

GUOSHENG YIN

Chair Professor, Patrick SC Poon Endowed Professor
Department of Statistics and Actuarial Science
University of Hong Kong, Pokfulam Road, Hong Kong
Tel: 852-3917-8313 Fax: 852-2858-9041 Email: gyin@hku.hk

PUBLICATIONS

Book

1. **Yin, G.** (2012). *Clinical Trial Design: Bayesian and Frequentist Adaptive Methods* by John Wiley & Sons (Wiley Series in Probability and Statistics). Hoboken, New Jersey, USA. Japanese version translated by Teramukai, S. and Daimon, T. (2014), Medical Publications, Tokyo, Japan.
2. **Yin, G.** and Shi, H. (2018). *Statistical Design and Analysis in Clinical Trials* (in Chinese). Higher Education Press, China.

AI and Machine Learning Conferences

3. Liu, B. and **Yin, G.*** (2024). Enhancing Semi-Supervised Domain Adaptation via Effective Target Labeling. The 38th Annual AAAI (Association for the Advancement of Artificial Intelligence) Conference on Artificial Intelligence February, 2024, Vancouver, Canada.
4. Hu, M., Liu, B. and **Yin, G.** (2023). Multi-site and multi-pollutant air quality data modeling. *Sustainability* **16** (1), 165.
5. Liu, B. and **Yin, G.** (2023). Graphmax for text generation. *Journal of Artificial Intelligence Research* **78**, 823–848.
6. Chan, T. H., Lau, K. W., Shen, J., **Yin, G.** and Yu, L. (2023). Adaptive Uncertainty Estimation via High-Dimensional Testing on Latent Representations. NeurIPS (Neural Information Processing Systems) December, 2023, New Orleans, USA.
7. Liu, B., He, J., Li, Z., Huang, X., Zhang, X. and **Yin, G.** (2023). Interpret ESG Rating’s Impact on the Industrial Chain Using Graph Neural Networks. IJCAI 2023. The 32nd International Joint Conference on Artificial Intelligence. August 19–25, 2023. Macao, S.A.R.
8. Chan, T. H., Cendra, F. J., Ma, L., **Yin, G.** and Yu, L. (2023). Histopathology Whole Slide Image Analysis with Heterogeneous Graph Representation Learning. CVPR 2023. Computer Vision and Pattern Recognition Conference. June 18–22, 2023. Vancouver, Canada.
9. Gu, J. and **Yin, G.** (2023). 3D-Polishing for Triangular Mesh Compression of Point Cloud Data. *KDD*. 29th SIGKDD Conference on Knowledge Discovery and Data Mining. August 6–10, 2023. Long Beach, California, USA.
10. Chan, T., Wong, C., Shen, J. and **Yin, G.** (2022). Source-aware embedding training on heterogeneous information networks. *Data Intelligence*. In press.

11. Du, W., Jin, H., Yu, C. and **Yin, G.** (2022). Deep reinforcement learning for bandit arm localization. 2022 IEEE International Conference on Big Data, Osaka, Japan, 2022 pp. 5243–5252. doi: 10.1109/BigData55660.2022.10020647
12. Tan, Z., Liu, B., and **Yin, G.** (2022). Asymmetric graph representation learning. 2022 IEEE International Conference on Big Data, Osaka, Japan, 2022 pp. 1369–1376. doi: 10.1109/BigData55660.2022.10021033
13. Chan, T., Wong, C., Shen, J. and **Yin, G.** (2022). SUMSHINE: Scalable unsupervised multi-source heterogeneous information network embeddings. 2022 IEEE International Conference on Knowledge Graph (ICKG), 32–39.
14. Liu, C., Cui, J., Gan, D., and **Yin, G.*** (2021). Beyond COVID-19 diagnosis: Prognosis with hierarchical graph representation learning. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) 2021.
15. Gu, J. and **Yin, G.*** (2021). Crystallization learning with the Delaunay triangulation. International Conference on Machine Learning (ICML) 2021 Spotlight. *Proceedings of the 38th International Conference on Machine Learning* **139**, 3854–3863, 2021.
16. Liu, B., Wang, L. and **Yin, G.** (2020). Learning distributed sentence vectors with bi-directional 3D convolutions. The 28th International Conference on Computational Linguistics (COLING) 2020.
17. Liu, B., Gao, X., He, M., Liu, L. and **Yin, G.** (2020). Online COVID-19 diagnosis with chest CT images: Lesion-attention deep neural networks. 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) 2020.
18. Liu, B. and **Yin, G.** (2020). Ventilation prediction for ICU patients with LSTM-based deep relative risk model. The 14th International Conference on Robotics and Artificial Intelligence (FLINS) 2020.
19. Liu, B. and **Yin, G.** (2020). Chinese document classification with bi-directional convolutional language model. Proceedings of the 43rd International ACM SIGIR, Conference on Research and Development in Information Retrieval (SIGIR) 2020.
20. Ma, D., Liu, B., Qiao, X., Cao, D. and **Yin, G.** (2020). Coarse-to-fine framework for music generation via generative adversarial networks. HPCCT & BDAI 2020: Proceedings of the 4th High Performance Computing and Cluster Technologies Conference & 3rd International Conference on Big Data and Artificial Intelligence.
21. Liu, Y. and **Yin, G.*** (2019). Nonparametric functional approximation with Delaunay triangulation learner. IEEE International Conference on Big Knowledge (ICBK) 2019.
22. Gu, J. and **Yin, G.*** (2019). Fast algorithm for generalized multinomial models with ranking data. 36th International Conference on Machine Learning (ICML) 2019. *Proceedings of Machine Learning Research* **97**, 2445–2453.
23. Zhang, C. and **Yin, G.*** (2019). Fast and stable maximum likelihood estimation for incomplete multinomial model. 36th International Conference on Machine Learning (ICML) 2019. *Proceedings of Machine Learning Research* **97**, 7463–7471.

24. Jiang, F., **Yin, G.** and Dominici, F. (2018). Bayesian model selection approach to boundary detection with non-local priors. *Proceeding of the 32nd Conference on Neural Information Processing Systems (NeurIPS) 2018*, 1978–1987.

Statistical Methodologies

25. Fang, J. and **Yin, G.*** (2024). Fractional accumulative calibration-free odds (f-aCFO) design for delayed toxicity in phase I clinical trials. *Statistics in Medicine*. In revision.
26. Gu, J. and **Yin, G.** (2024). Unit Information Dirichlet Process. *Biometrics*. In revision.
27. Zhang, C., Gu, J., Wu, Y. and **Yin, G.** (2024). Area Above Quantile: An Alternative Measure to Restricted Mean Survival Time. *Annals of Applied Statistics*. In revision.
28. Liu, C., Su, W., Liu, K.Y., **Yin, G.** and Zhao, X. (2024). Efficient Estimation for Functional Accelerated Failure Time Model. *Biometrika*. In revision.
29. Cheung, Y. Y. H., Lau, E. H. Y., **Yin, G.**, Lin, Y., Jiang, J., Cowling, B. J., and Lam, K. F. (2023). Joint analysis of vaccination effectiveness and antiviral drug effectiveness for COVID-19: A causal inference approach. *International Journal of Infectious Diseases*. In press.
30. Cheung, Y. Y. H., Lau, E. H. Y., **Yin, G.**, Lin, Y., Cowling, B. J., and Lam, K. F. (2023). Effectiveness of vaccines and antivirals in preventing severe and fatal COVID-19 in Hong Kong. *Emerging Infectious Diseases*. In press.
31. Fan, Y. and **Yin, G.** (2023). Gaussian mixture models with concave penalized fusion. *Statistica Sinica*. In press.
32. Su, W., Liu, L., **Yin, G.**, Zhang, Y. and Zhao, X. (2023). Semiparametric reversed mean model for recurrent event process with informative terminal event. *Statistica Sinica*. In press.
33. Zhang, H., **Yin, G.*** and Rubin, D. (2023). PCA rerandomization. *Canadian Journal of Statistics*. In press.
34. Tan, R., Zang, Y. and **Yin, G.** (2022). Nonlinear dimension reduction for functional data with application to clustering. *Statistica Sinica*. In press.
35. Wang, W., Jin, H., Y. D. Zhang, and **Yin, G.*** (2023). Two-dimensional calibration-free odds (2dCFO) design for phase I drug-combination trials. *Frontiers in Oncology, Section Cancer Epidemiology and Prevention*. **13** <https://doi.org/10.3389/fonc.2023.1294258>
36. Zhang, H., Su, W. and **Yin, G.** (2023). Quasi-rerandomization for observational studies. *BMC Medical Research Methodology* **23**, Article Number: 155. <https://doi.org/10.1186/s12874-023-01977-7>
37. Cheung, Y. Y. H., Lam, K. F., Zhang, H., Kwan, C. W., Wat, K. P., Zhang, Z., Zhu K., Chung, Y. K. and **Yin, G.*** (2023). A randomized controlled experiment for comparing face-to-face and online teaching during COVID-19 pandemic. *Frontiers in Education, Section Higher Education*, Volume **8** <https://doi.org/10.3389/educ.2023.1160430>

38. Gu, J. and **Yin, G.** (2023). Contribution to the Discussion of ‘Martingale Posterior Distributions’ by Fong, Holmes and Walker. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, <https://doi.org/10.1093/jrsssb/qkad092>
39. Jin, H. and **Yin, G.*** (2023). Time-to-event calibration-free odds: A robust efficient design for phase I trials with late-onset outcomes. *Pharmaceutical Statistics* **22**, 773–783.
40. Wang, X., Wang H., **Yin, G.** and Zhang, Y. (2023). Network-based drug repurposing for the treatment of COVID-19 patients in different clinical stages. *Heliyon*. 9(3):e14059. doi: 10.1016/j.heliyon.2023.e14059.
41. Gu, J., Fan, Y. and **Yin, G.*** (2023). Omnibus test for restricted mean survival time based on influence function. *Statistical Methods in Medical Research*. **32**, 1082–1099.
42. Gu, J., Zhang, Y. and **Yin, G.*** (2023). Bayesian log-rank test. *The American Statistician*. **77** (3) 292–300. <https://doi.org/10.1080/00031305.2022.2161637>
43. Gu, J. and **Yin, G.*** (2023). Bayesian SIR model with change points with application to the Omicron wave in Singapore. *Scientific Reports* **12** (1), 20864.
44. Zhang, H. and **Yin, G.*** (2023). Unit information prior for incorporating the real-world evidence into randomized controlled trials. *Statistical Methods in Medical Research* **32**, 229–241.
45. Fan, Y., Gu, J. and **Yin, G.** (2023). Sparse concordance-based ordinal classification. *Scandinavian Journal of Statistics* **50**, 934–961.
46. Wang, X., Huang, J., **Yin, G.**, Huang, J. and Wu Y. (2023). Double bias correction for high-dimensional sparse additive hazards regression with covariate measurement errors. *Lifetime Data Analysis* **29**, 115–141.
47. Su, W., **Yin, G.**, Zhang, J. and Zhao, X. (2023). A divide-and-conquer approach for the accelerated failure time model with massive time-to-event data. *Canadian Journal of Statistics*. **51**, 400–419.
48. He, B., Liu, Y., Wu, Y. and **Yin, G.** (2023). Model aggregation for doubly divided data with large size and large dimension. *Computational Statistics* **38**, 509–529.
49. Gu, J. and **Yin, G.** (2023). Triangular concordance learning of social networks. *Journal of Computational and Graphical Statistics* **32**, 434–447.
50. Lin, R., **Yin, G.**, and Shi, H. (2023). Bayesian adaptive model selection design for identifying optimal biological dose in phase I/II clinical trials. *Biostatistics* **24**, 277–294.
51. Zhang, C. and **Yin, G.** (2022). Safety of paclitaxel-coated devices in the femoropopliteal arteries: A systematic review and meta-analysis. *Plos One* 17(10):e0275888. doi:10.1371/journal.pone.0275888.
52. Gan, D., **Yin, G.** and Zhang, Y. (2022). The graphical R2D2 estimator for the precision matrices. *Briefings in Bioinformatics* 23(6):bbac426. doi:10.1093/bib/bbac426.
53. Zhang, C. and **Yin, G.*** (2022). Alternative measures to hazard ratio: Restricted mean survival time vs area above quantile. JAMA Oncology, comments. <https://jamanetwork.com/journals/jamaoncology/fullarticle/2793567>

54. Tian, P., Chan, T.H., Wang, Y.F., Yang, W., **Yin, G.** and Zhang, Y. (2022). Multiethnic polygenic risk prediction in diverse populations through transfer learning. *Frontiers in Genetics – Statistical Genetics and Methodology*. 13:906965. doi:10.3389/fgene.2022.906965.
55. Jin, H., Du, W., and **Yin, G.*** (2022). Approximate Bayesian computation design for phase I clinical trials. *Statistical Methods in Medical Research* **31**, 2310–2322.
56. Jin, H. and **Yin, G.*** (2022). CFO: Calibration-free odds design for phase I/II clinical trials. *Statistical Methods in Medical Research* **31**, 1051–1066.
57. Yu, C., Liu, J., Nemati, S. and **Yin, G.** (2022). Reinforcement learning in healthcare: A survey. *AMC Computing Surveys* **55**, 1–36. <https://doi.org/10.1145/3477600>
58. Lin, R., Shi, H., **Yin, G.**, Thall, P. F., Yuan Y. and Flowers R. C. (2022). Bayesian hierarchical random-effects meta-analysis and design of phase I clinical trials. *Annals of Applied Statistics* **16**, 2481–2504.
59. Zhang, C. and **Yin, G.*** (2022). Bayesian nonparametric analysis of restricted mean survival time. *Biometrics* **79**, 1383–1396.
60. Liu, L., Su, W., **Yin, G.***, Zhao, X. and Zhang, Y. (2022). Nonparametric inference for reversed mean models with panel count data. *Bernoulli*. **28**, 2968–2997.
61. Su, W., He, B., Zhang, Y. and **Yin, G.*** (2022). C-Index regression for recurrent event data. *Contemporary Clinical Trials* **118**, 106787.
62. Jin, H. and **Yin, G.*** (2022). Unit information prior for adaptive information borrowing from multiple historical datasets. *Statistics in Medicine* **40**, 5657–5672.
63. Gu, J., Fan, Y. and **Yin, G.** (2022). Reconstruction of Kaplan–Meier estimator as M-estimator. *The American Statistician*. **76**, 37–43.
64. Zhang, C. and **Yin, G.*** (2021). Prevention of Covid-19 with the BNT162b2 and mRNA-1273 Vaccines. *New England Journal of Medicine*, **385**, 1818–1819.
65. Zhang, C., Jin, H., Wen, Y. and **Yin, G.*** (2021). Efficacy of COVID-19 treatments: A Bayesian network meta-analysis of randomized controlled trials. *Frontiers in Public Health, Section Infectious Diseases Surveillance, Prevention and Treatment*. doi:10.3389/fpubh.2021.729559.
66. J. J. Lee and **Yin, G.** (2021). Principles of Bayesian trials and reporting. *Journal of Thoracic Oncology* **16**, 30–36.
67. Zhang, H. and **Yin, G.** (2021). Response-adaptive rerandomization. *Journal of Royal Statistical Society C – Applied Statistics* **70**, 1281–1298.
68. Jin, H., **Yin, G.**, Yuan, B. and Jiang, F. (2021). Bayesian hierarchical model for change point detection in multivariate sequences. *Technometrics* **64**, 177–186.
69. **Yin, G.** and Shi, H. (2021). Demystify Lindley’s paradox by connecting p-value and posterior probability. *Statistics and Its Interface* **14**, 489–502.
70. Fan, Y. and **Yin, G.*** (2021). Concordance index: Surrogacy of progression-free survival for overall survival. *Contemporary Clinical Trials* **104**, 106353.

71. Zhang, T. Yang, Z. and **Yin, G.*** (2021). Dynamic ordering design for dose finding in drug-combination trials. *Pharmaceutical Statistics* **20**, 348–361.
72. Zhang, C. and **Yin, G.*** (2021). Reduction in number to treat versus number needed to treat. *BMC Medical Research Methodology* **21** Article Number 48.
73. **Yin, G.**, Yang, Z., Odani, M. and Fukimbara, S. (2021). Bayesian hierarchical modeling and biomarker cutoff identification in basket trials. *Statistics in Biopharmaceutical Research* **13**, 248–258.
74. Lin, R. Yang, Z., Yuan, Y. and **Yin, G.*** (2021). Sample size re-estimation in adaptive enrichment design. *Contemporary Clinical Trials* 100:106216.
75. Shi, H. and **Yin, G.*** (2021). Reconnecting p-value and posterior probability under one- and two-sided tests. *The American Statistician* **75**, 265–275. <https://doi.org/10.1080/00031305.2020.1717621>
76. Yan, X., **Yin, G.** and Zhao, X. (2021). Subgroup analysis in censored linear regression. *Statistica Sinica* **31**, 1027–1054.
77. He, B., Liu, Y., Wu, Y., **Yin, G.**, and Zhao, X. (2020). Functional martingale residual process for high-dimensional Cox regression with model averaging. *Journal of Machine Learning Research*, **21** (207): 1–37.
78. **Yin, G.*** and Yang, Z. (2020). Fractional design: An alternative paradigm for late-onset toxicities in oncology dose-finding studies. *Contemporary Clinical Trials Communications* **19**, 100650.
79. **Yin, G.**, Zhang, C. and Jin, H. (2020). Statistical issues and lessons learned from COVID-19 clinical trials with lopinavir-ritonavir and remdesivir. *Journal of Medical Internet Research JMIR Public Health and Surveillance* 2020;6(3):e19538 doi:10.2196/19538.
80. **Yin, G.** and Jin, H. (2020). Comparison of transmissibility of coronavirus between symptomatic and asymptomatic patients: Reanalysis of the Ningbo COVID-19 data. *Journal of Medical Internet Research JMIR Public Health and Surveillance* 2020;6(2):e19464 doi:10.2196/19464.
81. Jin, H. and **Yin, G.*** (2020). Bayesian enhancement two-stage design with error control for phase II clinical trials. *Statistics in Medicine* **39**, 4452–4465.
82. Zhang, C., Wu, Y. and **Yin, G.*** (2020). Restricted mean survival time for interval-censored data. *Statistics in Medicine* **39**, 3879–3895.
83. Zhang, A., Zhang, H. and **Yin, G.** (2020). Adaptive iterative Hessian sketch via A-optimal subsampling. *Statistics and Computing* **30**, 1075–1090.
84. Liu, Y. and **Yin, G.** (2020). The Delaunay triangulation learner and its ensemble. *Computational Statistics & Data Analysis* **152**, 107030.
85. Lam, C. K., Jin, H., Jiang, F. and **Yin, G.*** (2020). A variable selection approach to multiple change-points detection. *Statistics and Its Interface* **13**, 251–260.
86. Zheng, S, Lin, R., Guo, J. and **Yin, G.** (2020). Testing homogeneity of high-dimensional covariance matrices. *Statistica Sinica* **30**, 35–53.

87. Shi, H., Zhang, T. and **Yin, G.*** (2020). START: Single-to-double arm transition design for phase II clinical trials. *Pharmaceutical Statistics* **19**, 454–467.
88. Wang, N. and **Yin, G.** (2020) Convergence rates of the blocked Gibbs sampler with random scan in the Wasserstein metric. *Stochastics* **92**, 265–274.
89. Jiang, F., Cheng, Q., **Yin, G.** and Shen, H. (2020). Functional censored quantile regression. *Journal of the American Statistical Association* **115**, 931–944.
90. **Yin, G.*** and Zhang C. (2019). Reanalysis of data comparing prophylactic cranial irradiation vs observation in patients with locally advanced non-small cell lung cancer. *Journal of American Medical Association Oncology* JAMA Oncology 2019;5(11):1638. doi:10.1001/jamaoncol.2019.3593
91. **Yin, G.*** and McCaw, Z. R. (2019). Design of non-inferiority trials for hypofractionated vs conventional radiotherapy among cancer patients. *Journal of American Medical Association Oncology* JAMA Oncology **5(10)**, 1508–1509. doi:10.1001/jamaoncol.2019.2391
92. McCaw, Z. R., **Yin, G.** and Wei, L. J. (2019). Using the restricted mean survival time difference as an alternative to the hazard ratio for analyzing clinical cardiovascular studies. *Circulation* **140**, 1366–1368.
93. Yang, Z. and **Yin, G.*** (2019). The number needed to treat for both binary and survival endpoints. *British Medical Journal*. <https://www.bmj.com/content/364/bmj.l525/rr-2>
94. Yang, Z., McCaw, Z. R. and **Yin, G.*** (2019). Caplacizumab for acquired thrombotic thrombocytopenic purpura. *New England Journal of Medicine*. 380(18): e32. doi: 10.1056/NEJMc1902336
95. Yang, Z., McCaw, Z. R. and **Yin, G.*** (2019). Interpreting the long-term benefit of radical prostatectomy in prostate cancer. *New England Journal of Medicine* **380**, 1083–1084.
96. Jiang, F., **Yin, G.** and Dominici, F. (2019). Bayesian model selection approach to multiple change-points detection with non-local prior distributions. *ACM Transactions on Knowledge Discovery from Data*. Article 48, 13: 1–17. <https://dl.acm.org/doi/pdf/10.1145/3340804>
97. Wang, Y., Zou, C., Wang, Z. and **Yin, G.** (2019). Multiple change-points detection in high dimension. *Random Matrices: Theory and Applications* **8** (4), 1950014.
98. **Yin, G.**, C. Zhang, and Yang, Z. (2019). Biostatistics pitfalls: Lessons learned from analysis of medical data. *Contemporary Clinical Trials*. **87**, 105875. doi: 10.1016/j.cct.2019.105875
99. Yang, Z. and **Yin, G.*** (2019). An alternative approach for estimating the number needed to treat for survival endpoints. PLoS ONE 14(10): e0223301. <https://doi.org/10.1371/journal.pone.0223301>
100. Lam, C. K., Lin, R. and **Yin, G.*** (2019). Non-parametric overdose control for dose finding in drug combination trials. *Journal of Royal Statistical Society C – Applied Statistics* **68**, 1111–1130.
101. Shi, H. and **Yin, G.*** (2019). Control of type I error rates in Bayesian sequential designs. *Bayesian Analysis* **14**, 399–425.
102. Lin, R. and **Yin, G.*** (2018). Uniformly most powerful Bayesian interval design for dose finding. *Pharmaceutical Statistics* **17**, 710–724.

103. Lam, C. K., Xu, Y., and **Yin, G.** (2018). Dynamic portfolio choice without cash. *Quantitative Finance* **19**, 313–326.
104. Liu, Y. and **Yin, G.** (2018). Average holding price. *Annals of Financial Economics* **13**, No. 01. 1850002 <https://doi.org/10.1142/S2010495218500021>
105. Shi, H. and **Yin, G.*** (2018). Bayesian enhancement two-stage design for single-arm phase II clinical trials with binary and time-to-event endpoints. *Biometrics* **74**, 1055–1064.
106. Jiang, F., Ma, Y., and **Yin, G.** (2018). Kernel-based adaptive randomization toward balance in continuous and discrete covariates. *Statistica Sinica* **28**, 2841–2856.
107. Li, H., Cao, Z., and **Yin, G.*** (2018). Varying-association copula models for multivariate survival data. *Canadian Journal of Statistics* **46**, 556–576.
108. Dong, F., and **Yin, G.** (2018). Maximum likelihood estimation for incomplete multinomial data via the weaver algorithm. *Statistics and Computing* **28**, 1095–1117.
109. **Yin, G.**, Chen, N., and Lee, J. J. (2018). Bayesian adaptive randomization and trial monitoring with predictive probability for time-to-event endpoint. *Statistics in Bioscience* **10**, pp 420–438.
110. Wang, G., Zou, C., and **Yin, G.** (2018). Change-point detection in multinomial data with a large number of categories. *Annals of Statistics* **46**, 2020–2044.
111. Shi, H. and **Yin, G.*** (2018). Boosting conditional logit model. *Journal of Choice Modeling* **26**, 48–63.
112. Zhang, J., **Yin, G.**, Liu, Y., and Wu, Y. (2018). Censored cumulative residual independent screening for ultrahigh-dimensional survival data. *Lifetime Data Analysis* **24**, 273–292.
113. Shi, H. and **Yin, G.*** (2018). Two-stage seamless transition design from open-label single-arm to randomized double-arm clinical trials. *Statistical Methods in Medical Research* **27**, 158–171.
114. Lin, R. and **Yin, G.*** (2017). STEIN: A simple toxicity and efficacy interval design for seamless phase I/II clinical trials. *Statistics in Medicine* **36**, 4106–4120.
115. Duan, X., and **Yin, G.*** (2017). Ensemble approaches to estimating the population mean with missing response. *Scandinavian Journal of Statistics* **44**, 899–917.
116. **Yin, G.***, Lam, C. K., and Shi, H. (2017). Bayesian randomized clinical trials: from fixed to adaptive design. *Contemporary Clinical Trials* **59**, 77–86.
117. Shi, H. and **Yin, G.*** (2017). Landmark cure rate models with time-dependent covariates. *Statistical Methods in Medical Research* **26**, 2042–2054.
118. Liu, Y. and **Yin, G.*** (2017). Partitioned log-rank tests for the overall homogeneity of hazard rate functions. *Lifetime Data Analysis* **23**, 400–425.
119. Lin, R. and **Yin, G.*** (2017). Bayesian optimal interval design for dose finding in drug-combination trials. *Statistical Methods in Medical Research* **26**, 2155–2167.
120. Zhao, X., Wu, Y., and **Yin, G.** (2017). Sieve maximum likelihood estimation for a general class of accelerated hazards models with bundled parameters. *Bernoulli* **23**, 3385–3411.

121. Lin, R. and **Yin, G.*** (2017). Nonparametric overdose control with late-onset toxicity in phase I clinical trials. *Biostatistics* **18**, 180–194.
122. Wu, Y. and **Yin, G.*** (2017). Cure rate quantile regression accommodating both finite and infinite survival times. *Canadian Journal of Statistics* **45**, 29–43.
123. Wu, Y. and **Yin, G.*** (2017). Multiple imputation for cure rate quantile regression with censored data. *Biometrics* **73**, 94–103.
124. Shi, H. and **Yin, G.*** (2017). Bayesian two-stage design for phase II clinical trials with switching hypothesis tests. *Bayesian Analysis* **12**, 31–51.
125. Lin, R. and **Yin, G.*** (2016). Bayesian optimal interval design for drug-combination trials. *Frontiers of Biostatistical Methods and Applications in Clinical Oncology*. Springer.
126. Lin, R. and **Yin, G.*** (2016). Robust optimal interval design for high-dimensional dose finding in multi-agent combination trials. ICSA book proceedings for the symposium at Calgary. Springer.
127. Lin, R., Liu, Z., Zheng, S., and **Yin, G.** (2016). Power computation for hypothesis testing with high-dimensional covariance matrices. *Computational Statistics & Data Analysis* **104**, 10–23.
128. Lin, R. and **Yin, G.*** (2016). Bootstrap aggregating continual reassessment method for dose finding in drug-combination trials. *Annals of Applied Statistics* **10**, 2349–2376.
129. Li, H., Duan, X., and **Yin, G.*** (2016). Generalized method of moments for additive hazards model with clustered dental survival data. *Scandinavian Journal of Statistics* **43**, 1124–1139.
130. Xu, J., **Yin, G.***, Ohlssen, D., and Bretz, F. (2016). Bayesian two-stage dose finding for cytostatic agents via model adaptation. *Journal of Royal Statistical Society C – Applied Statistics* **65**, 465–482.
131. Morita, S., Sakamaki, K., and **Yin, G.** (2015). Detecting overall survival benefit derived from survival postprogression rather than progression-free survival. *Journal of the National Cancer Institute* **107** (8), djv133.
132. Lin, R. and **Yin, G.*** (2015). Bayes factor and posterior probability: complementary statistical evidence to p-value. *Contemporary Clinical Trials* **44**, 33–35.
133. Ro, K., Zou, C., Wang, Z., and **Yin, G.** (2015). Outlier detection for high dimensional data. *Biometrika* **102**, 589–599.
134. Jin, I. H., Huo, L., Yuan, Y., and **Yin, G.** (2015). Phase I trial design for drug combinations with Bayesian model averaging. *Pharmaceutical Statistics* **14**, 108–119.
135. **Yin, G.** and Lin, R. (2015). Continual reassessment methods. *Modern Approaches to Clinical Trials Using SAS: Classical, Adaptive, and Bayesian Methods* Chapter 5, edited by R. Zink and S. Menon, pp. 133–162. SAS Institute Inc., Cary, NC, USA.
136. Lin, R. and **Yin, G.*** (2015). Overview of adaptive randomization. *Modern Approaches to Clinical Trials Using SAS: Classical, Adaptive, and Bayesian Methods* Chapter 9, edited by R. Zink and S. Menon, pp. 247–272. SAS Institute Inc., Cary, NC, USA.

137. Lin, R. and **Yin, G.*** (2015). Sample size re-estimation in adaptively randomized clinical trials with missing data. *Modern Adaptive Randomized Clinical Trials: Statistical and Practical Aspects*, edited by O. Sverdlov, pp. 269–286. Chapman & Hall/CRC Press.
138. Wu, Y., Ma, Y., and **Yin, G.*** (2015). Smoothed and corrected score approach to censored quantile regression with measurement errors. *Journal of the American Statistical Association* **110**, 1670–1683.
139. **Yin, G.** and Lin, R. (2015). Comments on ‘Competing designs for drug combination in phase I dose-finding clinical trials’ by M-K. Riviere, F. Dubois and S. Zohar *Letter to Editor – Statistics in Medicine* **34**, 13–17.
140. Wu, Y. and **Yin, G.*** (2015). Conditional quantile screening in ultrahigh-dimensional heterogeneous data. *Biometrika* **102**, 65–76.
141. Wang, X., Zhang, J., Yu, L., and **Yin, G.*** (2015). Generalized partially linear single-index model for zero-inflated count data. *Statistics in Medicine* **34**, 876–886.
142. Zou, C., **Yin, G.***, Feng, L., and Wang, Z. (2014). Nonparametric maximum likelihood approach to multiple change-point problems. *Annals of Statistics* **42**, 970–1002.
143. Xu, J. and **Yin, G.*** (2014). Two-stage adaptive randomization for delayed response in clinical trials. *Journal of Royal Statistical Society C – Applied Statistics* **63**, 559–578.
144. **Yin, G.**, Zeng, D., and Li, H. (2014). Censored quantile regression with varying coefficients. *Statistica Sinica* **24**, 855–870.
145. Wu, Y. and **Yin, G.*** (2013). Cure rate quantile regression for censored data with a survival fraction. *Journal of the American Statistical Association* **108**, 1517–1531.
146. Liu, S., **Yin, G.**, and Yuan, Y. (2013). Bayesian data augmentation dose finding with continual reassessment method and delayed toxicity. *Annals of Applied Statistics* **7**, 2138–2156.
147. **Yin, G.** and Ma, Y. (2013). Pearson-type goodness-of-fit test with bootstrap maximum likelihood estimation. *Electronic Journal of Statistics* **7**, 412–427.
148. Shi, Y. and **Yin, G.*** (2013). Escalation with overdose control for phase I drug-combination trials. *Statistics in Medicine* **32**, 4400–4412.
149. Ma, Y. and **Yin, G.*** (2013). Testing overall and subpopulation treatment effects with measurement errors. *Statistica Sinica* **23**, 1019–1042.
150. **Yin, G.**, Zheng, S., and Xu, J. (2013). Two-stage dose finding for cytostatic agents in phase I clinical trials. *Statistics in Medicine* **32**, 644–660.
151. **Yin, G.**, Zheng, S., and Xu, J. (2013). Fractional dose-finding methods with late-onset toxicity in phase I clinical trials. *Journal of Biopharmaceutical Statistics* **23**, 856–870.
152. Gu, X., **Yin, G.**, and Lee, J. J. (2013). Bayesian two-stage Lasso strategies for biomarker selection in personalized medicine development. *Contemporary Clinical Trials* **36**, 642–650.
153. Lee, J. J., Chen, N., and **Yin, G.** (2012). Worth adapting? Revisiting the usefulness of outcome-adaptive randomization. *Clinical Cancer Research* **18**, 4498–4507.

154. Huo, L., Yuan, Y., and **Yin, G.*** (2012). Bayesian dose finding for combined drugs with discrete and continuous doses. *Bayesian Analysis* **7**, 235–252.
155. Diao, G. and **Yin, G.** (2012). A general transformation class of semiparametric cure rate frailty models. *Annals of the Institute of Statistical Mathematics* **64**, 959–989.
156. **Yin, G.**, Chen, N., and Lee, J. J. (2012). Phase II trial design with Bayesian adaptive randomization and predictive probability. *Journal of Royal Statistical Society C – Applied Statistics* **61**, 219–235.
157. Yuan, Y. and **Yin, G.** (2011). On the usefulness of outcome-adaptive randomization. *Journal of Clinical Oncology* **29**, e390–e392.
158. Garcia, T., Ma, Y., and **Yin, G.** (2011). Efficiency improvement in a class of survival models through model-free covariate incorporation. *Lifetime Data Analysis* **7**, 552–565.
159. Yuan, Y. and **Yin, G.** (2011). Dose-response curve estimation: A semiparametric mixture approach. *Biometrics* **67**, 1543–1554.
160. Yuan, Y. and **Yin, G.** (2011). Robust EM continual reassessment method in oncology dose finding. *Journal of the American Statistical Association* **106**, 818–831.
161. **Yin, G.**, Ma, Y., Liang, F., and Yuan, Y. (2011). Stochastic generalized method of moments. *Journal of Computational and Graphical Statistics* **20**, 714–727.
162. Yuan, Y. and **Yin, G.** (2011). Bayesian hybrid dose-finding design in phase I oncology clinical trials. *Statistics in Medicine* **30**, 2098–2108.
163. Yuan, Y. and **Yin, G.*** (2011). Bayesian phase I/II adaptively randomized oncology trials with combined drugs. *Annals of Applied Statistics* **5**, 924–942.
164. Ma, Y. and **Yin, G.** (2011). Censored quantile regression with covariate measurement errors. *Statistica Sinica* **21**, 949–971.
165. Lei, X., Yuan, Y., and **Yin, G.*** (2011). Bayesian phase II clinical trial design with time-to-event adaptive randomization. *Lifetime Data Analysis* **17**, 156–174.
166. **Yin, G.** and Yuan, Y. (2010). Correspondence to the discussion of “Bayesian dose finding in oncology for drug combinations by copula regression.” *Journal of Royal Statistical Society C – Applied Statistics* **59**, 544–546.
167. Ma, Y. and **Yin, G.*** (2010). Semiparametric median residual life model and inference. *Canadian Journal of Statistics* **38**, 665–679.
168. Yuan, Y. and **Yin, G.** (2010). Bayesian quantile regression for longitudinal studies with non-ignorable missing data. *Biometrics* **66**, 105–114.
169. **Yin, G.** and Yuan, Y. (2010). Bayesian approach for adaptive design. *Handbook of Adaptive Designs in Pharmaceutical and Clinical Development*, A. Pong, and S.-C. Chow, (eds.), Chapter 3. pp. (3-1)–(3-19). Boca Raton, FL: Chapman & Hall/CRC.
170. **Yin, G.** and Nieto-Barajas, L. E. (2009). Bayesian cure rate model accommodating multiplicative and additive covariates. *Statistics and Its Interface* **2**, 513–521.

171. **Yin, G.** and Li, H. (2009). Least squares estimation of varying-coefficient hazard regression with application to breast cancer dose-intensity data. *Canadian Journal of Statistics* **37**, 659–674.
172. **Yin, G.** and Yuan, Y. (2009). Bayesian model averaging continual reassessment method in phase I clinical trials. *Journal of the American Statistical Association* **104**, 954–968.
173. **Yin, G.** (2009). Bayesian generalized method of moments (with discussion). *Bayesian Analysis* **4**, 191–208; and Rejoinder, 217–222.
174. Yuan, Y. and **Yin, G.** (2009). Bayesian dose finding by jointly modeling toxicity and efficacy as time-to-event outcomes. *Journal of Royal Statistical Society C – Applied Statistics* **58**, 719–736.
175. Li, H. and **Yin, G.*** (2009). Generalized method of moments for linear regression with clustered failure time data. *Biometrika* **96**, 293–306.
176. **Yin, G.** and Yuan, Y. (2009). Bayesian dose finding in oncology for drug combinations by copula regression. *Journal of Royal Statistical Society C – Applied Statistics* **58**, 211–224.
177. **Yin, G.** (2009). Bayesian chi-squared goodness-of-fit test for censored data. *Journal of Statistical Planning and Inference* **139**, 1474–1483.
178. **Yin, G.** and Yuan, Y. (2009). A latent contingency table approach to dose finding for combinations of two agents. *Biometrics* **65**, 866–875.
179. **Yin, G.** (2008). Bayesian transformation cure frailty models with multivariate failure time data. *Statistics in Medicine* **27**, 5929–5940.
180. Yuan, Y. and **Yin, G.** (2008). Sequential continual reassessment method for two-dimensional dose finding. *Statistics in Medicine* **27**, 5664–5678.
181. **Yin, G.**, Li, H., and Zeng, D. (2008). Partially linear additive hazards regression with varying coefficients. *Journal of the American Statistical Association* **103**, 1200–1213.
182. **Yin, G.**, Zeng, D., and Li, H. (2008). Power-transformed linear quantile regression with censored data. *Journal of the American Statistical Association* **103**, 1214–1224.
183. Ma, Y. and **Yin, G.** (2008). Cure rate model with mismeasured covariates under transformation. *Journal of the American Statistical Association* **103**, 743–756.
184. Nieto-Barajas, L. E. and **Yin, G.** (2008). Bayesian semiparametric cure rate model with an unknown threshold. *Scandinavian Journal of Statistics* **35**, 540–556.
185. Li, H., **Yin, G.**, and Zhou, Y. (2007). Local likelihood with time-varying coefficient additive hazards model. *Canadian Journal of Statistics* **35**, 321–337.
186. Ji, Y., **Yin, G.**, Tsui, K.-W., Kolonin, M. G., Sun, J., Arap, W., Pasqualini, R., and Do, K.-A. (2007). Bayesian mixture models for complex high-dimensional count data in phage display experiments. *Journal of Royal Statistical Society C – Applied Statistics* **56**, 1–14.
187. Cong, X., **Yin, G.**, and Shen, Y. (2007). Marginal analysis of correlated failure time data with informative cluster sizes. *Biometrics* **63**, 663–672.

188. **Yin, G.** (2007). Model checking for additive hazards model with multivariate survival data. *Journal of Multivariate Analysis* **98**, 1018–1032.
189. Ji, Y., Li, Y., and **Yin, G.** (2007). Bayesian dose finding in phase I clinical trials based on a new statistical framework. *Statistica Sinica* **17**, 531–547.
190. **Yin, G.** and Ibrahim, J. (2006). Bayesian transformation hazard models. *The Second Lehmann Symposium—Optimality, IMS Lecture Notes—Monographs Series*, 170–182.
191. **Yin, G.**, Li, Y., and Ji, Y. (2006). Bayesian dose-finding in phase I/II trials using toxicity and efficacy odds ratio. *Biometrics* **62**, 777–784.
192. Zeng, D., **Yin, G.**, and Ibrahim, J. (2006). Semiparametric transformation models for survival data with a cure fraction. *Journal of the American Statistical Association* **101**, 670–684.
193. **Yin, G.** and Zeng, D. (2006). Efficient algorithm for computing maximum likelihood estimates in linear transformation models. *Journal of Computational and Graphical Statistics* **15**, 228–245.
194. **Yin, G.** and Ibrahim, J. (2005). Cure rate models: a unified approach. *Canadian Journal of Statistics* **33**, 559–570.
195. Zeng, D., **Yin, G.**, and Ibrahim, J. (2005). Inference for a class of transformed hazard models. *Journal of the American Statistical Association* **100**, 1000–1008.
196. **Yin, G.** and Ibrahim, J. (2005). Bayesian frailty models based on Box-Cox transformed hazards. *Statistica Sinica* **15**, 781–794.
197. **Yin, G.** and Shen, Y. (2005). Self-designing trial combining with classical group sequential monitoring. *Journal of Biopharmaceutical Statistics* **15**, 667–675.
198. **Yin, G.** (2005). Bayesian cure rate frailty models with application to a root canal therapy study. *Biometrics* **61**, 552–558.
199. **Yin, G.** and Ibrahim, J. (2005). A general class of Bayesian survival models with zero and non-zero cure fractions. *Biometrics* **61**, 403–412.
200. **Yin, G.** and Shen, Y. (2005). Adaptive design and estimation in randomized clinical trials with correlated observations. *Biometrics* **61**, 362–369.
201. Zeng, D., Lin, D. Y., and **Yin, G.** (2005). Maximum likelihood estimation in proportional odds model with random effects. *Journal of the American Statistical Association* **100**, 470–483.
202. **Yin, G.** and Ibrahim, J. (2005). A class of Bayesian shared gamma frailty models with multivariate failure time data. *Biometrics* **61**, 209–217.
203. **Yin, G.** and Cai, J. (2005). Quantile regression models with multivariate failure time data. *Biometrics* **61**, 152–162.
204. **Yin, G.** and Zeng, D. (2005). Pair chart test for an early survival difference. *Lifetime Data Analysis* **11**, 117–129.
205. **Yin, G.** and Hu, J. (2004). Two simulation methods for constructing confidence bands under the additive risk model. *Journal of Biopharmaceutical Statistics* **14**, 389–402.

206. Hu, J., **Yin, G.**, Morris, J. S., Zhang, L., and Wright, A. F. (2004). Entropy and survival-based weights to combine Affymetrix array types and analyze differential expression and survival. *Methods of Microarray Data Analysis IV, Critical Assessment of Microarray Data Analysis*, eds. J. S. Shoemaker and S. M. Lin, pp. 95–108.
207. Morris, J. S., **Yin, G.**, Baggerly, K., Wu, C., and Zhang, L. (2004). Pooling information across different studies and oligonucleotide chip types to identify prognostic genes for lung cancer. *Methods of Microarray Data Analysis IV, Critical Assessment of Microarray Data Analysis*, eds. J. S. Shoemaker and S. M. Lin, pp. 51–66.
208. **Yin, G.** and Cai, J. (2004). Additive hazards model for multivariate failure time data. *Biometrika* **91**, 801–818.
209. Hu, J. and **Yin, G.** (2003). A semiparametric regression model for oligonucleotide arrays. *Journal of Modern Applied Statistical Methods* **2**, 256–267.
210. **Yin, G.**, Cai, J., and Kim, J. (2003). Quantile inference with multivariate failure time data. *Biometrical Journal* **45**, 602–617.

Collaborative Research

211. Wen, Y. F., Chen, M. X., **Yin, G.**, Lin, R., Yang, S., Zhong, Y. J., Dong, Q. Q. and Wong, H. M. (2020). The global, regional, and national burden of cancer among adolescents and young adults in 204 countries and territories, 1990-2019: A population-based study. *Journal of Hematology & Oncology* **14** (1), 1–14.
212. Zhong, Y. J., Wen, Y. F., Wong, H. M., **Yin, G.**, Lin, R. and Yang, S. Y. (2019). Trends and patterns of disparities in burden of lung cancer in the United States, 1974-2015. *Frontiers in Oncology, Thoracic Oncology* **9**, 404.
213. Cardó-Vila, M., Marchió, S., Sato, M., Staquicini, F., Bronk, J., **Yin, G.**, Zurita, A., Lee, J. J., Hong, W., Wistuba, I., Arap, W., and Pasqualini, R. (2016). Interleukin-11 receptor is a candidate target for ligand-directed therapy in lung cancer: analysis of clinical samples and BMTP-11 pre-clinical activity. *American Journal of Pathology* **186**, 2162–2170.
214. Wen, Y. F., Wong, H. M., Lin, R., **Yin, G.**, and C. P. McGrath (2015). Inter-ethnic/racial facial variations: A systematic review and Bayesian meta-analysis of photogrammetric studies. *Plos One*, DOI:10.1371/journal.pone.0134525
215. Arun, B. K., Dhingra, K., Valero, V., Kau, S.-W., Broglio, K., Booser, D., Guerra, L., **Yin, G.**, Walters, R., Sahin, A., Ibrahim, N., Buzdar, A. U., Frye, D., Sneige, N., Strom, E., Ross, M., Theriault, R., Vadhan-Raj, S., Hortobagyi, G. N. (2011). Randomized trial of dose intensive neoadjuvant chemotherapy with or without G-CSF in locally advanced breast cancer: long-term results. *Oncologist* **16**, 1527–1534.
216. Richards, K. L., Zhang, B., Sun, M., Dong, W., Churchill, J., Bachinski, L. L., Wilson, C. D., Baggerly, K. A., **Yin, G.**, Hayes, D. N., Wistuba, I. I. and Krahe, R. (2011). Methylation of the candidate biomarker TCF21 is very frequent across a spectrum of early stage non-small cell lung cancers. *Cancer* **117**, 606–617.

217. Yuan, P., Kadara, H., Behrens, C., Tang, X., Woods, D., Solis, L. M., Huang, J., Spinola, M., Dong, W., **Yin, G.**, Fujimoto, J., Kim, E., Xie, Y., Girard, L., Moran, C., Hong, W. K., Minna, J. D. and Wistuba, I. I. (2010). Sex determining region Y-box 2 (SOX2) is a potential cell-lineage gene highly expressed in the pathogenesis of squamous cell carcinomas of the lung. *Plos One* **5**, e9112. doi:10.1371/journal.pone.0009112
218. Arun, B. K., Granville, L. A., **Yin, G.**, Middleton, L. P., Dawood, S., Shu, W.-K., Kamal, A., Hsu, L., Hortobagyi, G. N. and Sahin, A. A. (2010). Glutathione-S-transferase-pi (GST-pi) expression in early breast cancer: associated with outcome and response to chemotherapy. *Cancer Investigation* **28**, 554–559.
219. Sun, M., Behrens, C., Feng, L., Ozburn, N., Tang, X., **Yin, G.**, Komaki, R., Varella-Garcia, M., Hong, W. K., Aldape, K. D. and Wistuba, I. I. (2009). HER family receptor abnormalities in lung cancer brain metastases and corresponding primary tumors. *Clinical Cancer Research* **15**, 4829–4837.
220. Rivera, E., Mejia, J., Arun, B., Adinin, R., Walters, R., Abenaa B., A., Broglio, K., **Yin, G.**, Hortobagyi, G. and Valero, V. (2008). Phase III study comparing the use of docetaxel on an every-three-week versus weekly schedule in the treatment of metastatic breast cancer. *Cancer* **112**, 1455–1461.
221. Arun, B., Valero, V., Logan, C., Broglio, K., Rivera, E., Brewster, A., **Yin, G.**, Green, M., Kuerer, H., Gong, Y., Browne, D., Hortobagyi, G. N. and Sneige, N. (2007). Comparison of ductal lavage and random periareolar fine needle aspiration as tissue acquisition methods in early breast cancer prevention trials. *Clinical Cancer Research* **13**, 4943–4948.
222. Gonzalez, R. J., Buzdar, A. U., Symmans, W. F., Yen, T. W., Broglio, K. R., Lucci, A., Esteva, F. J., **Yin, G.** and Kuerer, H. M. (2007). Novel clinical trial designs for treatment of ductal carcinoma in situ of the breast with trastuzumab (herceptin). *Breast Journal* **13**, 72–75.
223. Rivera, E., Meyers, C., Groves, M., Valero, V., Francis, D., Arun, B., Broglio, K., **Yin, G.**, Hortobagyi, G. N. and Buchholz, T. (2006). Phase I study of capecitabine in combination with temozolomide in the treatment of patients with brain metastases from breast carcinoma. *Cancer* **107**, 1348–1354.
224. Sneige, N., Liu, B., **Yin, G.** and Arun, B. K. (2006). Correlation of cytologic findings and chromosomal instability detected by fluorescence in situ hybridization in breast fine-needle aspiration specimens from women at high risk for breast cancer. *Modern Pathology* **19**, 622–629.
225. Klos, K. S., Sun, M., Tan, M., Zhou, X., Li, P., Yang, W., **Yin, G.** and Yu, D. (2006). ErbB2 increases VEGF protein synthesis via activation of the mTOR/p70S6K pathway leading to increased angiogenesis and spontaneous metastasis of human breast cancer cells. *Cancer Research* **66**, 2028–2037.
226. Tan, M., Li, P., Sun, M., **Yin, G.** and Yu, D. (2006). Upregulation and activation of PKC α by ErbB2 through Src promotes breast cancer cell invasion that can be blocked by combined treatment with PKC α and Src inhibitors. *Oncogene* **25**, 3286–3295.
227. Hanrahan, E. O., Broglio, K. R., Buzdar, A. U., Theriault, R. L., Valero, V., Cristofanilli, M., **Yin, G.**, Kau, S.-W., Hortobagyi, G. N. and Rivera, E. (2005). Combined-modality treatment for isolated recurrences of breast carcinoma. *Cancer* **104**, 1158–1171.

228. Caplan, D., Cai, J., **Yin, G.** and White, A. (2005). Root canal filled versus non-root canal filled teeth: A retrospective comparison of survival times. *Journal of Public Health Dentistry* **65**, 90–96.