance systems if we are to succeed in the fight against COVID-19 and to manage future epidemics. **AJPH**

Arnaud Chiolero, MD, PhD
Daniela Anker, MSc

Remember why we need data

- To assess the burden of disease

- To design and evaluate public health policies

- To provide information useful for decision and action in public health

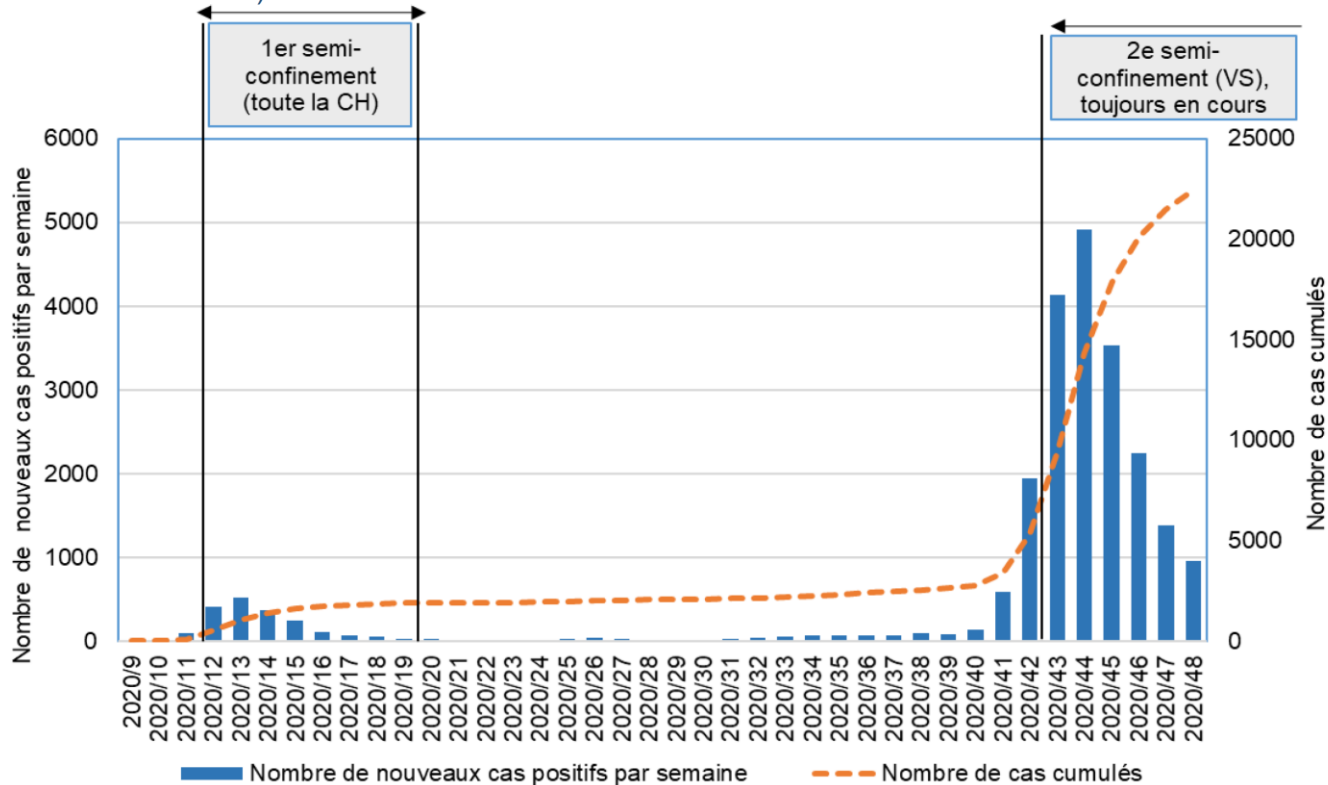# But (small as well as big) data do not speak by themselves

- No automatic production of information

Decision makers

Health data scientists

- data ≠ information ≠ knowledge

The case of Covid-19 (1)

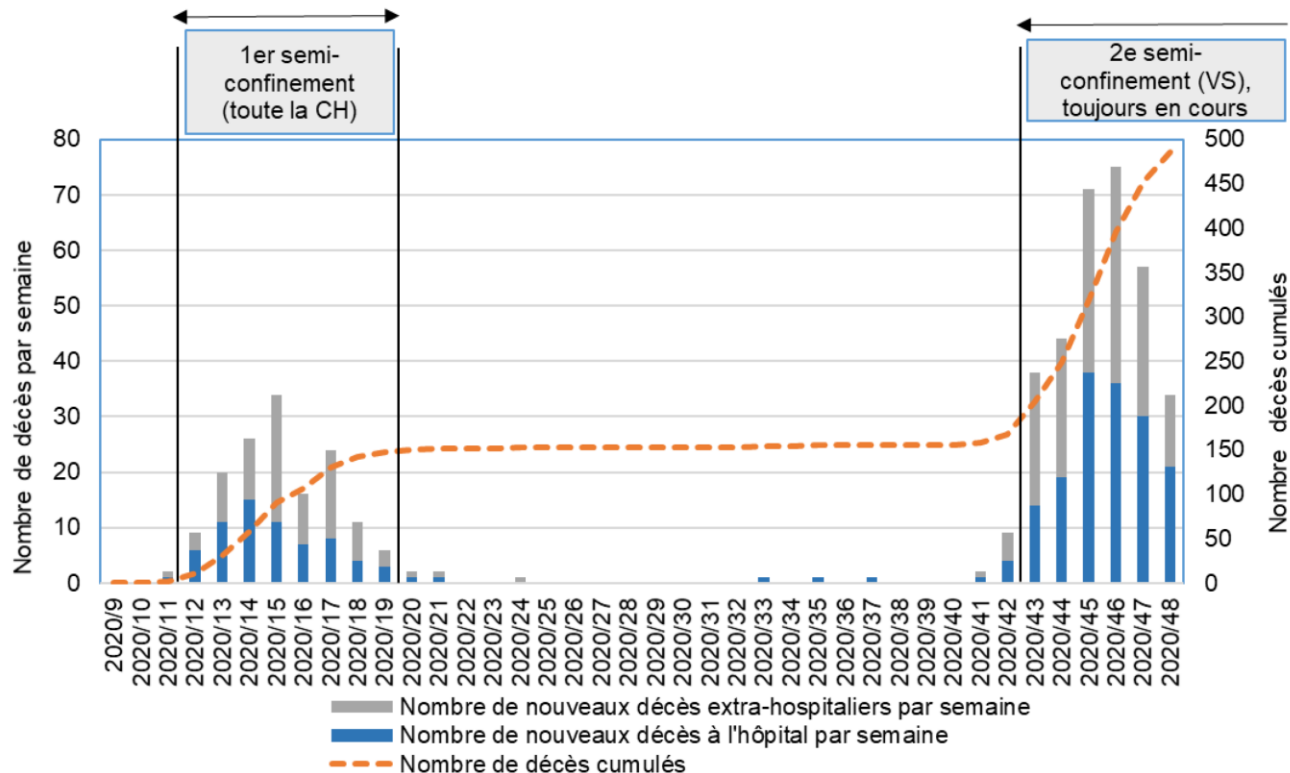F4.35 COVID-19 : Nombre de nouveaux cas déclarés par semaine en Valais (source : OVS/SSP état au 29.11.2020)

> ➢ **1st vs 2nd wave**
> ➢ **1 to 10**

www.ovs.ch

The case of Covid-19 (2)

F4.38 COVID-19 : Nombre de décès déclarés en Valais par semaine (source : OVS/SSP, état au 29.11.2020)

- ➤ **1st vs 2nd wave**
- ➤ **1 to 2**
- ➤ **Incoherence in the data?**

www.ovs.ch

Surveillance bias

- If conditions sensitive to the modality and intensity of detection activities are sought with differential intensity across populations or over time, or according to care setting

  o Differences explained by the frequency of detection - not the frequency of the conditions

  o Related concept of 'streetlight effect' if surveillance not concentrated on what matters, but on what is measurable
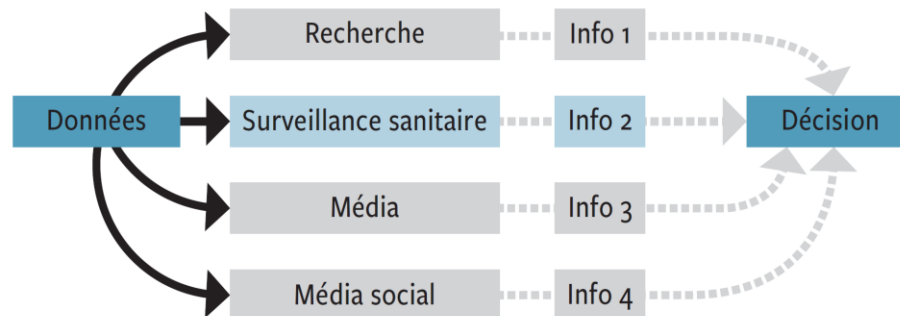
Infodemic

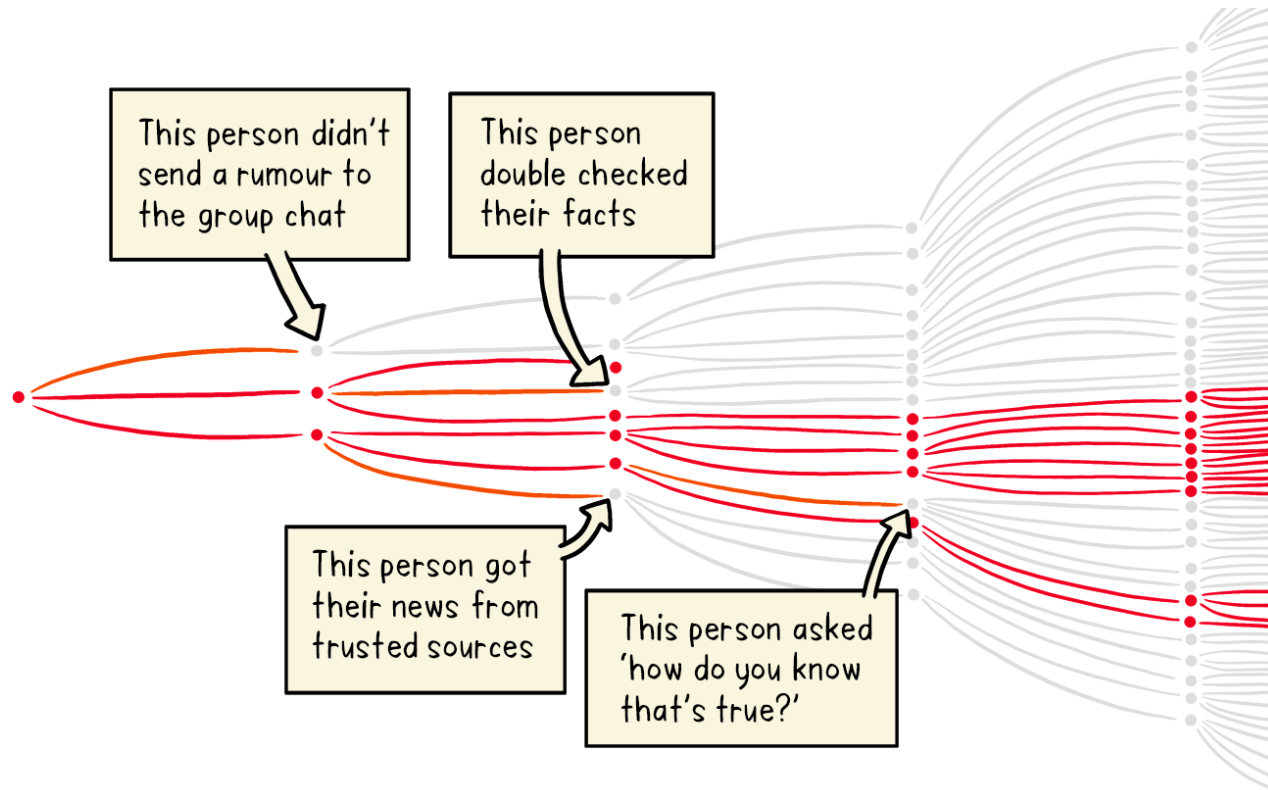- Too much data and too much information!
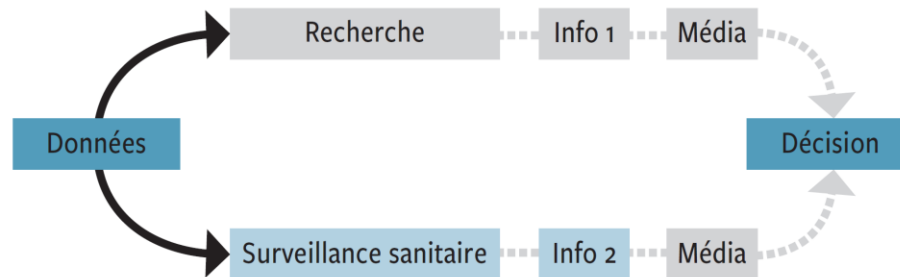
Rev Med Suisse 2021

- Information production process has changed

Infodemic

- Too much data and too much information

Source: WHO

- Information circulation has also changed!

Infodemic

Top tips for navigating the infodemic

World Health Organization

1. Assess the source:

Who shared the information with you and where did they get it from? Even if it is friends or family, you still need to vet their source.
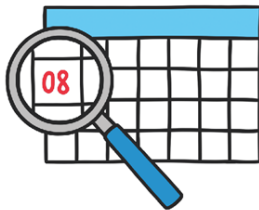
2. Go beyond headlines:

Headlines may be intentionally sensational or provocative.

3. Identify the author:

Search the author's name online to see if they are real or credible.

4. Check the date:

Is it up to date and relevant to current events? Has a headline, image or statistic been used out of context?

5. Examine the supporting evidence:

Credible stories back up their claims with facts.

6. Check your biases:

Think about whether your own biases could affect your judgment on what is or is not trustworthy.

7. Turn to fact-checkers:

Consult trusted fact-checking organizations, such as the International Fact-Checking Network and global news outlets focused on debunking misinformation.

Toward evidence-based and data-informed public health

Surveillance has failed

Disease surveillance for the COVID-19 era: time for bold changes

The COVID-19 pandemic has exposed weaknesses in disease surveillance in nearly all countries. Early identification of COVID-19 cases and clusters for rapid containment was hampered by inadequate diagnostic capacity, insufficient contact tracing, fragmented data systems, incomplete data insights for public health responders, and suboptimal governance of all these elements. Once SARS-CoV-2 became widespread, interventions to control community transmission were undermined by weak surveillance of cases and insufficient national capacity to integrate data for timely adjustment of public health measures.[1,2] Although some countries

WHO. Fifth, disease surveillance must be adequately financed.

Interpretation of disease surveillance data needs population representativeness, denominators, and historical baseline data. Civil registration and vital statistics (CRVS) systems are important for population estimation and understanding excess mortality but have historically taken years to build. Many countries that lack or have inadequate CRVS systems need to accelerate their development in alignment with the recommendations in the WHO *SCORE for Health Data Technical Package* report.[6] In the meantime, representative

Science vs politics

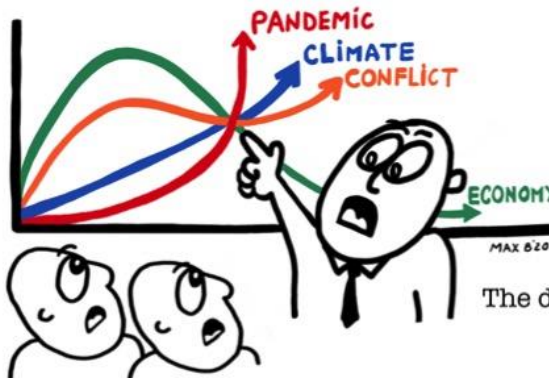## Faut-il réduire la task force scientifique au silence?

«Ich habe die Wissenschaft zu wenig hinterfragt»

...tés dès ce jeudi dans le cadre de la
...ut empêcher la task force scientifique de
...monde de la recherche est abasourdi

Bundesrat Alain Berset zeigte sich in einem SRF-Interview selbstkritisch. (Archiv)
Bild: Keystone

PANDEMIC
CLIMATE
CONFLICT
ECONOMY

MAX 820

The disasters are collaborating
better than we are!

LIVE
L REPORT
IEFING FROM CORONAVIRUS TASK FORCE
TES, GO TO NBCNEWS.COM/CORONAVIRUS
THE WHITE HOUSE

https://oxfamblogs.org/fp2p/why-confront-covid-19-with-cartoons-and-humour/; www.bluewin.ch;
www.letemps.ch https://www.businessinsider.fr/us/dr-anthony-fauci-did-a-facepalm-during-trumps-coronavirus-briefing-2020-3

From evidence to policy

- Path complex from evidence to the implementation of sound public health interventions
  - Evidence is never enough and politics more than science determines policies
- Play politics? [Chiolero Lancet 2020]
  - Fight opponents with competing interests, lobbying through interest groups, make elect leaders serving the PH agenda…
  - Scientists not trained
  - And not legitimate?

How to foster surveillance

- Health data scientists
  - Have to connect with policy makers
  - Need to be trained in risk communication
  - Have to assume their responsibility and accept that decisions are taken by others

- Policy makers & citizens
  - Need to be trained in population health data science
  - Need to put evidence-based public health at the center of policy making

In summary

- Surveillance = data → information → decision
    - ○ Surveillance ≠ research

- More and more (big) data…
    - ○ but they do not speak by themselves
    - ○ data are not enough

- We need to strengthen
    - ○ surveillance system & process
    - ○ surveillance training and culture of *both* health data scientists & decision makers

In summary

Decision makers

have to better collaborate

Health data scientists

Thank you for your interest

Prof Arnaud Chiolero MD PhD[1,2,3]

1. Population Health Laboratory (#PopHealthLab) University of Fribourg, Switzerland; 2. Observatoire valaisan de la santé, Sion; 3. School of Populaion and Global health, McGill University, Montreal
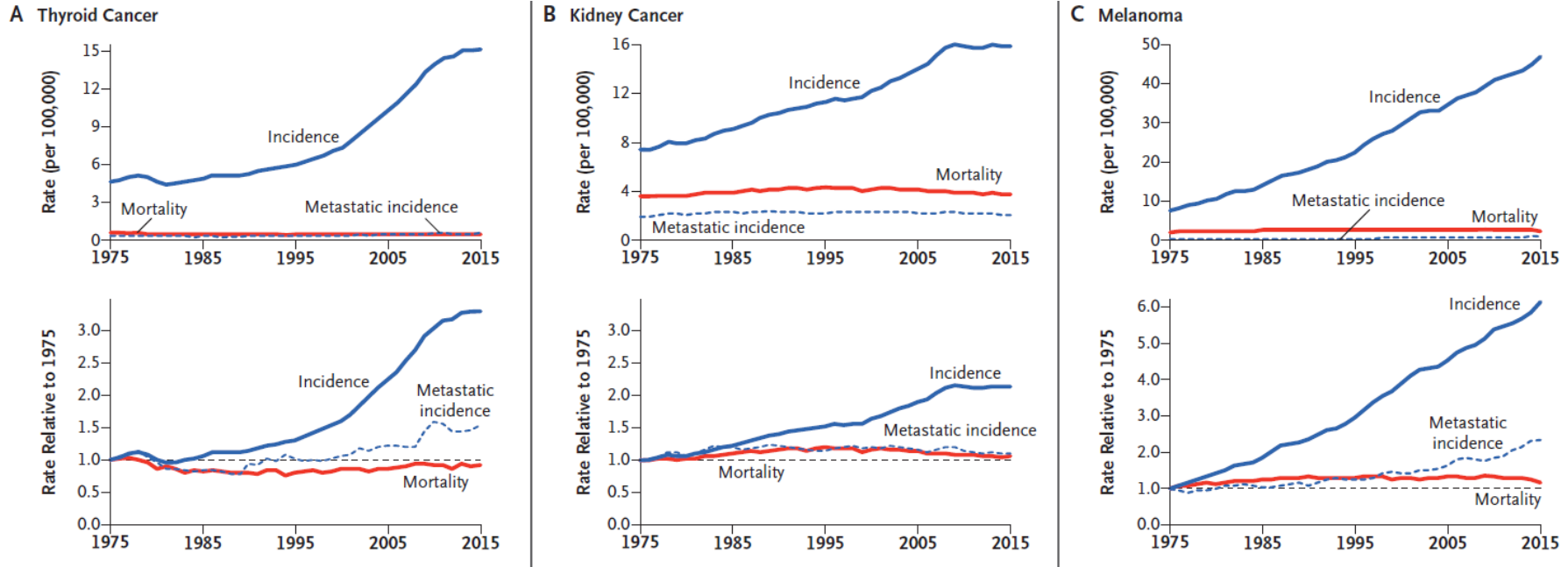
arnaud.chiolero@unifr.ch

BMed1 2021

#Pop Health Lab

UNI FR

Supplemental material

Scrutiny dependent cancer

Figure 3. Overdiagnosis with Stable True Cancer Occurrence.
Mortality is disease-specific — that is, the rate of death attributed to the specified cancer.

Welch NEJM 2020

What is public health surveillance

Data generation & collection → Data transformation & analyses → Information production → Dissemination → Decision making

Figure 2 Steps in the data processing of public health surveillance, from data generation and collection to information dissemination for decision-making.

Era of "big data"

- Surveillance activity: production of data $\rightarrow$ conversion into usable information $\rightarrow$ public health intervention

- Digital age: amount of data is increasing at a rapid pace, including health data

How much information is there in the "information society"?

2002 marked the start of the digital age. That was the year that humankind first stored more information in digital than in analogue form

Lopez & Hilbert Significance 2012

Big data hype

- New wording
  - Big data, massive data, data deluge, organic data, open data, data-intensive health care, data mining, data analysts, petabyte, exabyte, zettabyte, …

- Some health-related data become really analyzable
  - genetic data
  - electronic medical records (EMR)
  - internet queries (flu trends)
  - e-patient, self-tracker
  - and more, and more…

The Changing Face of Epidemiology

Is Size the Next Big Thing in Epidemiology?

Sengwee Toh and Richard Platt

Ok but…

Big Data in
Epidemiology

Chiolero A.
Epidemiology 2013

Too Big to Fail?

- Measurement error
- Misclassification
- (zillions of*) Selection bias
- Confounding…
- More than ever with big (=cheap) data

* Ioannidis JPA. Am J Bioethics 2013; 13:40-2

Only describe, don't explain.

— Ludwig Wittgenstein —

AZ QUOTES