# Le Bao

Address: Department of Statistics, The Pennsylvania State University, University Park, PA 16802

Homepage: https://lebao0215.github.io/

Email: lebao@psu.edu

## Education:

2006 – 2011 Ph.D. in Statistics, University of Washington, Seattle, WA

Doctoral dissertation, "Statistical Models for Estimating and Projecting HIV/AIDS Epidemics," (advisor: Adrian E. Raftery).

2004 – 2005 M.S. in Statistics, Dalhousie University, Canada

2000 – 2004 B.S. in Applied Mathematics, Peking University, China

## Professional Position:

2023 – present Professor of Statistics, The Pennsylvania State University

2023 – present Associate Head, Department of Statistics, The Pennsylvania State University

2014 – present Associate Director of Center for Advanced Data Assimilation and Predictability Techniques (ADAPT), The Pennsylvania State University

2017 – 2023 Associate Professor of Statistics, The Pennsylvania State University

2018 – 2019 Visiting Professor, Center of Statistical Science, Peking University

2011 – 2017 Assistant Professor of Statistics, The Pennsylvania State University

## Honors:

2023 Certificates of Excellence in Teaching and Making a Positive Impact on Students, Eberly College of Science, The Pennsylvania State University

2022 Rosenberger Outstanding Teaching Award, Department of Statistics, The Pennsylvania State University

2010 Chinese Government Award for Outstanding Students Abroad

2009 Z.W. Birnbaum Award for Outstanding General Exam, University of Washington

## Editorial Position:

2018 – present Associate Editor for The Annals of Applied Statistics

2021 – 2022 Guest Associate Editor for PNAS

2018 – 2021 Statistical Reviewer for Lancet HIV

2013 – 2016 Statistical Adviser for PLOS One

## Professional Activities:

2024 Program Chair-Elect, ASA Statistics in Epidemiology Section

2011 - present Key technical advisor of the UNAIDS Reference Group on HIV Estimates, Modelling and Projections who advises on the methods for calculating international AIDS statistics. http://www.epidem.org/about-us

2014 - present National Institute of Statistical Sciences (NISS) Liaison

2013 - 2016 Core project team leader of the Diagnostics Modelling Consortium who aims to utilize modelling to guide the effective use of diagnostic technologies in resource-poor settings. http://www.dxmodelling.org/people

2014, 2015, 2017 Expert Consultation on Meeting for HIV/AIDS Response of China, Ministry of Health (MoH), Beijing, China

2017 Expert Consultation on Estimating the Size of Key Populations in Resource Limited Settings, Atlanta, US Centers for Disease Control and Prevention

2010 Consultant, Regional Training on Methods for Size Estimation of Most-at-Risk Populations in the Asia-Pacific Region, Bangkok, Thailand

2009 Consultant, Training of Trainers Workshop on HIV/AIDS Estimates and Projections, Geneva, Switzerland

## Service to the University:

2014 – present Associate Director of the Center for Advanced Data Assimilation and Predictability Techniques (ADAPT).

2013 – present Bioinformatics and Genomics Training Program Committee.

2016 – 2018 Search Committee for Faculty Position in Data Ethics.

2012 –2013 Protocol reviewer at Clinical Research Center (CRC).

## Service to the Department:

2020 – present Schreyer Honors Advisor in Statistics.

2019 – present Award Committee.

2018 – present Tenure and Promotion Committee.

2020 – present Undergraduate Education Program.

2019 – present Graduate Program Curriculum Committee.

2020 – 2024 Stat Club Co-Advisor.

2018 – 2023 Chair of Undergraduate Research.

2020 – 2021 Co-Chair of Strategic Planning Committee, Undergraduate Education.

2019 – 2020 Chair of Search Committee for Faculty Position in Statistics.

2018 – 2020 World Campus / Online Education Committee.

2014 – 2017 First Year Graduate Student Adviser.

2013 – 2017 Undergraduate Student Adviser.

2011, 2012, 2015 PhD Qualifying Exam Committee.

2013 – 2015 Clifford C. Clogg Memorial Lecture Committee.

2012 – 2013 Colloquium Chair.

2012 – 2013 Khatri / Krishnaiah / Rao Prize Committee.

## Funding Support:

1. Project Title: Leveraging Big Data Science to Focus the HIV Response in Countries with Generalized HIV Epidemics

   Funding Agency: NIH National Institute of Allergy and Infectious Diseases – NIAID

   Role: Principal Investigator for PSU portion

   Amount: $1,042,390 (PSU portion) Period: 7/29/2022 to 06/30/2026

   Project Number: R01 AI170249 (P.I.: Stefan Baral, Johns Hopkins University)

2. Project Title: Estimating Causes of Death in High Data Quality Settings: A Bayesian Hierarchical Model Approach

   Funding Agency: World Health Organization (WHO)

   Role: Principal Investigator

   Amount: $30,000 Period: 12/01/2023 to 12/31/2023

3. Project Title: Statistical Models for Estimating and Projecting HIV/AIDS Epidemics

Funding Agency: NIH National Institute of Allergy and Infectious Diseases – NIAID

Role: Principal Investigator

Amount: $3,585,943 Period: 9/25/2017 to 8/31/2024 (2nd year extension)

Project Number: R01 AI136664 (P.I.: Le Bao, Penn State University)

4. Project Title: Computation, Bioinformatics, and Statistics (CBIOS) Training

   Funding Agency: NIH National Institute of General Medical Sciences

   Role: Co-Investigator

   Amount: $2,737,901 Period: 7/01/2013 to 6/30/2023

   Project Number: R01 AI136664 (P.I.: Chiaromonte, Francesca, Penn State University)

5. Project Title: Detecting COVID-19 Outbreaks Using Travelers Data

   Funding Agency: COVID-19 Seed Grants (Institute for Computational and Data Sciences, Huck Institutes for the Life Sciences, Penn State University)

   Role: Principal Investigator

   Amount: $29,600 Period: 4/7/2020 to 8/31/2021

6. Project Title: Small Area Estimation for HIV Key Population Sizes

   Funding Agency: SOAR (Supporting Operational AIDS Research)

   Role: Co-Principal Investigator

   Amount: $74,640 Period: 7/1/2018 to 2/28/2019

   Project Number: OSP 204956 (P.I.: Xiaoyue Niu and Le Bao, Penn State University)

7. Project Title: Strengthening Capacity for Assessment of HIV- Related Data Needs among Key Populations to Inform Evidence- Based Responses

   Funding Agency: John Hopkins University (USAID Prime)

   Role: Co-Principal Investigator

   Amount: $59,045 Period: 07/01/2017 to 06/30/2018

   Project Number: OSP 195250 (P.I.: Xiaoyue Niu and Le Bao, Penn State University)

8. Project Title: Quantifying Spatial Representatives and Uncertainty in Antenatal Care Sentinel Surveillance for HIV in Sub-Saharan Africa

   Funding Agency: Imperial College of Science, Technology and Medicine (NIH Prime)

   Role: Co-Investigator

   Amount: $1,481 Period: 12/15/2016 – 11/30/2018

Project Number: R03 AI125001-01A1 (P.I.: Jeff Eaton, Imperial College London)

9. Project Title: Development of Methods to Produce Fine Scale Estimates of HIV Epidemics
   Funding Agency: The Joint United Nations Programme on HIV/AIDS (UNAIDS)
   Role: Principal Investigator
   Amount: $353,689 Period: 7/24/2016 to 3/15/2018
   Project Number: OSP 183485 (P.I.: Le Bao, Penn State University)
10. Project Title: HWRF Initialization by Comparing with and Adoption of the PSU WRFENKF Method
    Funding Agency: National Oceanic and Atmospheric Administration (NOAA)
    Role: Co-Investigator
    Amount: $414,500 Period: 07/01/2016 to 12/31/2017
    Project Number: G-00459-1 (P.I.: Fuqing Zhang, Penn State University)
11. Project Title: New Statistical Models for Estimating and Projecting HIV/AIDS Epidemics
    Funding Agency: NIH National Institute of Allergy and Infectious Diseases – NIAID
    Role: Principal Investigator
    Amount: $383,449 Period: 8/15/2016 to 7/31/2017
    Project Number: R56AI120812-01A1 (P.I.: Le Bao, Penn State University)
12. Project Title: Penn State Clinical and Translational Sciene Institute
    Funding Agency: National Center for Research Resources
    Role: Investigator
    Amount: $27,000,000 Period: 03/01/2011 to 02/28/2016
    Project Number: UL1 TR00127 (P.I.: Lawrance Sinoway, Penn State University)
13. Project Title: Diagnostics Modelling Consortium -- Evaluating Impacts of Incidence Assays
    Funding Agency: Imperial College London and Bill & Melinda Gates Foundation
    Role: Principal Investigator
    Amount: $30,000 Period: 06/01/2014 to 04/30/2015
    Project Number: 161794 AWARD (P.I.: Le Bao, Penn State University)

14. Project Title: Development of Hierarchical Models for Estimating Health Indicators
    Funding Agency: World Health Organization (WHO)
    Role: Principal Investigator
    Amount: $30,000 Period: 10/20/2014 to 06/19/2015
    Project Number: SPHQ14-APW-4034 (P.I.: Le Bao, Penn State University)

## Refereed Publications [+PhD or Postdoc trainee] [*corresponding author]

1. Yang, W., **Bao, L.***, Liu D. and Li R. (2024). A Likelihood Approach to Incorporating Self-Report Data in HIV Recency Classification. Biometrics. https://doi.org/10.1093/biomtc/ujae147
2. Zhang A.+, Li C.+, Daniels M. and **Bao L.*** (2024). Approximate Cross-validated Mean Estimates for Bayesian Hierarchical Regression Models. To appear in Journal of Computational and Graphical Statistics. https://doi.org/10.1080/10618600.2024.2404711
3. **Bao, L.*,** Niu, X., Imai-Eaton, J. and Brown T. (2024). Dynamic Models Augmented by Hierarchical Data: An Application of Estimating HIV Epidemics at Sub-National Level. *Biostatistics.* https://doi.org/10.1093/biostatistics/kxae003
4. Lan, Z.+ and **Bao, L.*** (2024) Multivariate spatial modelling for predicting missing HIV prevalence rates among key populations. *Journal of the Royal Statistical Society: Series A.* 18 (2), 321-337. https://doi.org/10.1093/jrsssa/qnad113
5. Rucinski K., Knight J., Willis K., Wang L., Rao A., Roach M.A., Phaswana-Mafuya R., **Bao L.**, Thiam S., Arimi P., Mishra S. and Baral S. (2024) Challenges and Opportunities in Big Data Science to Address Health Inequities and Focus the HIV Response. Curr HIV/AIDS Rep 21, 208–219. https://doi.org/10.1007/s11904-024-00702-3
6. **Bao, L.*,** Niu, X., Mahy M. and Ghys P.D. (2023) Estimating HIV Epidemics for Sub-national Areas. *Annals of Applied Statistics.* 17 (3), 2515-2532. http://dx.doi.org/10.1214/23-AOAS1730
7. Zhang Y.+, Chen S. and **Bao L.*** (2023) Air pollution estimation under air stagnation -- a case study of Beijing, China air pollution. *Envirometrics.* 34 (6); e2819. https://doi.org/10.1002/env.2819

8. Laga, I.[+], Niu, X., … and **Bao, L.*** (2023) Mapping the number of female sex workers in countries across sub-Saharan Africa. *Proceedings of the National Academy of Sciences.* 120 (2) e2200633120. https://doi.org/10.1073/pnas.2200633120
9. Laga, I.[+], **Bao, L.,** and Niu, X. (2023) A Correlated Network Scale-up Model: Finding the Connection Between Subpopulations. *Journal of the American Statistical Association.* 118 (543): 2515-2524. https://doi.org/10.1080/01621459.2023.2165929
10. Sheng B.[+], Li, C.[+], **Bao, L.*** & Li, R. (2023) Probabilistic HIV Recency Classification -- A Logistic Regression without Labeled Individual Level Training Data. *Annals of Applied Statistics.* 17 (1): 108-129. DOI: 10.1214/22-AOAS1618
11. **Bao, L.*,** Zhang, Y.[+] and Niu, X. (2022) What Can We Learn from the Travelers Data in Detecting Disease Outbreaks--A Case Study of the COVID-19 Epidemic. *Annals of Epidemiology.* 75: 67-72. https://doi.org/10.1016/j.annepidem.2022.09.005
12. Parsons, J.[+], Niu, X. & **Bao, L.*** (2022) A Bayesian hierarchical modeling approach to combining multiple data sources: A case study in size estimation. *Annals of Applied Statistics.* 16(3): 1550-1562. DOI: 10.1214/21-AOAS1556
13. **Bao, L.** Li, C.[+], Li, R., & Yang S.[+] (2022) Causal Structural Learning on MPHIA Individual Dataset. *Journal of the American Statistical Association.* 117.540, 1642-1655; https://doi.org/10.1080/01621459.2022.2077209.
14. Li, X.[+], Zhang, A.[+], Al-Zaidy, R., Baral, S., **Bao, L.*** and Giles, C. L. (2022) Automating document classification with distant supervision to increase the efficiency of systematic reviews: A case study on identifying studies with HIV impacts on female sex workers. *PLOS ONE.* 17(6): e0270034. doi.org/10.1371/journal.pone.0270034
15. VanEvery H., Yang W., Su J., Olsen N., **Bao L.**, Lu B., Wu S., Cui L., Gao X. (2022) Low density lipoprotein cholesterol and risk of rheumatoid arthritis: a prospective study. *Nutrients,* 14(6), 1240; https://doi.org/10.3390/nu14061240.
16. Laga I.[+], **Bao L.**, & Niu X. (2021) Thirty Years of The Network Scale-up Method. *Journal of the American Statistical Association.* 116:535, 1548-1559; https://doi.org/10.1080/01621459.2021.1935267

17. Laga I.[+], Niu X., & **Bao L.*** (2021) Modeling the Marked Presence-only Data: A Case Study of Estimating the Female Sex Worker Size in Malawi. *Journal of the American Statistical Association.* 117.537, 27-37; https://doi.org/10.1080/01621459.2021.1944873
18. Parsons, J.[+] and **Bao, L.*** (2021) A Unified Approach for Outliers and Influential Data Detection – The Value of Information in Retrospect. *Stat.* doi.org/10.1002/sta4.442
19. VanEvery H., Yang W., Olsen N., **Bao L.**, Lu B., Wu S., Cui L., Gao X. (2021) Alcohol consumption and risk of rheumatoid arthritis: a prospective study. *Nutrients,* 13(7), 2231; https://doi.org/10.3390/nu13072231.
20. Wu Z., Huang Z., Lichtenstein A., Liu Y., Chen S., Jin Y., Na M., **Bao L.**, Wu S. and Gao X. (2021) The Risk of Ischemic Stroke and Hemorrhagic Stroke in Chinese adults with low density lipoprotein cholesterol concentrations<70 mg/dL. *BMC Medicine,* 16;19(1):142. DOI:10.1186/s12916-021-02014-4.
21. Niu, X. M., Rao, A., Chen, D., Sheng, B., Weir, S., Umar, E., ... & **Bao, L.*** (2021). Using factor analyses to estimate the number of female sex workers across Malawi from multiple regional sources. *Annals of Epidemiology*, 55, 34-40. https://doi.org/10.1016/j.annepidem.2020.12.001
22. Parsons, J.[+], Niu, X., & **Bao, L.*** (2020). Evaluating the relative contribution of data sources in a Bayesian analysis with the application of estimating the size of hard to reach populations. *Statistical Communications in Infectious Diseases*, 12(s1): 20190020; https://doi.org/10.1515/scid-2019-0020.
23. Sheng B.[+], Eaton J., Mahy M. and **Bao L.*** (2020). Comparison of HIV Prevalence Among Antenatal Clinic Attendees Estimated from Routine Testing and Unlinked Anonymous Testing, *Statistics in Biosciences,* 12: 279–294; https://doi.org/10.1007/s12561-020-09265-4
24. Eaton J., Brown T., Puckett R., Glaubius R., Mutai K., **Bao L.**, Salomon J., Stover J. Mahy M., Hallett T. (2019). The Estimation and Projection Package Age-Sex Model and the r-hybrid model: new tools for estimating HIV incidence trends in sub-Saharan Africa, *AIDS*. 33: S235–S244. DOI: 10.1097/QAD.0000000000002437

25. Datta A., Lin W., Rao A., Diouf D., Edwards J., **Bao L.**, Louis T. and Baral S. (2018). Bayesian estimation of MSM population size in Cote d'Ivoire, *Statistics and Public Policy*. 6(1): 1–13. DOI: 10.1080/2330443X.2018.1546634
26. Huang S., Li J., Wu Y., Ranjbar S., Xing A., Zhao H., Wang Y., Shearer G. C., **Bao L.**, Lichtenstein A. H., Wu S. and Gao X. (2018). Tea consumption and longitudinal change in high-density lipoprotein cholesterol concentration in Chinese adults, *Journal of the American Heart Association*. 7, 13, e008814. https://doi.org/10.1161/JAHA.118.008814
27. Cheng F.W., Gao X., **Bao L.**, Mitchell D.C., Wood C., Sliwinski M.J., Smiciklas-Wright H., Still C.D., Rolston D.D.K., and Jensen G.L. (2017). Obesity as a risk factor for developing functional limitation among older adults: A conditional inference tree analysis. *Obesity (Silver Spring)*. 25(7):1263-1269. DOI: 10.1002/oby.21861
28. Wu Z., Su X., Sheng H., Chen Y., Gao X., **Bao L.**, Jin W. (2017) Conditional Inference Tree for Multiple Gene-Environment Interactions on Myocardial Infarction Among Chinese Men. *Archives of Medical Research*. 48(6):546-552. doi.org/10.1016/j.arcmed.2017.12.001
29. Eaton J. and **Bao L.** (2017). Accounting for non-sampling error in estimates of HIV epidemic trends from antenatal clinic sentinel surveillance. *AIDS* 31: S61-S68. DOI: 10.1097/QAD.0000000000001419
30. Niu X., Zhang A.X.$^{+}$, Brown T., Puckett R., Mahy M., **Bao L.*** (2017). Incorporation of hierarchical structure into EPP fitting with examples of estimating sub-national HIV/AIDS dynamics. *AIDS* 31: S51-S59. doi: 10.1097/QAD.0000000000001426
31. Sheng B.$^{+}$, Marsh K., Slavkovic A.B., Gregson S., Eaton J., **Bao L.*** (2017). Statistical Models for Incorporating Data from Routine HIV Testing of Pregnant Women at Antenatal Clinics into HIV/AIDS Epidemic Estimates. *AIDS* 31: S87-S94. doi: 10.1097/QAD.0000000000001428
32. Hunter D.R., **Bao L.**, and Poss M. (2017). Assignment of Endogeneous Retrovirus Integration Sites Using a Mixture Model. *Annals of Applied Statistics* 11(2): 751-770. DOI: 10.1214/16-AOAS1016
33. Thomas J. and **Bao L.** (2016). Modeling the dynamics of an HIV epidemic. *Dynamic Demographic Analysis.* 91-144. DOI:10.1007/978-3-319-26603-9_6

34. Malhotra, R., Elleder, D., **Bao, L.**, Hunter, D. R., Poss, M., Acharya, R. (2016). A pipeline for identifying integration sites of mobile elements in the genome using next-generation sequencing. *Proceedings of the 8th International Conference on Bioinformatics and Computational Biology.* (BICOB 2016): 63-69.
35. Li R.$^{+}$, Dudek S.M., Kim D., Hall M.A., Bradford Y., Peissig P.L., Brilliant M.H., Linneman J.G., McCarty C.A., **Bao L.**, and Ritchie M.D. (2016) Identification of genetic interaction networks via an evolutionary algorithm evolved Bayesian network. *Bio Data Mining,* 9(18). DOI: 10.1186/s13040-016-0094-4.
36. **Bao L.***, Raftery A.E., Reddy A. (2015) Estimating the sizes of populations at risk of HIV infection from multiple data sources using a Bayesian hierarchical model. *Statistics and Its Interface*. 8(2): 125-136. doi: 10.4310/SII.2015.v8.n2.a1
37. **Bao L.**, Elleder D., Malhotra R., DeGiorgio M., Maravegias T., Horvath L., Carrel L., Gillin C., Hron T., Fábryová H., Hunter D. and Poss M. (2014) Computational and statistical analyses of insertional polymorphic endogenous retroviruses in a non-model organism. *Computation*. 2: 221-245. https://doi.org/10.3390/computation2040221
38. **Bao L**.*, Ye J.$^{+}$, Hallett T.B. (2014) Incorporating incidence information within the UNAIDS Estimation and Projection Package framework: A study based on simulated incidence assay data. *AIDS* 28: S515-S522. DOI: 10.1097/QAD.0000000000000434
39. Brown T., **Bao L.**, Eaton J.W., Hogan D.R., Mahy M., Marsh K., Mathers B.M., Puckett R. (2014) Improvements in prevalence trend fitting and incidence estimation in EPP 2013. *AIDS* 28: S415-S425. DOI: 10.1097/QAD.0000000000000454
40. Kamath P., Elleder D., **Bao L.**, Cross P., Poss M. (2013) The population history of endogenous retroviral elements in mule deer (Odocoileus hemionus). *Journal of Heredity*, 105: 173-187. DOI: 10.1093/jhered/est088
41. **Bao L.*** (2012). A new infectious disease model for estimating and projecting HIV/AIDS epidemics. *Sexually Transmitted Infections,* 88: i58-i65. doi: 10.1136/sextrans-2012-050689
42. **Bao L.***, Salomon J.A., Brown T., Raftery A.E., and Hogan D.R. (2012). Modelling national HIV/AIDS epidemics: Revised approach in the UNAIDS Estimation and

Projection Package 2011. *Sexually Transmitted Infections,* 88: i3-i10. http://dx.doi.org/10.1136/sextrans-2012-050637

43. Clark S.J., Thomas J., and **Bao L.** (2012). Estimates of age-specific reductions in HIV Prevalence in Uganda: Bayesian melding estimation and probabilistic population forecast with an HIV-enabled cohort component projection model. *Demographic Research* 27: 743-774. DOI: 10.4054/DemRes.2012.27.26
44. Meila M. and **Bao L.** (2010). An exponential model for infinite rankings. *Journal of Machine Learning Research*, 11: 3481-3518. http://jmlr.org/papers/v11/meila10a.html
45. Raftery A.E. and **Bao L.** (2010). Estimating and projecting trends in HIV/AIDS generalized epidemics using incremental mixture importance sampling. *Biometrics,* 66: 1162-1173. DOI: 10.1111/j.1541-0420.2010.01399.x
46. **Bao L.**, Raftery A.E. (2010). A stochastic infection rate model for estimating and projecting national HIV prevalence rates. *Sexually Transmitted Infections,* 86: ii93-ii99. http://dx.doi.org/10.1136/sti.2010.044529
47. Brown T., **Bao L.**, Raftery A.E., Salomon J.A., Baggaley R.F., Stover J., and Gerland P. (2010). EPP 2009: Bringing the UNAIDS Estimation and Projection Package into the ART era. *Sexually Transmitted Infections,* 86: ii3-ii10. http://dx.doi.org/10.1136/sti.2010.044784
48. **Bao L.**, Gneiting T., Grimit E., Guttrop P. and Raftery A.E. (2010). Bias correction and Bayesian model averaging for ensemble forecasts of surface wind direction. *Monthly Weather Review*, 138:1811-1821. https://doi.org/10.1175/2009MWR3138.1
49. **Bao L.,** Zhu, Z. and Ye, J. (2009). Modeling oncology gene pathways network with multiple genotypes and phenotypes via a copula method. *IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology,* 237-246. DOI: 10.1109/CIBCB.2009.4925734
50. Meila M. and **Bao L.** (2008). Estimation and clustering with infinite rankings. *Proceedings of the 24th Conference in Uncertainty in Artificial Intelligence,* 24:393-402. http://uai2008.cs.helsinki.fi/UAI_camera_ready/meila.pdf
51. **Bao L.**, Gu H., Dunn, K.A. and Bielawski J. (2008). Likelihood based clustering (LiBaC) for codon models, a method for grouping sites according to similarities in the underlying

process of evolution. *Molecular Biology and Evolution.* 25:1995-2007. https://doi.org/10.1093/molbev/msn145

52. **Bao L.**, Gu H., Dunn K.A. and Bielawski J. (2007). Methods for selecting fixed-effect models for heterogeneous codon evolution, with comments on their application to gene and genome data. *BMC Evolutionary Biology*. 7 Suppl 1:S5. doi:10.1186/1471-2148-7-S1-S5
53. Mitnitski A, **Bao L.** and Rockwood K. (2007). A cross-national study of transitions in deficit counts in two birth cohorts: Implications for modeling ageing. *Experimental Gerontology.* 42:241-246. DOI: 10.1016/j.exger.2006.10.001
54. Mitnitski A, **Bao L.** and Rockwood K. (2006). Going from bad to worse: a stochastic model of transitions in deficit accumulation, in relation to mortality. *Mechanisms of Ageing and Development.* 127: 490-493. DOI: 10.1016/j.mad.2006.01.007

## Reports:

55. UNAIDS (2015). Guidelines on monitoring the impact of the HIV epidemic using population-based surveys. http://www.unaids.org/en/resources/documents/2015/population_based_surveys
56. WHO (2015). Tracking universal health coverage: First global monitoring report. http://www.who.int/healthinfo/universal_health_coverage/report/2015/en/
57. **Bao L.,** Hogan D., Raftery A.E., and Salomon J.A. (2011). A flexible model for estimating and projecting trends in HIV/AIDS epidemics. *Technical Report, UNAIDS, Geneva*.

## Reviews and Comments:

58. **Bao L.**, (2015). Comment on Quasi-Monte Carlo sampling by Gerber and Chopin. *Journal of the Royal Statistical Society: Series B*, 77: 560.
59. **Bao L.**, Fricks J. and Haran M. (2012). Comment on the mechanistic modeling and inference for cell motility by Manolopoulou et al. *Journal of the American Statistical Association*, 107: 869−871.

## Statistical Software Packages:

1. IMIS: R-package for Incremental Mixture Importance Sampling.

2. SizeEstimation: R-package for estimating the size of hidden population with multiple data sources.
3. Estimation and Projection Package (EPP): it is used to estimate and project adult HIV prevalence and incidence from surveillance data.
4. Codeml_FE: Comprehensive set of fixed-effect models of codon evolution.
5. LiBaC: Clustering analysis under a probabilistic model of codon evolution.

## Invited talks:

*1.* Conference on Applied, Bayesian, Computational, Demographic and Environmental Statistics (ABCDE), University of Washington, 08/09/2024, *Dynamic Models Augmented by Hierarchical Data: An Application of Estimating HIV Epidemics at Sub-National Level*

*2.* Joint Conference on Statistics and Data Science, Kunming, China, 07/13/2023, *Dynamic Models Augmented by Hierarchical Data*

*3.* Hangzhou International Conference on Frontiers of Data Science, Zhejiang University, China, 07/10/2024, *Clinical Trial Recommendation Model with Large Language Model*

*4.* Institute of Statistics and Big Data, Renmin University of China, 03/26/2024, *Dynamic Models Augmented by Hierarchical Data*

*5.* Center for Statistical Science, Peking University of China, 03/07/2024, *Dynamic Models Augmented by Hierarchical Data*

*6.* Kenya Data Utilization Workshop, 11/08/2023, *Statistical Methods for Mapping the Distribution of Marginalized Groups*

*7.* Joint Conference on Statistics and Data Science, Beijing, China, 07/11/2023, *Mapping the Marginalized Groups with Presence-Only Data*

*8.* WNAR, Anchorage, Alaska, 06/19/2023, *A case-control sampling strategy for zero-inflated models with an application to female sex worker mapping in sub-Saharan Africa*

*9.* The ICSA 2023 Applied Statistics Symposium, University of Michigan, Ann Arbor, 06/13/2023, *Mapping the Marginalized Groups with Presence-Only Data*

*10.* UNAIDS Reference Group on Estimates, Modelling and Projections Meeting, Stellenbosch, South Africa, 05/17/2023, *Statistical methods for key population indicators in sub-Saharan Africa*

11. Academy of Mathematics and Systems Science, Beijing, China, 02/13/2023, *Probabilistic HIV Recency Classification -- A Logistic Regression without Individual Level Training Data*
12. Biostatistics, Epidemiology and Research Design Senimar, CTSI, Penn State University, 01/30/2023, *Statistical Methods for Mapping the Distribution of Marginalized Groups*
13. Social Data Analytics (SoDA) talk, Penn State University, 10/25/2022, *Statistical Methods for Mapping the Distribution of Marginalized Groups*
14. Department of Statistics, University of Florida, 11/11/2022, *Everyone Counts: Advanced Methods for Estimating Marginalized Populations*
15. Department of Statistics, Xia Men University, 10/14/2022, *Everyone Counts: Advanced Methods for Estimating Marginalized Populations*
16. UNAIDS Reference Group on Estimates, Modelling and Projections Meeting on "Key population stratified estimates across all HIV epidemic settings", Online Meeting, 04/20/2021, *Size Estimates of Female Sex Workers in Sub-Saharan Africa*
17. Centers for Disease Control and Prevention, Atlanta, 04/14/2021, *Probabilistic HIV Recency Classification -- A Logistic Regression without Labeled Individual Level Training Data*
18. International Society for Bayesian Analysis, 2021 World Meeting, 06/30/2021, *Mapping the population size of female sex worker in countries across sub-Saharan Africa*
19. Center for Infectious Disease Dynamics, Penn State University, 1/14/2021, *Detecting Disease Outbreak Using Travelers Data -- A Case Study of COVID-19 Epidemic*
20. 2020 JSM, Virtual Conference, 08/04/2020, *Detecting Disease Outbreak Using Travelers Data -- A Case Study of COVID-19 Epidemic*
21. Center for Infectious Disease Dynamics, Penn State University, 1/9/2020, *Assessing Data Contribution for Estimation of HIV Epidemics*
22. Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine, The University of Pennsylvania, 10/17/2019, *Assessing Data Contribution for Estimation of HIV Epidemics - The Value of Information Method*
23. 2019 JSM, Denver, 08/01/2019, *Assessing Data Contribution for Estimation of HIV Epidemics - The Value of Information Method*

*24.* Saw Swee Hock School of Public Health, National University of Singapore, 7/8/2019, *Value of Information in Retrospect*

*25.* 2019 Hangzhou International Conference on Frontiers of Data Science, Hangzhou, China, 5/26/2019, *The Value of Information Method*

*26.* Workshop "Statistical Challenges and Opportunities in HIV/AIDS Research in the Era of Getting-to-Zero HIV Infections", Philadelphia, 3/23/2019, *Modeling the Recovery of Immune Function in HIV Patients by a Hidden Markov Model*

*27.* Workshop "Statistical Challenges and Opportunities in HIV/AIDS Research in the Era of Getting-to-Zero HIV Infections", Philadelphia, 3/23/2019, *HIV prevalence in key populations: A semiparametric Bayesian hierarchical model for scarce and imbalanced data*

*28.* Department of Biostatistics, School of Public Health, Peking University, China, 2/27/2019, *Leaving No One Behind: Size Estimation of People at High Risk for HIV Infections*

*29.* Academy of Mathematics and System Sciences, Chinese Academy of Sciences, 1/7/2019, *Approaches to population size estimation and opportunities for leveraging existing data*

*30.* Center for Statistical Science, Peking University, China, 11/1/2018, *Leaving No One Behind: Size Estimation of People at High Risk for HIV Infections*

*31.* Center for Statistical Science, Tsinghua University, China, 10/29/2018, *Leaving No One Behind: Size Estimation of People at High Risk for HIV Infections*

*32.* Machine Learning: Dimension reduction, Summer School in Astroinformatics, 06/04/2018, Center for Astrostatistics, PSU.

*33.* ICSA 2018 Applied Statistics Symposium, New Brunswick, New Jersey, 06/15/2018, *Size Estimation of People at High Risk for HIV Infections*

*34.* UNAIDS Reference Group on Estimates, Modelling and Projections Meeting on "New data, tools, and methods for estimating HIV incidence patterns and trends", Atlanta, USA, 05/31/2018, *Understand Contribution of Key Populations in HIV/AIDS epidemics*

*35.* UNAIDS Reference Group on Estimates, Models and Projections, London, 10/17/2017, *Discussion of the Bayesian Hierarchical Models for Size Estimation of Key Populations*

*36.* 2017 JSM, Baltimore, 08/01/2017, *Statistical Models for Estimating HIV/AIDS Epidemics with Multiple Types of Prevalence Data*

*37.* International Conference on Big Data in Biological and Medical Sciences, Xi'An, China, 07/04/2017, *Identification of endogenous retrovirus integration sites using a mixture model*

*38.* Department of Financial Mathematics, Peking University, China, 06/15/2017, *Challenges of Quantitative Analysis in Global Health*

*39.* Academy of Mathematics and System Sciences, Chinese Academy of Sciences, 06/07/2017, *Challenges of Quantitative Analysis in Global Health*

*40.* UNAIDS Reference Group on Estimates, Models and Projections, Geneva, 05/15/2017, *Missing Data Issue in HIV Surveillance Data*

*41.* Small Area Estimation and Spatial Statistics Working Group, Department of Biostatistics, Johns Hopkins University, 03/09/2017, *Leaving No One Behind: Estimating HIV Epidemics at Sub-National and Sub-Population Level*

*42.* Centers for Disease Control and Prevention, Atlanta, 01/31/2017, *Estimating the size of populations at high risk of HIV using a Bayesian hierarchical model*

*43.* UNAIDS Reference Group on Estimates, Models and Projections, New York, 11/08/2016, *Modelling using a Hierarchical Approach; Incidence Assay Update*

*44.* Latent Variables 2016 Conference, University of South Carolina, 10/13/2016, *Identification of endogenous retrovirus integration sites using a mixture model*

*45.* Department of Statistics, The Ohio State University, 10/06/2016, *Incorporating Hierarchical Structure into Dynamic Systems: An Application of Estimating HIV Epidemics at Sub-National and Sub-Population Level*

*46.* Department of Statistics, UCLA, 10/04/2016, *Incorporating Hierarchical Structure into Dynamic Systems via Auxiliary Data Approach*

*47.* ADAPT seminar, The Pennsylvania State University, 09/01/2016, *Incorporating Hierarchical Structure into Dynamic Systems via Auxiliary Data Approach*

*48.* 2016 JSM, Chicago, 08/03/2016, *Estimating HIV Epidemics at Fine Scales*

*49.* The Third Taihu International Statistics Forum, Shanghai, China, 07/10/2016, *Statistical models for Estimating and Predicting HIV Epidemics*

*50.* The School of Mathematics, Peking University, 06/27/2016, *Statistical models for Estimating and Predicting HIV Epidemics*

*51.* The School of Economics, Xiamen University, 06/22/2016, *Statistical models for Estimating and Predicting HIV Epidemics*

*52.* National Center for AIDS/STD Control and Prevention, China CDC, 06/16/2016, *Estimating HIV Epidemics at Fine Scales*

*53.* Hershey/Institute for Personalized Medicine, The Pennsylvania State University, 04/29/2016, *Network-based Discriminant Analysis*

*54.* CTSI BERD seminar, The Pennsylvania State University, 03/29/2016, *Leaving No One Behind -- Estimating HIV Epidemics at Fine Scales*

*55.* Department of Global Health, University of Washington, 02/29/2016*, Estimating HIV Epidemics at Sub-National and Sub-Population Level*

*56.* Department of Statistics, Temple University, 11/20/2015*, An Efficient Way of Estimating HIV Epidemics in Sub-National Areas and Sub-Populations*

*57.* UNAIDS Reference Group on Estimates, Models and Projections, London, 10/26/2015, *Incorporating PMTCT data in EPP: Investigation of Level of Continuity Required, Data Quality Requirements; Further Testing of the Hierarchical Model*

*58.* Department of Statistics, Columbia University, 09/28/2015*, Estimating HIV Epidemics for Sub-National Areas*

*59.* Big Data Social Science IGERT program, The Pennsylvania State University, 09/09/2015*, Estimating HIV Epidemics for Sub-National Areas*

*60.* Bioinformatics and Genomics Retreat, The Pennsylvania State University, 08/29/2015*, Assigning Viruses from Sequence Count Data via a Mixture Model*

*61.* UNAIDS Reference Group on Estimates, Models and Projections, Boston, 06/04/2015, *Hierarchical Model - Revisions, Accuracy, Efficiency, Use with Key Populations*

*62.* UNAIDS Reference Group on Estimates, Models and Projections, Boston, 06/03/2015, *Incorporating PMTCT Data into EPP Fitting*

*63.* ENAR, Miami, 03/17/2015, *Compression of Complex Data with an Example of Time Series Gene Expression and Biomarker*

64. Department of Genomics, The Pennsylvania State University, 02/27/2015, *Compression of Complex Data with an Example of Time Series Gene Expression*
65. Bill & Melinda Gates Foundation, Seattle, 02/17/2015, *Towards a Target Product Profile for HIV Incidence Assay Development*
66. Information Communication Technology for Development (ICT4D), The Pennsylvania State University, 12/17/2014, *HIV Epidemics Study*
67. The Methodology Center, The Pennsylvania State University, 11/13/2014, *A Hierarchical Model for Estimating HIV/AIDS Epidemics*
68. UNAIDS Reference Group on Estimates, Models and Projections, UNAIDS, Geneva, 10/29/2014, *Considerations for Incorporating PMTCT as an Additional Data Source with Calibrating Parameter*
69. UNAIDS Reference Group on Estimates, Models and Projections, UNAIDS, Geneva, 10/28/2014, *Hierarchical Approach for Generating Sub-National Estimates within the EPP Framework*
70. UNAIDS Reference Group on Estimates, Models and Projections, Seattle, 04/25/2014, *Bayesian Hierarchal Model for Sharing Information Across Areas and Countries*
71. Institute for Health Metrics and Evaluation, University of Washington, Seattle, 04/23/2014, *Statistical Models for Estimating and Predicting HIV/AIDS Epidemics*
72. National Institutes of Health (NIH), NICHD, Bethesda, Maryland, 02/20/2014, *Compression of Complex Data with an Example of Time Series Gene Expression*
73. UNAIDS Reference Group on Estimates, Models and Projections, Spain, 08/12/2013, *Use of Incidence Assays within the EPP framework*
74. The 2nd Taihu International Statistics Forum, Soo Chow University, China, 07/07/2013, *Inference of Gene Associations using Model-based Clustering and Adjusted Rand Index in Time-course Gene Expression Data*
75. IMS-China, International Conference on Statistics and Probability, Chengdu, China, 07/02/2013, *Inference of Gene Associations using Model-based Clustering and Adjusted Rand Index in Time-course Gene Expression Data*
76. ENAR, Orlando, FL, 03/11/2013, *A New Infectious Disease Model for Estimating and Projecting HIV/AIDS Epidemics*

77. Working group on Stochastic Modelling and Computational Statistics, State College, PA, 11/15/2012, *Incremental Mixture Importance Sampling*
78. UNAIDS Reference Group on Estimates, Models and Projections, London, 09/24/2012, *A New Infectious Disease Model for Estimating and Projecting HIV/AIDS Epidemics*
79. Second Biostatistics Symposium, Renmin University, Beijing, 07/09/2012, *New Methods for Estimating and Projecting National HIV/AIDS Prevalence Rates*
80. Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, 06/15/2012, *Incremental Mixture Importance Sampling for Estimating and Predicting HIV/AIDS Epidemics*
81. Department of Mathematics and Statistics, University of Minnesota, Duluth, MN, 05/04/2012, *Statistical Models for Estimating and Predicting HIV/AIDS Epidemics*
82. UNAIDS Reference Group on Estimates, Models and Projections, Boston, MA, 04/02/2012, *New Improvements on R-Flexible Model*
83. Student Advisory Committee (SAC) Seminar, Department of Statistics, The Pennsylvania State University, University Park, PA, 03/21/2012, *New Methods for Estimating and Projecting National HIV/AIDS Prevalence Rates*
84. Department of Statistics, University of Washington, Seattle, WA, 05/17/2011, *Statistical Models for Estimating and Projecting HIV/AIDS Epidemics*
85. Center of Statistical Science, Peking University, Beijing, China, 04/12/2011, *Statistical Models for Estimating and Projecting HIV/AIDS Epidemics*
86. Department of Biostatistics and Bioinformatics, Emory University, Atlanta, 03/31/2011, *Statistical Models for Estimating and Projecting HIV/AIDS Epidemics*
87. Department of Statistics, The Pennsylvania State University, University Park, 02/10/2011, *Statistical Models for Estimating and Projecting HIV/AIDS Epidemics*
88. Department of Statistics, National University of Singapore, Singapore, 01/21/2011, *Estimating HIV At-Risk Population Size Using a Bayesian Hierarchical Model*
89. UNAIDS Reference Group on Estimates, Models and Projections, Boston, 10/26/2010, *A Flexible Model for Estimating and Projecting HIV Prevalence Rates*
90. Joint Statistical Meetings, Vancouver, Canada, 08/01/2010, *The Bayesian Hierarchical Model for Estimating the Size of HIV At-Risk Populations in Bangladesh*

*91.* UNAIDS Reference Group on Estimates, Models and Projections, Glastonbury, 05/21/2010, *A Stochastic Infection Rate Model for Estimating and Projecting National HIV Prevalence Rates*

*92.* UNAIDS Reference Group on Estimates, Models and Projections, London, 10/14/2009, *The Bayesian Model for Estimating the Size of HIV At-Risk Population*

*93.* Working group on Model-Based Clustering Summer Session, Seattle, WA, 07/10/2008, *Incremental Mixture Importance Sampling*

*94.* Pacific Northwest Weather Workshop, Seattle, WA, 02/29/2008, *Probabilistic Wind Direction Forecasting Using Bayesian Model Averaging*

## Student Supervision:

*Current Ph.D. Students at The Pennsylvania State University*

Tengjie Tang (Department of Statistics)

Ying Zhang (Department of Statistics)

Wenlong Yang (co-advised with Runze Li, Department of Statistics)

Yifan Jiang (co-advised with Runze Li, Department of Statistics)

Xihui Xu (Department of Statistics)

Ryan Halstater (Department of Statistics)

Jiahao Zhang (Department of Statistics)

Yuankai Ma (co-advised with Runze Li, Department of Statistics)

*Graduated Ph.D. Students*

Sanam Sanei (Department of Statistics, 2024)

Ian Laga (co-advised with Xiaoyue Niu, Department of Statistics, 2022)

Ben Sheng (Department of Statistics, 2021)

Amy Zhang (Department of Statistics, 2021)

Jacob Parsons (co-advised with Xiaoyue Niu, Department of Statistics, 2019)

Jingyi Ye (co-advised with Runze Li, Department of Statistics, 2017)

*Postdoc Advisees*

Lei Wang (Department of Statistics, 2023-2024)

Changcheng Li (Department of Statistics, 2019-2021)

Zhou Lan (Department of Statistics, 2018-2019)

*Ph.D. Thesis Committees at The Pennsylvania State University*

Vivian Cheng (Department of Statistics, 2024)

Alex Zhao (Department of Statistics, 2022)

Xiaoxiao Li (Department of Statistics, 2022)

Hannah VanEvery (Department of Nutrition Science, 2021)

Dhanushi Wijeyakulasuriya (Department of Statistics, 2020)

Chaoran Ma (Department of Nutrition Science, 2020)

Xiaoheng Cheng (Department of Biology, 2020)

Yuji Samizo (Department of Statistics, in progress)

Jaewoo Park (Department of Statistics, 2019)

ChingChi Yang (Department of Statistics, 2019)

Theodora Kaiser (Department of Biology, 2018)

Ruowang Li (Department of Genomics, 2016)

Wenjie Hu (College of Information Sciences and Technology, 2016)

Joshua Goldstein (Department of Statistics, 2015)

Xin Chen (College of Information Sciences and Technology, 2015)

Rashmi Bomiriya (Department of Statistics, 2014)

Xiaotian Zhu (Department of Statistics, 2014)

Yihan Li (Department of Statistics, 2014)

Wenyu Hua (Department of Statistics, 2014)

Qi Fang (College of Information Sciences and Technology, 2013)

Wei Wang (Department of Statistics, 2013)

James Yonamine (Department of Political Science, 2013)

*Undergraduate Honor Thesis at The Pennsylvania State University*

Angkai Li (Department of Statistics, 2024)

Yuxuan Wu (Department of Statistics, 2024)

Grant Thomas Hopkins (Department of Statistics, 2022)

David Chen (Department of Statistics, 2021)

Yiyang Wang (Department of Statistics, 2020)

Shanglun Li (Department of Statistics, 2018)

Yuan Tang (Department of Mathematics, 2015)

## Teaching:

1. STAT/IST 557, Data Mining: A PhD-level course introducing data mining, statistical/machine learning, and their applications in large complex data sets.
2. STAT554, Categorical Data Analysis: A PhD-level course discussing how to formulate, fit, select, and assess a wide variety of models for discrete data.
3. STAT897D, Applied Data Mining: A master-level course providing techniques and software to automate big data analysis.
4. STAT/MATH415, Introduction to Mathematical Statistics: An undergraduate-level course introducing statistical inference, e.g., estimation, hypothesis testing, regression, non-parametric statistics, and Bayesian statistics.
5. STAT497, Introduction to Statistical Research: An undergraduate-level course providing new researchers the necessary skills to engage in independent research projects, critically evaluate research outcomes, and effectively communicate research findings.

## Professional Memberships:

Institute of Mathematical Statistics

American Statistical Association

International Chinese Statistical Association

International Society for Bayesian Analysis

## Referee:

Journals in Statistics: Annals of Applied Statistics, Biostatistics, Journal of the American Statistical Association, Journal of the Royal Statistical Society: Series B, Journal of Computational and Graphical Statistics, Journal of Machine Learning Research, Computational Statistics and Data Analysis, Journal of Agricultural, Biological, and Environmental Statistics, Pattern Recognition, Statistics and Computing, Statistics and Its Interface, Statistics in Medicine, Technometrics.

Journals in Bioinformatics, Epidemiology, Global Health and Social Science: BMC Genetics, BMC Bioinformatics, BioData Mining, Bioinformatics, PLOS ONE, Epidemiology, International Journal of Infectious Diseases, Journal of Acquired Immune Deficiency

Syndromes, Journal of AIDS and Clinical Research, Journal of the International AIDS Society, Statistical Communications in Infectious Diseases, Sexually Transmitted Infections, The Lancet HIV, Annals of LGBTQ Public and Population Health, PLOS Global Health, Sociological Methods and Research, The Proceedings of the National Academy of Sciences (PNAS), The Journal of Nutrition, Health and Aging, Population Health Metrics, The South Pacific Journal of Natural and Applied Sciences.