Michele Santacatterina

Personal Data

email:	michele.santacatterina@nyulangone.org
Skype ID:	miksanta86
Profiles:	Personal Website, Google Scholar, LinkedIn, Research Gate
Research Interests:	Biostatistics, Causal Inference, Data Science, Healthcare, Real-World Data

Academic Appointments

Current 2021 Oct Assistant Professor of Biostatistics,
Division of Biostatistics, Department of Population Health,
NYU Langone Health

Current 2021 Sep Assistant Research Professor of Biostatistics and Bioinformatics,
The Biostatistics Center, Department of Biostatistics and Bioinformatics, Milken Institute School of Public Health,
George Washington University
Co-Investigator of the Data Coordinating Center for the COVID-19 Community Research Partnership Study and the Hurricane Maria Mortality Project.

Education

Academic Degrees

2018 Apr 2013 Dec Ph.D., Medical Science, Karolinska Institutet, Stockholm, Sweden
Main topics: Biostatistics, Causal Inference, Data Science, Optimization.

2012 Mar 2009 Oct M.S., Biostatistics, University of Milano-Bicocca, Milan, Italy
Main topics: Classical and Bayesian inference, Survival analysis, Longitudinal data analysis, Statistical methods for Epidemiology, and Design of Experiments.

2009 Sep 2006 Sep B.S., Statistics and Computer Science, University of Padua, Padua, Italy
Selected topics: Probability, Statistical inference, Statistical models, Computational Statistics, Algebra, Linear Optimization, and Information Technology.

Relevant Courses

2016 Mar 2015 Nov Course SF3961 "Statistical Inference" at KTH Royal Institute of Technology
Organized and taught by Prof. Henrik Hult.
15-ECTS Ph.D.-level course within the Mathematical Statistics Ph.D. Program at KTH.

2015 Jun 2015 Mar Course SF2822 "Applied nonlinear optimization" at KTH Royal Institute of Technology
Organized and taught by Prof. Anders Forsgren.
7.5-ECTS MSc-level course within the Applied Mathematics Program at KTH.

Other Education

2019 Jan — Auditing: CS6780 - Advanced Machine Learning
Taught by Prof. Thorsten Joachims, Cornell University, Department of Computer Science and Department of Information Science.

2014 Dec — Course on "Regression Models" by John Hopkins University on Coursera
Organized and taught by Prof. Brian Caffo, Prof. Roger D. Peng and Assoc. Prof. Jeff Leek, Johns Hopkins University.
Certification and grades.

2013 Dec — Course on "Causal inference from observational data" at Karolinska Institutet
Organized by the Swedish Interdisciplinary Graduate School at Karolinska Institutet. Taught by Prof. Miguel Hernán, Harvard School of Public Health.

2013 Jun — Summer School in "Modern Methods in Biostatistics and Epidemiology", Treviso, Italy
Organized jointly by Harvard School of Public Health and Karolinska Institutet.
Course attended:
"Causal Inference" by Prof. Andrea Rotnitzky, Harvard School of Public Health

2010 Feb — SAS Masterclass in Business Intelligence, Milan, Italy
Organized jointly by SAS and Accenture.
Certification obtained:
SAS Certified Base Programmer for SAS 9

2009 Aug
2009 Feb — Visiting Student, Department of Mathematics, University of Barcelona
Course attended: Probability, Statistical inference.

Research Experience

2018 Aug- 2020 Aug	Postdoctoral Associate, TRIPODS Center for Data Science and Cornell Tech, Cornell University Development of novel statistical methods for optimal decision making using experimental and observational data with applications in public health and medicine. Mentors: Prof. Nathan Kallus, Prof. Thorsten Joachims, Prof. Kilian Weinberger
2013 Dec - 2018 Apr	Ph.D. student, Unit of Biostatistics, Karolinska Institutet Study and development of statistical methods based on mathematical programming techniques with application in public health and medical research. Teaching assistant. Advisor: Prof. Matteo Bottai
2017 Jun - 2017 Sept	Visiting Scholar, Cornell Tech, Cornell University Collaborate on research into novel optimization approaches to causal inference. Faculty Sponsor: Prof. Nathan Kallus
2013 Jan - 2013 Nov	Research assistant, Unit of Biostatistics, Karolinska Institutet Application of statistical methods in public health research, with focus on HIV research. Writing of grant applications (KID funding). Teaching assistant.
2012 Mar - 2012 Dec	Research assistant, Department of Public Health, Karolinska Institutet Application of statistical methodologies for medical and public health research.
2011 Sept - 2012 Feb	Master's Thesis, Department of Public Health, Karolinska Institutet Title: Access to antiretroviral therapy and treatment effectiveness among injection drug users: results from a Swedish population based study. Advisor: Prof. Rino Bellocco
2008 Nov - 2009 Jan	Internship, Department of Pediatrics, University of Padua. Application and evaluation of methods for survival analysis using data from patients with anaplastic large-cell lymphoma. Contact persons: Prof. Laura Ventura and Dr. Gloria Tridello

Manuscripts and Publications

Work in progress

2021	Prediction of severe hypoglycemia: machine learning approaches in the EDIC Cohort **Santacatterina, M.**, Uschner, D., Bebu, I. Application of machine learning approaches to predict hypoglycemia
2021	Analysis of cause-specific mortality of the Hurricane Maria in Puerto Rico **Santacatterina, M.**, et. al., Evaluation of the excess cause-specific mortality after Hurricane Maria in Puerto Rico.
2021	Predicting the risk of bleeding disorders **Santacatterina, M.** and Vaughan, R. Application of novel machine learning methods to help understand and predict the risk of bleeding disorders.

2020 | *Quantile based regression for life expectancy estimation with censored data.* Garcia-Pareja, C., and **Santacatterina, M.**, et. al.
Novel statistical method for estimating life expectancy.

Statistical and data science methodology

2021 | *Deep survival analysis with longitudinal X-rays for COVID-19*
Shu, M., Bowen, R.S., Herrmann, C., Qi, G., **Santacatterina, M.** and Zabih, R. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4046-4055. 2021
Development of a deep learning method that handles images longitudinally for better predictions of COVID-19 outcomes

2021 | *Robust weights that optimally balance confounders for estimating marginal hazard ratios*
Santacatterina, M., Acceptable after revision in Statistical Methods in Medical Research, 2021 (*)
Novel optimization method for estimating marginal hazard ratios for Cox models.

2021 | *More robust estimation of average treatment effects using kernel optimal matching in an observational study of spine surgical interventions*
Kallus, N., Pennicooke, B., **Santacatterina, M.** Statistics in medicine 40, no. 10 (2021): 2305-2320. (*)
Extension and application of Kernel Optimal Matching to robustly estimate sample average treatment effects under lack of overlap.

2020 | *Optimal weighting for estimating generalized average treatment effects*
Kallus, N., and **Santacatterina, M.**, Acceptable after revision in the Journal of Causal Inference, 2020 (*)
Formulation of a new general causal estimand and development of an optimization method for its estimation.

2020 | *Optimal balancing of time-dependent confounders for marginal structural models*
Kallus, N., and **Santacatterina, M.**, Acceptable after revision in the Journal of Causal Inference, 2020 (*)
Novel method that estimates causal effects affected by time-dependent confounders from longitudinal observational studies.

2019 | *Kernel optimal orthogonality weighting: a balancing approach to estimating effects of continuous treatments*
Kallus, N., and **Santacatterina, M.**, arXiv preprint arXiv:1910.11972, 2019 (*)
Novel optimization method for estimating causal dose-response curves.

2019 | *CAB: Continuous Adaptive Blending estimator for policy evaluation and learning*
Su, Y., Wang, L., **Santacatterina, M.**, Joachims, T, In International Conference on Machine Learning, pp. 6005-6014. PMLR, 2019.
Novel method for estimating optimal targeted policies.

2019 | *Optimal probability weights for estimating causal effects of time-varying treatments with marginal structural Cox models*
Santacatterina, M., Garcia-Pareja, C., Bellocco, R., Sönnerborg, A., Ekström, A.M., Bottai, M., Statistics in medicine 38, no. 10 (2019): 1891-1902.
Use of optimal probability weights when estimating treatment effects with longitudinal data. Comparisons, through simulations, with existing methods.

2018 | *Optimal probability weights for inference with constrained precision*

Santacatterina, M., and Bottai, M., Journal of the American Statistical Association 113, no. 523 (2018): 983-991.

Proposal of a method for obtaining optimal probability weights based on solving a constrained nonlinear optimization problem.

(*) The order of authors follows alphabetical order.
Santacatterina, M. is the corresponding author of those manuscripts.

Healthcare

2021 | *Prevention of Covid-19 with the BNT162b2 and mRNA-1273 Vaccines*

Santacatterina, M., et. al., The New England journal of medicine, 10.1056/NEJMc2113575, 2021

Letter to the Editor about the evaluation of vaccine effectiveness in real-world settings in the United States

2021 | *Using Repeated Antibody Testing to Minimize Bias from False Positives Estimates of Prevalence and Incidence of SARS-CoV-2 Infection*

Santacatterina, M. , et. al., Under Review in the American Journal of Epidemiology, 2021

Novel proposal of a combined parallel and sequential testing framework to reduce overestimation of prevalence and incidence

2021 | *Antibody responses among participants in the COVID-19 Community Research Partnership receiving one or two doses of COVID-19 mRNA vaccine*

Friedman-Klabanoff, D. J., Berry, A. A., Tjaden, A. H., **Santacatterina, M.** , et. al., Under Review in Clinical Infectious Diseases, 2021

Examination of seroconversion rates among vaccinated people

2021 | *The effect of patient age on discharge destination and complications after lumbar spinal fusion*

Pennicooke, B., **Santacatterina, M.**, et al., Journal of Clinical Neuroscience 91 (2021): 319-326.

A medical study evaluating the effect of age on surgical outcomes.

2021 | *Analysis of Accumulated SARS-CoV-2 Seroconversion in North Carolina: The COVID-19 Community Research Partnership*

Williamson, J.C., Wierzba, T.F., **Santacatterina, M.**, et al., medRxiv (2021)

A public health study on COVID-19 seroconversion.

2021 | *Duration of SARS-CoV-2 Sero-Positivity in a Large Longitudinal Sero-Surveillance Cohort: The COVID-19 Community Research Partnership*

COVID-19 Community Research Partnership Study Group, BMC Infect Dis 21, 889 (2021).

A medical study on the duration of antibodies among COVID-19 positive participants.

2020 | *Does state malpractice environment affect outcomes following spinal fusions? A robust statistical and machine learning analysis of 549,775 discharges following spinal fusion surgery in the United States*

Chan, A.K., **Santacatterina, M.**, et al., Neurosurgical focus 49, no. 5 (2020): E18.

A medical study on the effect of malpractice claims on clinical outcomes.

2017 | *Risk behaviour determinants among people who inject drugs in Stockholm, Sweden over a 10-year period, from 2002 to 2012*

Karlsson, N., **Santacatterina, M.**, et al., Harm reduction journal 14, no. 1 (2017): 1-11.

A study on understanding the underlying causes that drive various types and levels of risk behaviours among people who inject drugs.

2017 | *Effect of therapy switch on time to second-line antiretroviral treatment failure in HIV-infected patients*

Häggblom, A., **Santacatterina, M.**, (a) et al., PloS one 12, no. 7 (2017): e0180140.

A medical study on the effect of therapy switch on time to second-line antiretroviral treatment failure among HIV-infected patients.

2016 | *Impact of peer support on virologic failure in HIV-infected patients on antiretroviral therapy-a cluster randomized controlled trial in Vietnam*

Sönnerborg, A., Tam, V.V., El-Khatib, Z., **Santacatterina, M.**, et al., BMC infectious diseases 16, no. 1 (2016): 1-14.

A randomized control trial evaluating the impact of peer support on clinical outcomes in Vietnam.

2016 | *Inferences and conjectures in clinical trials: a systematic review of generalizability of study findings*

Santacatterina, M., and Bottai, M., Journal of internal medicine 279, no. 1 (2016): 123-126.

Systematic review on generalizability and heterogeneity of study findings in randomized clinical trials.

2015 | *Correlates of mobile phone use in HIV care: Results from a cross-sectional study in South Africa*

Madhvani, N., Longinetti, E., **Santacatterina, M.**, et al., Preventive medicine reports 2 (2015): 512-516.)

A cross-sectional study on the evaluation of the use of mobile phones as reminders in HIV care.

2014 | *Temporal Trends in the Swedish HIV-1 Epidemic: Increase in Non-B Subtypes and Recombinant Forms over Three Decades*

Neogi, U., Häggblom, A., **Santacatterina, M.**, et al., PloS one 9, no. 6 (2014): e99390.

A study on the temporal trend of the subtype distribution from the beginning of the HIV-1 epidemic in Sweden.

2014 | *The State-Led Large Scale Public Private Partnership 'Chiranjeevi Program' to Increase Access to Institutional Delivery among Poor Women in Gujarat, India: How Has It Done? What Can We Learn?*

De Costa, A., Kranti, S.V., Ryan, K., Sankara Raman, P., **Santacatterina, M.**, et al., PLoS One 9, no. 5 (2014): e95704.

A study determining risk factors associated with disclosing HIV status among antiretroviral therapy recipients in South Africa.

2014 | *Gender perspective of risk factors associated with disclosure of HIV status, a cross-sectional study in Soweto, South Africa*

Longinetti, E., **Santacatterina, M.**, El-Khatib, Z., PLoS One 9, no. 4 (2014): e95440.

A cross-sectional study on risk factors for viremia and drug resistance.

2013 | *Antibiotic prescribing in women during and after delivery in a non-teaching, tertiary care hospital in Ujjain, India: a prospective cross-sectional study*

Sharma, M., Sanneving, L., Kalpana, M., **Santacatterina, M.**, et al., Journal of pharmaceutical policy and practice 6, no. 1 (2013): 1-7.

A study that aimed at present antibiotic prescribing among inpatients during and after delivery in rural India.

(a) These authors contributed equally to this work.

2012 | *Monitoring the efficacy of antiretroviral therapy by a simple reverse transcriptase assay in HIV-infected adults in rural Vietnam*
Cuong, D.D., Agneskog, E., Nguyen Thi Kim, C., **Santacatterina, M.**, et al., Future Virology 7, no. 9 (2012): 923-931.
Determination of the feasibility of viral load monitoring in a cohort of HIV treatment-naive adult patients initiating antiretroviral therapy in rural Vietnam.

Conferences and Invited Talks

Organizer

MAR 14, 2021 | ENAR 2021
Organizer and chair of the invited paper session: *Leveraging real-world data for improved medical decision-making: challenges, opportunities, and recent developments*

DEC 14, 2019 | NeurIPS, 2019
Organizer of the workshop: *"Do the right thing": machine learning and causal inference for improved decision making*, together with Thorsten Joachims (Cornell), Nathan Kallus (Cornell), Adith Swaminathan (Microsoft Research), David Sontag (MIT), and Angela Zhou (Cornell)

MAY 23, 2019 | Atlantic Causal Inference Conference 2019
Organizer of the symposium: *Optimization methods for causal inference*, together with Adam Kapelner (CUNY), Nathan Kallus (Cornell), Nikos Ignatiadis (Stanford), and Stefan Wager (Stanford)

Talks

SEP 29, 2021 | NYU Langone Health - Department of Population Health Seminar Series *Real-World SARS-CoV-2 Vaccine Effectiveness in North Carolina: The COVID-19 Community Research Partnership*. Invited Talk.

AUG 09, 2021 | JSM 2021. *Optimal Weighting for Estimating Generalized Average Treatment Effects*. Talk.

MAR 15, 2021 | ENAR 2021. *Robust Weights that Optimally Balance Confounders for Estimating the Effect of Binary and Continuous Treatments with Time-to-event Data*. Talk.

FEB 19, 2021 | GWU - Statistics Seminar Series. *Optimal weighting for estimating treatment effects*. Invited talk.

DEC 01, 2020 | Harvard - Machine Learning and Causal Inference Reading group. *Optimal weighting for estimating treatment effects*. Invited talk.

Aug 03, 2020	JSM 2020 - Health policy statistics section. *Optimal estimation of generalized average treatment effects using Kernel Optimal Matching.* Invited talk.
Mar 22, 2020	ENAR 2020 - IMS invited session on challenges for precision medicine. *Kernel optimal orthogonality weighting: a balancing approach to estimating effects of continuous treatments.* Invited talk.
Jan 15, 2020	McGill Department of Biostatistics - Biostatistics Seminar. *Kernel optimal orthogonality weighting: a balancing approach to estimating effects of continuous treatments.* Invited talk.
Sep 23, 2019	Cornell Machine Learning in Medicine. *Kernel optimal orthogonality weighting: a balancing approach to estimating effects of continuous treatments.* Invited talk.
May 23, 2019	Atlantic Causal Inference Conference 2019. *Optimal estimation of generalized average treatment effects using Kernel Optimal Matching.* Talk and poster.
Oct 24, 2018	Second TRIPODS PI meeting. *Optimal balancing of time-dependent confounders for marginal structural models* . Talk.
Sep 7, 2018	Cornell - AI Seminar. *Optimal Weighting for Causal Inference* . Invited talk.
Apr 11, 2018	EUROCIM Causal Inference 2018. *Optimal balancing of time-dependent confounders for marginal structural models* . Talk.
Oct 18, 2017	MELODEM Selection Group Meeting. *Optimal probability weights for inference with constrained precision* . Invited talk.
Sep 7, 2017	Royal Statistical Society International Conference. *Estimating treatment effects with optimal inverse probability weighting.* Talk.
April 5, 2017	UK Causal Inference Meeting. *Estimating treatment effects with optimal inverse probability weighting.* Poster.
Sep 7, 2016	Royal Statistical Society International Conference. *Optimal probability weights for inference with constrained precision* . Talk.
Sep 4, 2015	Nordic and Baltic Stata Users Group meeting. *Weight watchers: How to optimize your weight* . Invited talk.
Sep 28, 2014	HIV Nordic conference. *Antiretroviral therapy among HIV-infected people who inject drugs in Sweden: access and treatment response* . Poster.

Advising and Mentoring

Master Students

2021 | Nolan Kuenster, M.Ph. in Epidemiology - George Washington University (Advisor). *Application of statistical methods for excess mortality estimation*

2019 | Yaniv Ravid, M.Eng. in Operations Research and Information Engineering - Cornell Tech (Advisor). *A Python library for Kernel Optimal Matching*

2017 | Claudia Carlucci, M.Sc. in Biostatistics - Univeristy of Milano-Bicocca and Karolinska Institutet (Mentor). *Modeling additive interaction with continuous variables: smoking and the risk for rheumatoid arthritis*

2015 | Chiara Chiavenna, M.Sc. in Biostatistics - Univeristy of Milano-Bicocca and Karolinska Institutet (Mentor). *Laplace regression with censored data: an overview and application to observational data in cardiovascular epidemiology*

Teaching Experience

2021 Jul – 2021 Jun | Lecturer at the George Washington University

Teaching, preparation of material and computer laboratory sessions =.

- PUBH 6002 - Biostatistical Applications for Public Health
- Summer Institute in Health Data Science and Biostatistics.

2018 Apr – 2015 Apr | Teaching Assistant at the Karolinska Institute

Preparation of material and teaching of computer laboratory sessions.

- "Biostatistics I" for the Doctoral Program in Epidemiology,
- "Biostatistics II" for the Doctoral Program in Epidemiology,
- "Biostatistics I" for the Research School for Clinical Epidemiology, and,
- "Biostatistics II" for the Research School for Clinical Epidemiology.

The courses are given once per semester. They cover notions of statistics and biostatistics and the use of `STATA` for data analysis.

Funding

2021 Nov – 2025 Nov | CO-PI, NSF 21-530 - Smart Health and Biomedical Research in the Era of Artificial Intelligence and Advanced Data Science

Project title: Interpretable survival analysis of complex longitudinal data. 1,199,980 USD.

Proposal submitted - Written in collaboration with Prof. Ramin Zabih (PI), Dr. Yifan Peng, and Dr. Alexander Rush.

2017 Dec – 2013 Dec | Karolinska Institutet Doctoral (KID) funding

Project title: Novel methods for estimating optimal dynamic treatment regimes for HIV-infected patients. 1,280,000 SEK.

Written in collaboration with Prof. Matteo Bottai (PI), Prof. Anna Mia Ekström, Prof. Anders Sönnerborg and Prof. Rino Bellocco.

Other awards and merits

2017 Jun 2017 Aug	KI Travel Grant Visiting Scholar at Cornell Tech. 14,000 sek.
2012 Feb 2011 Sept	Programme Extra Plus - Fondazione Cariplo Visiting student with the aim of working on the Master's thesis at the Division of Global Health, Karolinska Institutet. 4,500 EUR.
2009 Aug 2009 Feb	Erasmus+ Grant Visiting student at the Department of Mathematics, University of Barcelona. 1,500 EUR.

Service

Current 2020 Nov	Chair and organizer of the working group for a summer Institute in Health Data Science and Biostatistics
Current 2020 Oct	Chair of the Data Analysis and Visualization Working Group of the COVID-19 Community Research Partnership Study
Current 2020 Sept	Active member of the Department Faculty Meetings
2020 Aug 2019 Feb	Involved in the project WiTNY, Women in Technology and Entrepreneurship in New York.
Current 2017 Apr	Member of the Steering committee for the EURO-CIM
Current 2018 Aug	Reviewer for the *Journal of the Royal Statistical Society - Series A, Biometrical Journal, Computational Statistics and Data Analysis, NeurIPS, ICML, AISTAT, Nature Machine Intelligence, BMC Medical Research Methodology, Epidemiological methods, Journal of the American Statistical Association, Statistics in Biopharmaceutical Research, Clinical Infectious Diseases, Annals of Applied Statistics.*

Languages

English:	Fluent (C1/2).
Italian:	Mother tongue.
Spanish:	Basic knowledge (A2).
Swedish:	Basic knowledge (A2).

Computer Skills

Operative Systems
Advanced user of Linux, MacOS and Windows.

Programming Languages

Advanced:	`R` and `STATA`
Intermediate:	`Python`, `scikit-learn` and `SAS`
Notions of:	`Mathematica`, `SageMath`, `SQL`, `TensorFlow` and `Matlab`

Character and Skills

- Independent
- Cooperative
- Devoted

- Open-minded

Other Interests

When I am not reading about biostatistics and data science, I like to learn more about contemporary history. I also enjoy keeping myself active by playing sports and going to the gym. I especially enjoy running in the forest. It is my way of releasing stress and enjoy life. I love to play and listen to music, especially electronic and film music. During the past few years I started several electronic music projects that included production using Ableton Live. Finally, together with my wife Alice, I have a dog named Mister Pom Princess "PK" Cake, which keeps me happily busy.