Population Health Monitoring and Surveillance Bias

Study protocol for a PhD thesis

Date: 28th February 2022

Student: Stefano Tancredi, MD

Supervisor: Dr. Stéphane Cullati, PhD

Thesis director: Prof. Arnaud Chiolero, MD PhD

Background

Public health surveillance is the ongoing collection, analysis and dissemination of health-related data and is one of the ten essential public health operations according to the World Health Organization (1). A good public health surveillance system allows citizens, healthcare providers, and policy makers to evaluate population health, set priorities, plan policies, and assess the need and the effects of interventions.

Data availability and quality are key to produce information useful for decision. Despite the massive growth in the amount of health-related data available for surveillance (2), translating them into meaningful information is hampered by several issues that can bias the surveillance process, including limited data harmonization, lack of clear methods or surveillance approaches, and privacy issues. One major problem that challenges surveillance data interpretation is surveillance bias. This bias arises when differences in the frequency of a condition are due to differences in the modality of detection rather than to a change in the actual risk of the condition (3, 4). When this happens, it becomes difficult to assess the real burden of diseases and their trends.

Although many conditions are exposed to this bias, it still has to be clearly defined and characterized and its impact is often not taken into consideration when dealing with surveillance data. Therefore, our research aims to increase the knowledge on this bias, providing a clear and comprehensive definition and estimating its impact focusing on two conditions, that is, coronavirus disease 2019 (COVID-19) and prostate cancer. Below, we summarize the specific sub projects that will be carried out within this PhD project.

Research plan

This research project is divided in two parts. The first part aims to retrospectively describe the SARS-CoV-2 spread dynamic in Switzerland using seroprevalence data, that are less exposed to surveillance bias than other epidemic indicators. In the second part, we will use the same data to estimate the scale of surveillance bias of COVID-19 and we will eventually estimate its impact on prostate cancer surveillance.

Part 1 *Monitoring the COVID-19 epidemic in Switzerland*

Background. Serological studies, i.e., repeated cross-sectional or longitudinal studies, are useful tools to estimate the extent and the dynamics of the COVID-19 pandemic (5) . They allow accounting for a- or pauci-symptomatic infections, often missed by surveillance based on reported cases, whose numbers are dependent of different screening and diagnosis strategies across time and regions. Moreover, serological studies allow identifying high-risk groups and risk factors for seropositivity (6-10). For instance, seropositivity can be associated with patients' characteristics, different settings, individual behaviors, or mitigation measures. This information can inform decision making and guide the public health response.

Objectives. Taking advantages of a population based seroprevalence study, our objectives are therefore [1] to assess the seroprevalence trends of Sars-cov-2 in Switzerland and [2] to investigate risk factors for seropositivity and their changes over time.

Methods. We will use data from Corona Immunitas, a population based nationally coordinated research program implemented by the Swiss School of Public Health (SSPH+) (11). The program included several cross-sectional and longitudinal studies conducted in Switzerland using a standardized methodology, with the aim to estimate the number of people who have been infected with SARS-CoV-2 and have developed antibodies. It consisted of four phases that took place in March 2020, between April 2020 and October 2020, between November 2020 and May 2021 and between May 2021 and August 2021, respectively. Randomly selected participants provided a blood sample and filled out a questionnaire on demographic and socioeconomic characteristics, adherence to COVID-19 preventive measures and health status. For this specific study, we will include participants from 11 cantons. For objective 1, we will estimate seroprevalence of IgG antibodies using a Bayesian logistic regression model, adjusted for the antibody test sensitivity and specificity performances. Estimates will be weighted by age and sex of the population of each canton. We will describe seroprevalence per phase and region. We will also describe seroprevalence of neutralizing antibodies across three regions for phase 4. For objective 2, we will assess the association of seropositivity with participants' characteristics and potential risk factors, selected based on findings of previous studies and background expert knowledge, using univariable and multivariable regression analyses.

Relevance and impact. We will provide a nationwide picture of seroprevalence over the various phases of the pandemic that will help to grasp the changes in time and space of the immunological status of the Swiss population from the first phases of the pandemic. Identifying high-risk groups and risk factors for seropositivity will provide useful information for decision makers involved in the pandemic response.

Strengths and limitations. A major strength of this study will be to be part of a national research project made of multiple population-based studies with a standardized research protocol. The same high sensitivity and specificity test was used in each region. Limitations will be selection bias (e.g., higher participation among people who are more wary about risk of COVID-19) and the possible underestimation of seroprevalence because of waning immunity (12) and people failing to produce antibodies (13).

Data science approach. We will conduct descriptive and predictive analysis (14).

Part 2 ***Surveillance bias***

This section is divided in 4 subsections. First, we will provide an overview of what is surveillance bias and its consequences through a narrative review. Second, we will report a case study of surveillance bias of COVID-19 in the Canton of Fribourg, estimating the scale of this bias by the comparison between the 1st and 2nd wave of the pandemic. Third, we will estimate the degree of surveillance bias of COVID-19 in Switzerland, comparing seroprevalence, used as the marker of epidemic size, with the incidence of COVID-19 cases, hospitalizations, ICU hospitalizations, and deaths. Fourth, we will investigate surveillance bias of prostate cancer, assessing the trend in the PSA-based screening uptake in Switzerland and comparing it with the trends in prostate cancer incidence and mortality.

2.1 A narrative review of surveillance bias

Surveillance bias occurs when differences in the frequency of a condition are due to variations in the modality of detection rather than to a change in the actual risk of the condition (3, 4). It can impair population health trends monitoring and it also often affect quality of care assessment, hindering comparability across time, regions, or health care providers.

Many diseases are prone to surveillance bias such as COVID-19, prostate cancer, melanoma, hypertension or chlamydia. In this narrative review, we will provide selected examples of patients' conditions and quality of care indicators at risk of surveillance bias and we will propose solutions to cope with such a bias.

2.2 Surveillance bias of COVID-19: a case study

Background. COVID-19 surveillance based on diagnosed cases is exposed to surveillance bias because the number of diagnosed cases is influenced, among others, by screening and diagnostic strategies and testing capacity (15, 16). By contrast, population-based seroprevalence studies can estimate the actual burden of disease as they account for all infections independently of diagnostic and screening strategies.

Objectives. In this case study, we will estimate the scale of surveillance bias in the Canton of Fribourg comparing the 1st and 2nd wave of the pandemic, before the deployment of the vaccination campaign**.**

Methods. We will use data from two serosurveys conducted in the Canton of Fribourg as part of Corona Immunitas, after the 1st wave (July - October 2020) and after the 2nd wave of the pandemic (November 2020 - February 2021). Data on the incidence of COVID-19 diagnosed cases (positive PCR or rapid antigen test) will be retrieved from the Federal Office of Public Health.

Relevance and impact. This study will help illustrating the concept of surveillance bias and will provide insight into the scale of its impact during the COVID-19 pandemic.

Strengths and limitations. Both seroprevalence data and routine surveillance data will be retrieved from reliable data sources. Main limitations include selection bias (e.g., higher participation among people who suspected to have COVID-19) and underestimation of SARS-CoV-2 seroprevalence because of waning immunity (12) and people failing to produce antibodies (13).

Data science approach. In this work we will perform descriptive analysis.

2.3 Estimating surveillance bias of COVID-19

Background. Several markers of epidemic size have been used during the COVID-19 pandemic, that is, the number of COVID-19 cases, hospitalizations, or deaths. All these epidemic indicators are prone to surveillance bias to different degrees. Seroprevalence is a less biased marker, as it allows accounting for a- or pauci-symptomatic undiagnosed infections and is not dependent on screening and diagnostic strategies.

Objectives. Our objective is to estimate the scale of surveillance bias of COVID-19 using different types of data, comparing seroprevalence with the incidence of COVID-19 cases, hospitalizations, ICU hospitalizations, and deaths in Switzerland.

Methods. We will use the prevalence of SARS-CoV-2 antibodies (before vaccination campaign) as a "true" marker of epidemic size. We will retrieve seroprevalence data from Corona Immunitas. To compare seroprevalence with the incidence of COVID-19 cases, hospital admissions and mortality we will use routine surveillance data from the Federal Office of Public Health. We will

made comparisons both at the national level and cantonal level, to account for interregional variability.

Relevance and impact. This study will help design a population health monitoring epidemic tool using different types of data. A better understanding and quantification of surveillance bias will provide useful information that could ease data interpretation and allow a more accurate COVID-19 surveillance activity.

Strengths and limitations. Both seroprevalence data and routine surveillance data will be retrieved from reliable data sources. Main limitations include selection bias (e.g., higher participation among people who suspected to have COVID-19) and underestimation of SARS-CoV-2 seroprevalence because of waning immunity (12) and people failing to produce antibodies (13).

Data science approach. We will perform descriptive analysis.

2.4 Incidence, mortality and screening trends for prostate cancer in Switzerland

Background. Prostate cancer is the second most frequent cancer in men, accounting for about 8% of all new cases of cancer and 7% of all cancer deaths in men in 2020 (17). Given this high burden, a lot of effort has been put into screening activities since 1990s (18). However, several trials have demonstrated that, despite possible benefits, PSA-based screening can be harmful, mainly because of overdiagnosis and overtreatment (19-22). Therefore, the US Preventive Service Task force recommends against PSA-based screening for prostate cancer in men 70 years and older and recommends individual decision to undergo screening in men aged from 55 to 69 years (18, 21). Nevertheless, other organizations have other recommendations (23, 24), and differences in the use of the PSA test over time and among countries result in differences in the incidence rates of prostate cancer.

Objectives. In this study, we will assess the trend in screening uptake in Switzerland and we will compare it with the trends in prostate cancer incidence and mortality.

Methods. Data on the uptake of the PSA-based screening in Switzerland will be retrieved from the Swiss Health Survey (25), a repeated survey conducted by the Swiss Federal Statistical Office every five years (from 1992 to 2017) on a representative sample of individuals aged 15 and older permanently residing in Switzerland. Data on incidence and mortality will be retrieved from the Swiss National Agency for Cancer Registration (NICER) for the period 1992 to 2016 (26).

Relevance and impact. By comparing the trends in incidence and mortality rates with the trend in screening uptake (both at the national and cantonal level) this study will allow to determine if surveillance bias occurred.

Strengths and limitations. Incidence trends may be influenced by the gradual introduction of cancer registries in Switzerland and estimates may be affected by under registration. Data from the Swiss Health Survey are self-reported and information bias is possible.

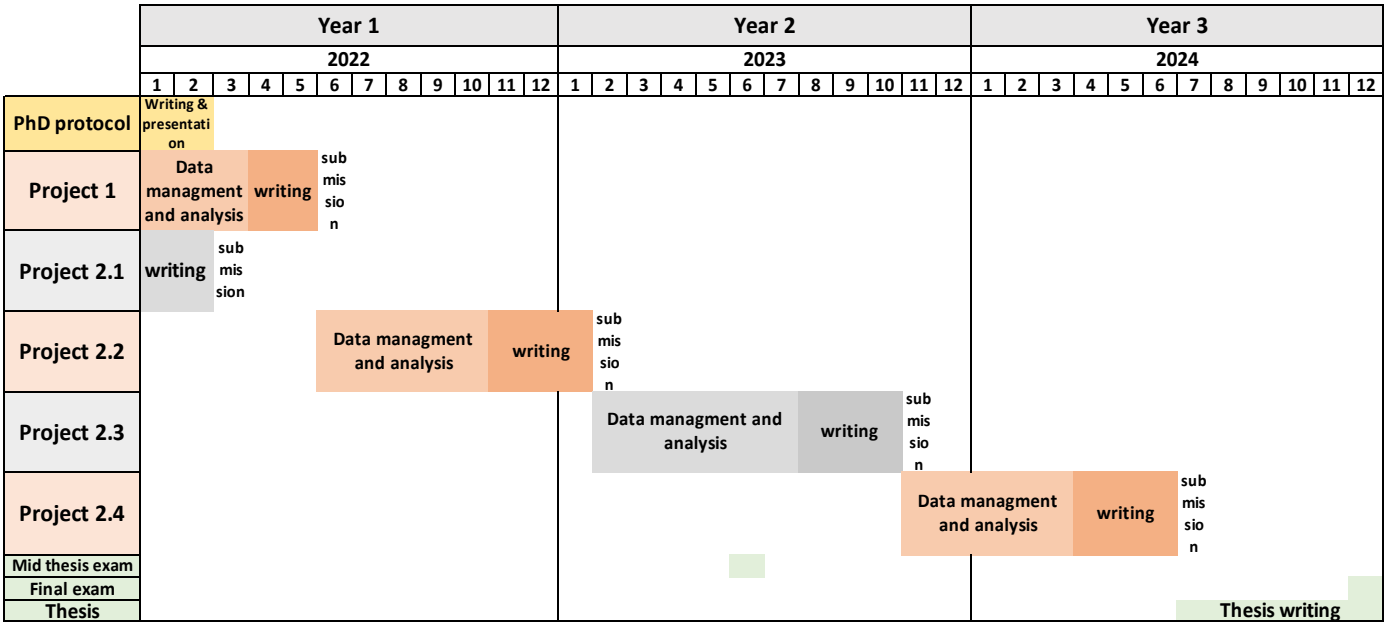
Data science approach. We will perform descriptive analysis.

Expected Outcomes

This project will address the issue of surveillance bias. Healthcare professionals, data scientists and policy makers should be aware of this bias: accounting for it is necessary for accurate public health surveillance and monitoring activity. Our goals are to provide a clear and comprehensive definition of surveillance bias, suggest possible approaches to cope with it and to estimate its impact in two commons conditions. The results of this PhD projects will be published in peer reviewed journals.

Timeframe

The project will start in January 2022 and last for 3 years. Below, a provisional timetable.



References

1. WHO. The 10 essential public health operations [Accessed February 24, 2022] [Available from: https://www.euro.who.int/en/health-topics/Health-systems/public-health-services/policy/the-10-essential-public-health-operations.]
2. Chiolero A. How infodemic intoxicates public health surveillance: from a big to a slow data culture. J Epidemiol Community Health. 2022.
3. Haut ER, Pronovost PJ. Surveillance bias in outcomes reporting. Jama. 2011;305(23):2462-3.
4. Chiolero A, Buckeridge D. Glossary for public health surveillance in the age of data science. J Epidemiol Community Health. 2020;74(7):612-6.
5. World Health Organization. Population-based age-stratified seroepidemiological investigation protocol for COVID-19 virus infection, 17 March 2020. Geneva: World Health Organization; 2020 2020. Contract No.: WHO/2019-nCoV/Seroepidemiology/2020.1.
6. Richard A, Wisniak A, Perez-Saez J, Garrison-Desany H, Petrovic D, Piumatti G, et al. Seroprevalence of anti-SARS-CoV-2 IgG antibodies, risk factors for infection and associated symptoms in Geneva, Switzerland: a population-based study. Scand J Public Health. 2022;50(1):124-35.
7. Basto-Abreu A, Carnalla M, Torres-Ibarra L, Romero-Martínez M, Martínez-Barnetche J, López-Martínez I, et al. Nationally representative SARS-CoV-2 antibody prevalence estimates after the first epidemic wave in Mexico. Nat Commun. 2022;13(1):589.
8. Bobrovitz N, Arora RK, Cao C, Boucher E, Liu M, Donnici C, et al. Global seroprevalence of SARS-CoV-2 antibodies: A systematic review and meta-analysis. PLoS One. 2021;16(6):e0252617.
9. Vos ERA, den Hartog G, Schepp RM, Kaaijk P, van Vliet J, Helm K, et al. Nationwide seroprevalence of SARS-CoV-2 and identification of risk factors in the general population of the Netherlands during the first epidemic wave. J Epidemiol Community Health. 2020;75(6):489-95.
10. Galanis P, Vraka I, Fragkou D, Bilali A, Kaitelidou D. Seroprevalence of SARS-CoV-2 antibodies and associated factors in healthcare workers: a systematic review and meta-analysis. J Hosp Infect. 2021;108:120-34.
11. West EA, Anker D, Amati R, Richard A, Wisniak A, Butty A, et al. Corona Immunitas: study protocol of a nationwide program of SARS-CoV-2 seroprevalence and seroepidemiologic studies in Switzerland. Int J Public Health. 2020;65(9):1529-48.

12. ECDC. Immune responses and immunity to SARS-COV-2 [Accessed February 24, 2022] [Available from: https://www.ecdc.europa.eu/en/covid-19/latest-evidence/immune-responses.]
13. Wei J, Matthews PC, Stoesser N, Maddox T, Lorenzi L, Studley R, et al. Anti-spike antibody response to natural SARS-CoV-2 infection in the general population. Nat Commun. 2021;12(1):6250.
14. Hernán MA, Hsu J, Healy B. A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks. CHANCE. 2019;32(1):42-9.
15. Tancredi S, Anker D, Rosella L, Chiolero A. Elimination of covid-19: beware of surveillance bias. Bmj. 2021;374:n2126.
16. Wu SL, Mertens AN, Crider YS, Nguyen A, Pokpongkiat NN, Djajadi S, et al. Substantial underestimation of SARS-CoV-2 infection in the United States. Nat Commun. 2020;11(1):4507.
17. Ferlay J EM LF, Colombet M, Mery L, Pineros M, Znaor A, Soerjomataram I. Global cancer observatory: cancer today. Lyon, France: International Agency for Research on Cancer [Accessed February 24, 2022] [Available from: https://gco.iarc.fr/today]
18. Moyer VA. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2012;157(2):120-34.
19. Pinsky PF, Prorok PC, Yu K, Kramer BS, Black A, Gohagan JK, et al. Extended mortality results for prostate cancer screening in the PLCO trial with median follow-up of 15 years. Cancer. 2017;123(4):592-9.
20. Andriole GL, Crawford ED, Grubb RL, 3rd, Buys SS, Chia D, Church TR, et al. Prostate cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: mortality results after 13 years of follow-up. J Natl Cancer Inst. 2012;104(2):125-32.
21. Fenton JJ, Weyrich MS, Durbin S, Liu Y, Bang H, Melnikow J. Prostate-Specific Antigen-Based Screening for Prostate Cancer: Evidence Report and Systematic Review for the US Preventive Services Task Force. Jama. 2018;319(18):1914-31.
22. Martin RM, Donovan JL, Turner EL, Metcalfe C, Young GJ, Walsh EI, et al. Effect of a Low-Intensity PSA-Based Screening Intervention on Prostate Cancer Mortality: The CAP Randomized Clinical Trial. Jama. 2018;319(9):883-95.
23. Parker C, Castro E, Fizazi K, Heidenreich A, Ost P, Procopio G, et al. Prostate cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2020;31(9):1119-34.
24. Wolf AM, Wender RC, Etzioni RB, Thompson IM, D'Amico AV, Volk RJ, et al. American Cancer Society guideline for the early detection of prostate cancer: update 2010. CA Cancer J Clin. 2010;60(2):70-98.
25. Office federal de la statistique. Enquête suisse sur la santé [Accessed February 24, 2022] [Available from: https://www.bfs.admin.ch/bfs/fr/home/statistiques/sante/enquetes/sgb.html.]
26. National statistics on cancer incidence. NICER [Accessed February 24, 2022] [Available from: https://www.nicer.org/en/statistics-atlas/cancer-incidence/.]