

TIANXI CAI

Business Address:

Department of Biostatistics, Harvard University
655 Huntington Ave, Boston, MA 02115

Phone: (617) 432-1056

Fax: (617) 432-5619

E-mail: tcai@hsph.harvard.edu

Education

ScD	Biostatistics	Harvard University (1996–1999)	Boston, MA
BSc	Mathematics	University of Science and Technology of China (1991–1995)	P.R. China

Professional Experience

2012–present	Professor , Dept of Biostatistics, Harvard University.
2007–2008	Joy Foundation Fellow , Radcliffe Institute for Advanced Studies.
2006–2011	Associate Professor , Dept of Biostatistics, Harvard University.
2002–2006	Assistant Professor , Dept of Biostatistics, Harvard University.
2000–2002	Assistant Professor , Dept of Biostatistics, Univ. of Washington.
1999–2000	Research Fellow , Dept of Biostatistics, Harvard University.
06–08/1997	Summer Assistant Research Statistician , Schering Plough Research Institute.

Honors and Awards

Myrto Lefkopoulou Distinguished Lecturer (2014)
Fellow of American Statistical Association (2011)
Joy Foundation Fellow, Radcliffe Institute for Advanced Studies (2007–2008)
Distinguished Visiting Junior Scholar, Peter Wall Institute of Advanced Study (2001)
Student Travel Award, ENAR International Biometrics Society (1999)
Authur T. Lyman and Henry S. Grew Scholarship, Harvard University (1998–1999)
Robert Reed Award for Academic Excellence in Biostatistics, Harvard University (1997)
Pfizer Scholarship (1997–1998)
Genetics Institute Scholarship (1996–1997)

Professional Activities

National and International Committee

George W. Snedecor Award Selection Committee (2012–2016)
Program Committee Chair, International Chinese Statistical Association (2012 –2013)
Recombinant DNA Advisory Committee (RAC), NIH (2011 –)
Education Advisory Committee, International Biometric Society ENAR (2010)
Program Committee Member and Biometrics Section Chair, International Chinese Statistical Association (2011–2012)
Executive Committee co-Chair, 2012 ICSA Applied Statistical Symposium
ASA Biometrics Section Program Chair 2011 Joint Statistical Meeting
Organizing Committee, IMS-China International Conference on Stat. & Prob. (2009)
IMS Program Chair, International Biometrics Society ENAR (2008 – 2009)

Editorial Board

Guest Editor

Lifetime Data Analysis (Special Issue on Risk Prediction and Evaluation 2013, Joint with Drs. M. Gail and R. Pfeiffer)

Statistics and Its Interface (Special Issue 2010, Joint with Drs. J. Zhu and H. Zhou)

Associate Editor

Journal of the American Statistical Association *Theory and Methods* (2011 –)

Life Time Data Analysis (2010 –)

Statistics in Biosciences (2009 –)

Journal of the Royal Statistical Society Series B (Statistical Methodology) (2014 –)

Journal of the American Statistical Association *Applications and Case Studies* (2010 –2012)

Biometrics (2008 – 2012)

Organizer and Chair of Invited Conference Sessions and Symposiums

Organizer/Co-organizer

Harvard Symposium on "Paths to Precision Medicine: the role of statistics" (2014)

PQG conference on "Sequencing and Complex Traits: beyond 1000 Genomes" (2012)

NSF FRG Workshop on "Recent Advances on High Dimensional Data Analysis" (2011)

Radcliffe Institute for Advanced Studies Exploratory Seminar on "High Dimensional Data Analysis" (2008)

Harvard Schering Plough Workshop (2010,2007)

Invited Session Organizer

ENAR (2015)

ICSA international symposium (2013)

New England Statistical Symposium (2011)

IMS-China International conference (2011)

First Joint Biostatistics Symposium in Beijing, China (2010)

International Conference on Statistics and Society (2010)

Joint Statistical Meeting (2010)

Organizer and Chair, IMS Medallion Lecture at ENAR (2009)

ICSA Applied Statistics Symposium (2008)

Joint Statistical Meeting (2007)

ICSA Applied Statistics Symposium (2005)

Review Panel for National and International Grants

Joint DMS/NIGMS National Science Foundation/National Institutes of Health review panel (2013)

Susan G. Komen Investigator Initiated Research - Prevention/Early Detection Panel (2010)

Doctoral Program Plus for Austrian National Science Foundation (2009)

Susan G. Komen Grant Foundation - Prevention/Risk Reduction Panel (2007–2009)

Clinical and Integrative Cardiovascular Sciences Study Section, NIH (2007–2009)

Small Grants for Cancer Epidemiology (2008)

Research Grants Council of HongKong (2006)

Biostatistical Methods and Research Design Study Section, NIH (2004)

Referee

Annals of Statistics, Biometrika, Biometrics, Biostatistics, Journal of the American Statistical Association, Journal of Computational Statistics and Data Analysis, Journal of the Royal Statistical Society, Medical Decision Making, Statistica Sinica, Statistics in Medicine, American Journal of Epidemiology, Circulation, Functional and Integrative Genomics

Publications

Methodology

1. Q. Zhou, Y. Zheng, L. Chibnik, E. Karlson and **T. Cai** “Assessing Incremental Value of Biomarkers with Multi-phase Nested case-control Studies” *Biometrics*, 2015, *accepted*.
2. Y. Shen, **T. Cai**, Y. Yang, and J. Chen “Retrospective Likelihood Based Methods for Analyzing Case-Cohort Genetic Association Studies” *Biometrics*, 2015, *accepted*.
3. S. Yu, K. Liao, S. Shaw, V. Gainer, S. Churchill, P. Szolovits, S. Murphy, I. Kohane, and **T. Cai**, “Toward High-throughput Phenotyping: Unbiased Automated Feature Extraction and Selection from Knowledge Sources”. *Journal of the American Medical Informatics Association*, 2015, *accepted*.
4. Y. Xia, **T. Cai**, and T.T. Cai “Testing Differential Networks with Applications to Detecting Gene-by-Gene Interactions”. *Biometrika*, 2015, *accepted*.
5. **T. Cai**, T.T. Cai, A. Zhang “Structured Matrix Completion with Applications to Genomic Data Integration”. *Journal of the American Statistical Association*, 2015, *accepted*.
6. J. Minnier, M. Yuan, J. Liu and **T. Cai**[†] “Risk Classification with an Adaptive Naive Bayes Kernel Machine Model.” *Journal of the American Statistical Association*, 2015, 110(509): 393-404. ([†]: a PhD student thesis paper)
7. Y. Shen, K. Liao, and **T. Cai**[†] “Sparse Kernel Machine Regression for Ordinal Outcomes”. *Biometrics*, 2015 Mar;71(1):63-70. ([†]: a PhD student thesis paper)
8. J. Sinnott, W. Dai, K. Liao, S. Shaw, A. Ananthakrishnan, V. Gainer, E. Karlson, S. Churchill, P. Szolovits, S. Murphy, I. Kohane, R. Plenge, **T. Cai**. “Improving the Power of Genetic Association Tests with Imperfect Phenotype Derived from Electronic Medical Records”. *Human Genetics*, 2014 133(11), 1369-1382.
9. R. Matsouaka, J. Li, and **T. Cai**[†] “Evaluating Marker-Guided Treatment Selection Strategies.” *Biometrics*, 2014, 70(3), 489-499. ([†]: a PhD student thesis paper)
10. L. Parast, L. Tian and **T. Cai**[†] “Landmark Estimation of Survival and Treatment Effect in a Randomized Clinical Trial” *Journal of the American Statistical Association*, 2014, 109(505):384-394. ([†]: a PhD student thesis paper)
11. **T. Cai** and Y. Zheng, Resampling Procedures for Making Inference under Nested Case Control Studies. *Journal of the American Statistical Association*, 2013, 108(504): 1532-1544.
12. Q. Zhou, Y. Zheng and **T. Cai** “Biomarker assessment with NCC studies”. *Clinical Trials*, 2013, 10(5):677-9.
13. Y. Zheng, **T. Cai** and M. S. Pepe “Adopting Nested Case-control Quota Sampling Designs for Risk Markers Evaluation”. *Lifetime Data Analysis*, 2013, 19(4):568-88.

14. J. Sinnott and **T. Cai**[†], “Omnibus Risk Assessment via Accelerated Failure Time Kernel Machine Modeling”. *Biometrics*, 2013, 69(4): 861–73. (†: a PhD student thesis paper).
15. **T. Cai**, L. Tian, D. Lloyd-Jones and L.J. Wei, “Evaluating subject-level incremental values of new markers for risk classification rule.” *Lifetime Data Analysis*, 2013, 19(4):547-67.
16. L. Parast and **T. Cai**[†], “Landmark risk prediction of residual life for breast cancer survival”. *Statistics in Medicine*, 2013, 32(20): 3459–71. (†: a PhD student thesis paper)
17. L. Zhao, L. Tian, **T. Cai**, B. Claggett, and L. J. Wei. “Effectively Selecting A Target Population for A Future Comparative Study”. *Journal of the American Statistical Association*, 2013, 108(502):527-539.
18. Y. Zheng, L. Parast, **T. Cai**, M. Brown “Evaluating Incremental Values from New Predictors with Net Reclassification Improvement in Survival Analysis” *Lifetime Data Analysis*, 2013, 19(3):350-70.
19. Q. Zhou, Y. Zheng, and **T. Cai**^{*}, “Subgroup Specific Incremental Value of New Markers for Risk Prediction.” (^{*}: corresponding author) *Lifetime Data Analysis*, 2013, 19(2), 142-169.
20. H. Uno, L. Tian, **T. Cai**, I. Kohane, and L. J. Wei, “A unified inference procedure for a class of measures to assess improvement in risk prediction systems with survival data.” *Statistics in Medicine*. 2013;32(14):2430-42.
21. L. Parast, S. Cheng and **T. Cai**[†] “Landmark Prediction of Long Term Survival Incorporating Short Term Event Time Information.” *Journal of the American Statistical Association*, 2012, 107, 1492-1501. (†: a PhD student thesis paper)
22. L Xue, H Zou and **T. Cai** . “Nonconcave penalized composite conditional likelihood estimation of sparse Ising models.” *Annals of Statistics*, 2012, 40, 1403–29.
23. D. Liu, **T. Cai**, and Y. Zheng, Evaluating the Predictive Value of Biomarkers with Stratified Case-Cohort Design. *Biometrics*, 2012, 68, 1219–27,.
24. **T. Cai**, X. Lin and R. J. Carroll, “Identifying Genetic Marker Sets Associated with Phenotypes via an Efficient Adaptive Score Test”. *Biostatistics*, 2012, 13 (4): 776–90.
25. **T. Cai** and Y. Zheng, “Evaluating Prognostic Accuracy of biomarkers under Nested Case-control Studies. *Biostatistics*, 2012, 13:89–100.
26. Y. Zheng, **T. Cai**, Y. Jin, and Z. Feng, “Evaluating Prognostic Accuracy of Biomarkers under Competing Risk. *Biometrics*, 2012, 68, 388–96.
27. L. Tian, **T. Cai**, L. Zhao and L.J. Wei, “On the covariate-adjusted estimation for an overall treatment difference with data from a randomized comparative clinical trial”. *Biostatistics*, 2012, 13: 256-73.
28. L. Leon and **T. Cai**[†], “Model Checking Techniques for Assessing Functional Form Specifications in Censored Linear Regression Models. *Statistica Sinica*, 2012, 509-30. (†: a PhD student thesis paper)
29. X. Lin, **T. Cai**, M. Wu, Q. Zhou, D. Christiani, and X. Lin, “Kernel Machine SNP-set Analysis for Censored Survival Outcomes in Genome-wide Association Studies”. *Genetic Epidemiology*, 2011, 35:620–31.
30. H. Uno, **T. Cai**, L. Tian, and L.J. Wei, “Graphical Procedures for Evaluating Overall and Subject-specific Incremental Values from New Predictors with Censored Event Time Data. *Biometrics*, 2011, 67:1389–96.

31. M. Wu, S. Lee, **T. Cai**, Y. Li, M. Boehnke and X. Lin, “Rare Variant Association Testing Using the Sequence Kernel Association Test (SKAT) for Sequence Studies.” *American Journal of Human Genetics*, 2011, 8, 82–93.
32. J. Minnier, L. Tian, and **T. Cai**[†], “A Perturbation Method for Inference on Regularized Regression Estimates. *Journal of the American Statistical Association*, 2011, , 106 (496) 1655-61. ([†]: a PhD student thesis paper)
33. **T. Cai** and Y. Zheng, “Non-parametric Evaluation of Biomarker Accuracy under Nested Case-control Studies.” *Journal of the American Statistical Association*, 2011, 106(494): 569-580.
34. **T. Cai**, L. Tian, and D. Lloyd-Jones, “Comparing Costs Associated with Risk Stratification Rules for t -year Survival”. *Biostatistics*, 2011, 12: 597–609.
35. L. Parast, S. Cheng and **T. Cai**[†], “Incorporating short-term outcome information to predict long-term survival with discrete markers.” *Biometrical Journal*, 53: 294–307. ([†]: a PhD student thesis paper)
36. **T. Cai**, G. Tonini and X. Lin, “Kernel Machine Approach to Testing the Significance of Genomic Pathways.” *Biometrics*, 2011, 67: 975–86.
37. H. Uno, **T. Cai**, M. Pencina, R. D’Agostino and L.J. Wei, “On the C-Statistics for Evaluating Overall Adequacy of Risk Prediction Procedures with Censored Survival Data.” *Statistics in Medicine*, 2011, 30: 1105–17.
38. S. McDaniel, J. Minnier, R. Betensky, G. Mohapatra, Y. Shen, J. Gusella, D. Louis, and **T. Cai**[†]: “Assessing Population Level Genetic Instability via Moving Average.” *Statistics in Biosciences*, 2011, 2: 120–36. ([†]: a PhD student thesis paper)
39. **T. Cai**, L. Tian, P. Wong and L. J. Wei, “Analysis of Randomized Comparative Clinical Trial Data for Personalized Treatment Selections.” *Biostatistics*, 2011: 12: 270–82.
40. L. Tian, R. Wang, **T. Cai** and L. J. Wei, “ The Highest Confidence Density Region and Its Usage for Inferences About the Survival Function with Censored Data.” *Biometrics*, 2010, 2010, 67: 604–10.
41. **T. Cai**, T. Gerds, Y. Zheng, and J. Chen, “Robust Prediction of t -Year Survival with Data from Multiple Studies. ” *Biometrics*, 2010, 67: 436–44.
42. **T. Cai**, L. Parast, and L. Ryan “A Meta-Analysis For Rare Events.” *Statistics in Medicine*, 2010, 29 (20): 2078–89.
43. R. Wang, L. Tian, **T. Cai** and L. J. Wei, “Nonparametric Inference Procedure for Percentiles of the Random Effect Distribution in Meta Analysis”. *Annals of Applied Statistics*, 2010, 4(1):520–32.
44. Y. Zheng, **T. Cai**, J. Stanford and Z. Feng, “Semiparametric Models of Time-dependent Predictive Values of Prognostic Biomarkers”. *Biometrics*, 2010, 66(1): 50–60.
45. **T. Cai**, L. Tian, H. Uno, S. Solomon, and L. J. Wei “Calibrating Parametric Subject-Specific Risk Estimation”. *Biometrika*, 2010, 97(2):389-404.
46. L. Leon, **T. Cai** and L. J. Wei, “Robust Inferences for Covariate Effects on Survival Time with Censored Linear Regression Models.” *Statistics in BioSciences*, 2009, 1: 50–64.
47. L. Tian, **T. Cai**, M. Pfeffer, N. Piankov, P. Cremieux and L. J. Wei, “Exact and efficient inference procedure for meta analysis and its application to the analysis of independent 2×2 tables with all available data but without artificial continuity correction.” *Biostatistics*, 2009, 10: 275–81.

48. L. Tian, **T. Cai** and L.J. Wei, "Identifying subjects who benefit from additional information for better prediction of the outcome variables." *Biometrics*, 2009, 65: 894–902.
49. Y. Zheng, **T. Cai**, M. S. Pepe and W. Levy, "Time-dependent Predictive Values of Prognostic Biomarkers with Failure Time Outcome." *Journal of the American Statistical Association*, 2008, 103: 362–368.
50. T. Gerds, **T. Cai** and M. Schumacher, "The Performance of Risk Prediction Models." *Biometrical Journal* 2008, 50: 457–79.
51. B. Jiang, X. Zhang, and **T. Cai**, "Estimating the Confidence Interval for Prediction Errors of Support Vector Machine Classifiers." *Journal of Machine Learning*, 9:521–40 2008.
52. **T. Cai**, J. Huang and L. Tian, "Regularized Estimation for the Accelerated Failure Time Model." *Biometrics*, 2008, 65: 394–404.
53. **T. Cai** and L. Dodd "Regression Analysis for the Partial Area Under the ROC Curve." *Statistica Sinica* 2008, 18: 817–36.
54. **T. Cai** and S. Cheng, "Robust Combination of Multiple Diagnostic Tests for Classifying Censored Event Times." *Biostatistics*, 2008, 9(2):216-233.
55. **T. Cai**, L. Tian, S. Solomn and L.J. Wei, "Predicting Future Responses Based on Possibly Misspecified Working Models." *Biometrika*, 95(1):75-92 2008.
56. J. Ware and **T. Cai**, "Comments on *Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond* by M. J. Pencina et al." *Statistics in Medicine* 27(2), 185-187 2007.
57. H. Uno, **T. Cai**, L. Tian, and L.J. Wei, "Evaluating Prediction Rules for t-Year Survivors With Censored Regression Models." *Journal of the American Statistical Association*, 102:527-37, 2007.
58. L. Tian, **T. Cai**, E. Goetghebeur and L. J. Wei, "Model Evaluation Based on the Distribution of Estimated Absolute Prediction Error". *Biometrika*, 94:297-311 2007.
59. **T. Cai** and Y. Zheng, "Model Checking for ROC Regression Analysis." *Biometrics*, **63**: 152–63 2007.
60. **T. Cai**, P. Gilbert and S. Self, "Joint Inferences on Vaccine Efficacy Against Infection and Disease with Application to the First HIV Vaccine Efficacy Trial". *Journal of Biopharmaceutical Statistics*, **16**: 517–38 2006.
61. L. Tian and **T. Cai**, "On the Accelerated Failure Time Model for Current Status and Interval Censored Data." *Biometrika*, , **93**: 329–42 2006.
62. Y. Zheng, **T. Cai** and Z. Feng, "Application of the time-dependent ROC Curves for Prognostic Accuracy with Multiple Biomarkers". *Biometrics*, **62**: 279–8, 2006.
63. **T. Cai**, M. S. Pepe, Y. Zheng, T. Lumley and N. S. Jenney, "The Sensitivity and Specificity of Markers for Event Times." *Biostatistics*, **72**: 182–97 2006.
64. M. S. Pepe, **T. Cai** and G. Longton, "Robust Binary Regression for Optimally Combining Predictors". *Biometrics*, **62**: pp 221–9 2006.
65. **T. Cai**, L. Tian and L. J. Wei , "Semiparametric Box-Cox Power Transformation Models for Censored Survival Observations." *Biometrika*, **92**: 619–32 2005.
66. **T. Cai** and C. Moskowitz, : "Semiparametric Estimation of the Binormal ROC Curve". *Biostatistics*, **5**: 573–86, 2004.

67. M. S. Pepe and **T. Cai***, “Analysis of Placement Values for Evaluating Discriminatory Measures.” *Biometrics*, **60**: 528 – 35 2004 (*: *equal contributors*).
68. **T. Cai**, “Semi-parametric ROC Regression Analysis with Placement Values.” *Biostatistics*, **5**: 45–60, 2004.
69. **T. Cai** and S. C. Cheng , “Semi-parametric Regression for Doubly Censored Data.” *Biometrika*, **91**: 277 – 90, 2004.
70. **T. Cai** and R. A. Betensky , “Hazard Regression for Interval Censored Data with Penalized Spline.” *Biometrics*, **59**: pp 570–9 2003.
71. **T. Cai** and M. S. Pepe , “Semi-parametric ROC Analysis to Evaluate Biomarker for Disease.” *Journal of the American Statistical Association*, **97**: 1099 – 107 2002.
72. **T. Cai**, R.J. Hyndman and M. Wand , “Mixed Model-based Hazard Estimation.” *Journal of the Computational and Graphical Statistics*, **11**: 784–98 2002.
73. **T. Cai**, S. Cheng and L.J. Wei , “Semi-Parametric Random Effects Models for Clustered Failure Time Data.” *Journal of the American Statistical Association*, **97**: 514–2, 2002.
74. **T. Cai** , L.J. Wei and M. Wilcox , “Semi-Parametric Regression Analysis for Clustered Failure Time Data.” *Biometrika*, 867–878 2000.
75. **T. Cai** and L.J. Wei , “Regression Analysis for Multivariate Failure Time Observations.” *Asymptotics in Statistics and Probability*, 47–70 2000.
76. **T. Cai** and J. Shen, “Boundedness is Redundant in a Theorem of Daubechies.” *Appl. Comp. Harm. Anal.*, volume 6, 400–40, 1999.

Applications

77. H. Uno, J. Wittes, H. Fu, S. Solomon, B. Claggett, L. Tian, **T. Cai**, M. Pfeffer, S. Evans and L.J. Wei “Alternatives to hazard ratios for comparing safety of therapies in non-inferiority studies” *Annals of Internal Medicine*, 2015, *accepted*.
78. Ananthakrishnan AN, Cagan A, **Cai T**, Gainer VS, Shaw SY, Churchill S, Karlson EW, Murphy SN, Kohane I, Liao KP, Xavier RJ. Common genetic variants influence circulating vitamin D levels in inflammatory bowel diseases. *Inflammatory Bowel Diseases* 2015 (in press)
79. A. Ananthakrishnan, A. Cagan, **T. Cai**, V. Gainer, S. Shaw, S. Churchill, E. Karlson, S. Murphy, I. Kohane, K. Liao “Diabetes and risk of infections with immunomodulator therapy in inflammatory bowel diseases”. *Alimentary Pharmacology Therapeutics*, 2015, *accepted*.
80. K. Liao, **T. Cai**, G. Savova, S. Murphy, E. Karlson, A. Ananthakrishnan, V. Gainer, S. Shaw, Z. Xia, P. Szolovits, S. Churchill, I. Kohane “Methods to develop electronic medical record phenotype algorithms incorporating natural language processing”. *BMJ*, 2015, *accepted*.
81. V. Castro, J. Minnier, S. Murphy, I. Kohane, S. Churchill, V. Gainer, **T. Cai**, Y. Dai, S. Block, M. Nadal-Vicens, A. Pollastri, J. Rosenquist, A. Hirschberg, S. Goryachev, P. Sklar, R. Perlis, J. Smoller. Validation of Electronic Health Record Phenotyping of Bipolar Disorder and Controls. *American Journal of Psychiatry*, 2015 Apr 1;172(4):363-72.
82. Ananthakrishnan AN, Cheng A, Cagan A, **Cai T**, Gainer VS, Shaw SY, Churchill S, Karlson EW, Murphy SN, Kohane I, Liao KP. Mode of child-birth and long-term outcomes in women with inflammatory bowel diseases. *Digestive Diseases and Sciences* 2015 Feb;60(2):471-7.

83. Kessler, R.C., Warner, C.H., Ivany, C., Petukhova, M.V., Rose, S. Bromet, E.J., Brown III, M., **Cai, T.**, Colpe, L.J., Cox, K.L., Fullerton, C.S., Gilman, S.E., Gruber, M.J., Heeringa, S.G., Lewandowski-Romps, L., Li, J., Millikan-Bell, A.M., Naifeh, J.A., Nock, M.K., Rosellini, A.J., Sampson, N.A., Schoenbaum, M., Stein, M.B., Wessely, S., Zaslavsky, A.M., Ursano, R.J. Predicting U.S. Army suicides after hospitalizations with psychiatric diagnoses in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *JAMA Psychiatry*, 2015, 72(1), 49-57.
84. Ananthakrishnan AN, Cagan A, Gainer VS, Cheng SC, **Cai T**, Scoville E, Konijeti GG, Szolovits P, Shaw SY, Churchill S, et al. Thromboprophylaxis Is Associated With Reduced Post-hospitalization Venous Thromboembolic Events in Patients With Inflammatory Bowel Diseases. *Clin Gastroenterol Hepatol*. 2014 Nov;12(11):1905-10.
85. Yu, S., Kumamaru, K., George, E., Dunne, R.M., Bedayat, A., Neykov, M., Hunsaker, A.R., Dill, K.E., **Cai, T.**, and Rybicki, F.J. . Classification of CT Pulmonary Angiography Reports by Presence, Chronicity, and Location of Pulmonary Embolism with Natural Language Processing. *J. Biomed. Inform.*, 2014, 52, 386-393.
86. Ananthakrishnan AN, Cagan A, **Cai T**, Gainer VS, Shaw SY, Churchill S, Karlson EW, Murphy SN, Kohane I, Liao KP. Colonoscopy is associated with a reduced risk for colon cancer and mortality in patients with inflammatory bowel diseases. *Clinical Gastroenterology and Hepatology* 2014, 13(2), 322-329.
87. Wardenaar, K.J., van Loo, H.M., **T. Cai**, T., Fava, M., Gruber, M.J., Li, J., de Jonge, P., Nierenberg, A.A., Petukhova, M.V. Rose, S., Sampson, N.A., Schoevers, R.A., Wilcox, M.A., Alonso, J., Bromet E.J., Bunting, B., Florescu, S.E., Fukao, A., Gureje, O., Hu, C., Huang, Y.Q., Karam, A.N., Levinson, D., Medina Mora, M.E., Posada-Villa, J., Scott, K.M., Taib, N.I., Viana, M.C., Xavier, M., Zarkov, Z., Kessler, R.C. The effects of comorbidity in defining major depression subtypes associated with long-term course and severity. *Psychological Medicine*, 2014, 44(15), 3289-3302.
88. A. Ananthakrishnan, S. Cheng, **T. Cai**, A. Cagan, V. Gainer, P. Szolovits, S. Shaw, S. Churchill, E. Karlson, S. Murphy, I. Kohane, K. Liao. Serum Inflammatory Markers and Risk of Colorectal Cancer in Patients with Inflammatory Bowel Diseases. *Clinical Gastroenterology and Hepatology*, *Clin Gastroenterol Hepatol.*, 2014, 12(8), 1342-1348.
89. Ananthakrishnan, A., Cagan, A., Gainer, V., Cheng, S., **Cai, T.**, Szolovits, P., Shaw, S., Churchill, S., Karlson, E. W., Murphy, S., Kohane, I. and Liao, K. P. (2014), Higher plasma vitamin D is associated with reduced risk of Clostridium difficile infection in patients with inflammatory bowel diseases. *Alimentary Pharmacology Therapeutics*, 2014, 39(10), 1136-1142
90. M. Ong, Isaac Kohane, **T. Cai** M. Gorman, K. Mendl “Population-Level Evidence for an Autoimmune Etiology of Epilepsy”, *JAMA Neurology*, 2014, 71(5), 569-574.
91. Ananthakrishnan AN, Cagan A, Gainer VS, Cheng S, **Cai T**, Szolovits P, Shaw SY, Churchill S, Karlson EW, Murphy SN, Kohane I, Liao KP. Mortality and Extraintestinal cancers in patients with Primary sclerosing cholangitis and inflammatory bowel disease. *Journal of Crohn's and Colitis* 2014, 8(9), 956-963.
92. P. S. Lai, J. Hang, F. Zhang, X. Lin, H. Dai, **T. Cai**, D. Christiani. “Occupational endotoxin exposure, impaired lung function, and death in Shanghai textile workers: a 30-year prospective cohort study”. *Occupational and Environmental Medicine*, 2014, 71(2), 118-125.

93. Ananthakrishnan, A. N., Cheng, S. C., Cai, T., Cagan, A., Gainer, V. S., Szolovits, P., ... Liao, K. P. Association Between Reduced Plasma 25-Hydroxy Vitamin D and Increased Risk of Cancer in Patients With Inflammatory Bowel Diseases. *Clinical Gastroenterology and Hepatology*, 2014, 12(5), 821-827.
94. A. Ananthakrishnan, V., Gainer, **T. Cai**, R. Perez, S. Cheng, G. Savova, G., ... K. P. Liao. "Similar risk of depression and anxiety following surgery or hospitalization for Crohn's disease and ulcerative colitis." *The American journal of gastroenterology*, 2013, 108(4), 594-601.
95. H.M. van Loo, **T. Cai**, M. Gruber, J. Li, P. de Jonge, M. Petukhova, S. Rose, N.A. Sampson, R.A. Schoevers, K.J. Wardenaar, M.A. Wilcox, A.O. Al-Hamzawi, L.H. Andrade, E.J. Bromet, B. Bunting, J. Fayyad, S.E. Florescu, O. Gureje, C. Hu, Y. Huang, D. Levinson, M. Medina-Mora, Y. Nakane, J. Posada-Villa, K.M. Scott, M. Xavier, Z. Zarkov, R.C. Kessler, "Major depressive disorder subtypes to predict long-term course. Depression and Anxiety." *Depression and Anxiety*, to appear.
96. Z. Xia, E. Secor, L. Chibnik, R. Bove, S. Cheng, T. Chitnis, A. Cagan, V. S Gainer, P. Chen, K. Liao, S. Shaw, A. Ananthakrishnan, P. Szolovits, H. Weiner, E. Karlson, S. N Murphy, G. Savova, **T. Cai**, S. Churchill, R. Plenge, I. S Kohane, and P. De Jager. "Modeling Disease Severity in Multiple Sclerosis Using Electronic Health Records." *Plos One*, 2013, 8(11): e78927. doi:10.1371/journal.pone.0078927.
97. G. Altschuler, O. Hofmann, I. Kalatskaya, R. Payne, S. Sui, U. Saxena, A. Krivtsov, S. Armstrong, **T. Cai**, L. Stein and W. Hide "Pathway Fingerprinting: An integrative approach to understand the functional basis of disease". *Genome Medicine*, 2013, 5:68 doi:10.1186/gm472.
98. K. Liao, D. Diogo, J. Cui, **T. Cai**, Y. Okada, V. Gainer, S. Murphy, N. Gupta, D. Mirel, A. Ananthakrishnan, P. Szolovits, S. Shaw, S. Raychaudhuri, S. Churchill, I. Kohane, E. Karlson, R. Plenge, "Association between low density lipoprotein and rheumatoid arthritis genetic factors with low density lipoprotein levels in rheumatoid arthritis and non-rheumatoid arthritis controls." *Ann Rheum Dis*. 2013, doi:10.1136/annrheumdis-2012-203202.
99. K. Liao, **T. Cai**, V Gainer, S Murphy, A Cagan, C Liu, S Churchill, S Shaw, I. Kohane, D. Solomon, R. Plenge and E. Karlson, "Lipid and lipoprotein levels and trends in rheumatoid arthritis compared to the general population". *Arthritis Care Res (Hoboken)*. 2013 Dec;65(12):2046-50.
100. A. Ananthakrishnan, A. Cagan, V. Gainer, **T. Cai**, S. Cheng, G. Savova, P. Chen, P. Szolovits, Z. Xia, P. De Jager, S. Shaw, S. Churchill, E. Karlson, I. Kohane, R. Plenge, S. Murphy, K. Liao, "Normalization of Plasma 25-hydroxy Vitamin D is Associated with Reduced Risk of Surgery in Crohn's Disease". *Inflamm Bowel Dis*. 2013 Aug;19(9):1921-7.
101. Gandhi RT, Bosch RJ, Aga E, Bedison MA, Bastow B, Schmitz JL, Siliciano JD, Siliciano RF, Eron JJ, Mellors JW; ACTG A5173 team. Residual plasma viraemia and infectious HIV-1 recovery from resting memory CD4 cells in patients on antiretroviral therapy: results from ACTG A5173. *Antivir Ther*. 2013;18(4):607-13.
102. A. Ananthakrishnan, V. Gainer, R. Perez, **T. Cai**, S. Cheng, G. Savova, P. Chen, P. Szolovits, Z. Xia, P. De Jager, S. Shaw, S. Churchill, E. Karlson, I. Kohane, R. Plenge, S. Murphy, K. Liao. "Psychological co-morbidity is Associated with Increased risk of Surgery in Crohn's disease". *Aliment Pharmacol Ther*. 2013 Feb;37(4):445-54.
103. K. Liao, F. Kurreeman, G. Li, G. Duclos, S Murphy, R Guzman, **T. Cai**, Namrata Gupta, V.

- Gainer, P. Schur, J. Cui, J. Denny, P. Szolovits, S. Churchill, I. Kohane, E. Karlson, R. Plenge. "Associations of autoantibodies, autoimmune risk alleles, and clinical diagnoses from the electronic medical records in rheumatoid arthritis cases and non-rheumatoid arthritis controls". *Arthritis Rheum.* 2013 Mar;65(3):571-81.
104. A. Ananthakrishnan, **T. Cai**, G. Savova, S. Cheng, P. Chen, R. Guzman Perez, V. Gainer, S. Murphy, P. Szolovits, Z. Xia, S. Shaw, S. Churchill, E. Karlson, I. Kohane, R. Plenge, K. Liao. "Improving Case Definition of Crohn's Disease and Ulcerative Colitis in Electronic Medical Records Using Natural Language Processing: a Novel Informatics Approach ". *Inflamm Bowel Dis.* 2013;19(7):1411-1420.
105. S. Wassertheil-Smoller, A. McGinn, M. Allison, **T. Cai**, D. Curbe, C. Eaton, S. Hendrix, R. Kaplan, M. Ko, L. Martin, X. Xue. "Improvement in stroke risk prediction: role of C-reactive protein and lipoprotein-associated phospholipase A2 in the women's health initiative." *Int J Stroke.* 2012 Oct 23. doi: 10.1111/j.1747-4949.2012.00860.x.
106. A. Ananthakrishnan, R. Guzman-Perez, V. Gainer, **T. Cai**, S. Churchill, I. Kohane, R. M. Plenge, S. Murphy. "Predictors of severe outcomes associated with Clostridium difficile infection in patients with inflammatory bowel disease". *Aliment Pharmacol Ther.* 2012;35(7):789-95.
107. R. Carroll, W. Thompson, A. Eyler, A. Mandeli, **T. Cai**, R. Zink, J. Pacheco, C. Boomersshine, E. Karlson, R. Perez, E. Ruderman, R. Pope, R. Plenge, A. Kho, K. Liao, J. Denny. "Portability Of An Algorithm To Identify Rheumatoid Arthritis In Electronic Health Records". *J Am Med Inform Assoc.* 2012 Jun;19(e1):e162-9.
108. M. Lu, S. Demetri, **T. Cai**, L. Parast, H. Ersoy, A. Whitmore, S. Goldaber, and F. Rybicki, Axial and reformatted 4-chamber RV/LV diameter ratios on CT pulmonary angiography as predictors of death after acute pulmonary embolism. *AJR Am J Roentgenol.* 2012 Jun;198(6):1353-60.
109. R. Perlis, D. Iosifescu, V. Castro, S. Murphy, V. Gainer, J. Minnier, **T. Cai**, S. Goryachev, P. Gallagher, M. Fava, J. Weilburg, S. Churchill, I. Kohane, and J. Smoller, Using Electronic Medical Records to Enable Large-Scale Studies in Psychiatry: Treatment Resistant Depression as a Model. *Psychological Medicine*, 2011, 41, 41–50.
110. L. Qi, L. Parast, **T. Cai**, C. Powers, E. Gervino, T. Hauser, F. Hu, A. Doria Genetic Susceptibility to Coronary Heart Disease in Type 2 Diabetes: Three Independent Studies. *Journal of the American College of Cardiology*, 2011, 58:2675–82.
111. F. Kurreeman, K. Liao, L. Chibnik, B. Hickey, E. Stahl, V. Gainer, G. Li, L. Bry, S. Mahan, K. Ardlie, B. Thomson, P. Szolovits, S. Churchill, S. Murphy, **T. Cai**, S. Raychaudhuri, I. Kohane, E. Karlson, R. Plenge. Genetic basis of autoantibody positive and negative rheumatoid arthritis risk in a multi-ethnic cohort derived from electronic health records. *American Journal of Human Genetics*, 2011, 88, 57–69.
112. T. Love, **T. Cai**, E. Karlson, "Validation of an algorithm to accurately define psoriatic arthritis using natural language processing of patient visit notes in a large database". *Seminars in Arthritis and Rheumatism*, 2011, 40(5):413-20.
113. K. Liao, **T. Cai**, V. Gainer, S. Goryachev, Q. Zeng, S. Raychaudhuri, P. Szolovits, S. Churchill, S. Murphy, I. Kohane, MD, E. Karlson, R. Plenge. "Utilizing electronic medical records for discovery research in rheumatoid arthritis". *Arthritis Care Research* 2010, 62(8):1120-7.
114. X. Gao, C. Lai, T. Scott, **T. Cai**, K. Tucker . "Urinary 8-hydroxy-2-deoxyguanosine and cognitive

- function in Puerto Rican adults”. *American Journal of Epidemiology* 2010, 172(3):271-8.
115. F. Rybicki, D. Mitsouras, C. Owens, A. Whitmore, M. Gerhard-Herman, N. Wake, **T. Cai**, Q. Zhou, M. Conte, M. Creager, R. Mulkern. “Multi-contrast high spatial resolution black blood inner volume three-dimensional fast spin echo MR imaging in peripheral vein bypass grafts. *Int J Cardiovasc Imaging* 2010; 26(6), 683-91.
116. A. Hunsaker, I. Oliva, **T. Cai**, B. Trotman-Dickenson, R. Gill, H. Hatabu, F. Rybicki “Contrast Opacification Using a Reduced Volume of Iodinated Contrast Material in CT Pulmonary Angiography: Objective and Subjective Evaluation of Larger and Smaller Pulmonary Arteries.” *American Journal of Roentgenology* 2010; 195:W118-24.
117. C. Wu, A. Sodickson, **T. Cai**, N. Levit, F. Rybicki, S. Ledbetter S. “Comparison of Respiratory Motion Artifact in 64-MSCT Pulmonary Angiography from Cranio-Caudal Versus Caudocranial Scanning.” *American Journal of Radiology* 2010; 195:155-9.
118. M. Steigner, D. Mitsouras, A. Whitmore, H. Otero, C. Wang, O. Buckley, N. Levit, A. Hussain, **T. Cai**, R. Mather, . Smedby, M. DiCarli, F. Rybicki “Iodinated Contrast Opacification Gradients in Normal Coronary Arteries Imaged with Prospectively ECG-Gated Single Heart Beat 320-Detector Row Computed Tomography.” *Circulation* 2010; 3:179-86.
119. M. Steigner, H. Otero, **T. Cai**, D. Mitsouras, L. Nallamshetty, A. Whitmore, H. Ersoy, M. Di Carli, F. Rybicki, “Narrowing the Phase Window in Prospectively ECG-Gated Single Heart Beat 320-Detector Row Coronary CT Angiography”, *International Journal of Cardiovascular Imaging* 2009; 25(1), 85-90.
120. C. Wu, **T. Cai**, D. Merberg, S. Kasif, and M. Steffen, “A Predictive Phosphorylation Signature Of Lung Cancer”. *PLoS ONE*, 2009, 4(11): e7994.
121. M. Cornelis, L. Qi, C. Zhang, P. Kraft, J. Manson, **T. Cai**, D. Hunter and F. Hu, “Joint Effects of Common Genetic Variants on the Risk of Type 2 Diabetes in U.S. Men and Women”. *Annals of Internal Medicine*, 2009, 150: 541–50.
122. M. Lu, **T. Cai**, H. Ersoy, A. Whitmore, S. Goldhaber and F. Rybicki, “Comparison of ECG-gated versus non-gated CT ventricular measurements in thirty patients with acute pulmonary embolism”. *International Journal of Cardiovascular Imaging*, 2009, 25: 101–7.
123. K. Chien, **T. Cai**, H. Hsu, T. Su, M. Chen, Y. Lee, F. Hu, “A Prediction Model for Type 2 diabetes Risk in Chinese”. *Diabetologia*, 2009, 52: 443–50.
124. F. Rybicki, H. Otero, M. Steigner, G. Vorobiof, L. Nallamshetty, D. Mitsouras, H. Ersoy, R. Mather, P. Judy, **T. Cai**, K. Coyner, K. Schultz, A. Whitmore, M. Di Carli, “Initial Evaluation of Coronary Images from 320-detector Row Computed Tomography. *International Journal of Cardiovascular Imaging*, 24: 535–46.
125. M. Lu, **T. Cai**, H. Ersoy, A. Whitmore, R. Quiroz, S. Goldhaber, and F. Rybicki, “Prediction of 30-day mortality after acute pulmonary embolism (PE) using interval increase in RV/LV diameter ratio between a negative prior and a positive diagnostic CT. *Radiology*, 246(1):281–7 2008.
126. J. MacKenzie, J. Nazario-Larrieu, **T. Cai**, S. Ledbetter, M. Duran-Mendicuti, T. Vargas, P. Judy, F. Rybicki. “Reduced dose computed tomography - Effect on reader evaluation for the detection of pulmonary embolism.” *American Journal of Roentgenology*, 189:1371-9 2007.
127. H. Ersoy, S. Goldhaber, **T. Cai**, T. Luu, J. Rosebrook, R. Mulkern, R and F. Rybicki, “Time resolved MR angiography: a primary screening exam in patients with suspected pulmonary

embolism and contraindication to iodinated contrast material. *American Journal Roentgenology*, 188(5):1246-54 2007.

128. S. K. Chuang and **T. Cai**, “Predicting Clustered Dental Implant Survival Using Frailty Methods.” *Journal of Dental Research*, 85(12):1147-1151 2006.
129. S. K. Chuang, **T. Cai**, C. Douglass, L. J. Wei, T. Dodson, “Frailty Approach for the Analysis of Clustered Failure Time Observations in Dental Research”, *J. of Dental Research*, **84**: 54–8 2005.
130. J. Rogus, **T. Cai** and L.J. Wei, “Issues in genomic screening: Critical values, sample sizes, and the ability to detect linkage.” *Genet. Epidemiol.*, **17**: S697-S701, 1999.

Recent Invited Talks

Joint Statistical Meeting (Aug 2015)
60th World Statistics Congress, Rio de Janeiro, Brazil (July, 2015)
IMS-Microsoft Research Workshop: Foundations of Data Science (June, 2015)
Big Data Science symposium, University of Ghent, Belgium (May 2015)
Plenary Speaker, First International Conference on Big Data & Applied Statistics, Beijing China (Nov 2014)
Distinguished Lecture, MD Anderson Cancer Center (Nov 2014)
University of Michigan (Oct 2014)
International Society of Clinical Biostatistics, Vienna Austria (Aug 2014)
IMS, Sydney Australia (Jul 2014)
Banff workshop on Statistical genetics (Jun 2014)
University of Wisconsin Madison (Apr 2014)
Columbia University (Feb 2014)
NIH/NICHD (Jan 2014)
ICSA International Symposium (Dec 2013)
University of Huston (Dec 2013)
University of Washington (Oct 2013)
University of Georgia (Oct 2013)
FDA, Center for Devices and Radiological Health (Sept 2013)
IMS China, Chengdu, China (Jul 2013)
INSERM Workshop on the Evaluation of predictive models, France (May 2013)
Eastern Mediterranean Region IBS conference (Apr 2013)
Joint Statistical Meeting (Aug 2013)
ENAR Spring Meeting (Mar 2013)
Rand Corporation (Jan 2013)
Department of Industrial and Systems Engineering, USC (Jan 2013)
Johnson and Johnson Annual Statistical Conference (Nov 2012)
Division of Biostatistics, Albert Einstein College of Medicine (Nov 2012)
FDA/Industry Workshop 2011, Washington DC (Sep 2012)
Joint Statistical Meeting (Aug 2012)
Workshop on statistical inference in complex/high-dimensional problems, Vienna, Austria (Jul 2012)
ISBIS, Bangkok Thailand (Jun 2012)
ICSA Applied Symposium, Boston MA (Jun 2012)

Conference on “Emerging Statistical Issues in Biomarker Validation for Clinical Trials”, University of Pennsylvania (Apr 2012)
Department of Biostatistics, Yale University (Apr 2012)
Department of Biostatistics, Johns Hopkins University (Apr 2012)
National Cancer Institute, NIH (Mar 2012)
Department of Biostatistics, UNC (Mar 2012)
DIA/FDA Co-sponsored Conference on Tailored Therapeutics (Oct 2011)
Risk Assessment Conference, University of Maryland, College Park (Oct 2011)
IMS Asia Pacific Rim Meeting, Tokyo, Japan (Jul 2011)
Canada CRM workshop on Analysis of survival and event history data (May 2011)
Dept of Statistics, University of Missouri (Apr 2011)
2011 International Indian Statistical Association Conference (Apr 2011)
ASA Connecticut Chapter annual mini-conference, (Mar 2011)
Dept of Statistics, Brown University (Mar 2011)
Dept of Statistics, University of Connecticut (Feb 2011)
Dept of Mathematics and Statistics, Boston University (Dec 2010)
Ann. Conf. of the International Society for Clinical Biostatistics, Montpellier, France (Aug 2010)
Joint Statistical Meeting (Aug 2010)
International Conference on Statistical Analysis of Complex Data, Kunming, China (Jul 2010)
International Conference on Statistics and Society, Beijing, China (Jul 2010)
First Joint Biostatistics Symposium, Beijing, China (Jul 2010)
Dept of Statistics, University of Vienna, Austria (Jun 2010)
DFG German Research Foundation Symposium “Statistical Modeling and Data Analysis in Clinical Epidemiology”, Freiburg, Germany (Jun 2010)
ENAR Spring Meeting (Mar 2010)
Conference on “Resampling Methods and High Dimensional Data”, Texas A& M (Mar 2010)
Oberwolfach Workshop on “Statistical Issues in Prediction: what can be learned for individualized predictive medicine?”, Mathematisches Forschungsinstitut Oberwolfach, Germany (Jan 2010)
Dept of Statistics, University of Urbana Champaign (Sep 2009)
Joint Statistical Meeting (Aug 2009)
International Conference on “Nonparametric Modeling & Its Application” Chengdu, China (Jul 2009)
IMS Asia Pacific Rim Meeting, Korea (Jun 2009)
Workshop on “Innovation and Inventiveness in Statistics Methodologies”, Yale University (May 2009)
Dept of Biostatistics, University of Michigan (Nov 2008)
Dept of Biostatistics, University of Pittsburgh (Oct 2008)
Joint Statistical Meeting (Aug 2008)
IMS-China International Conference on Statistics and Probability (Jun, 2008)
International Chinese Statistical Association Conference (Jun 2008)
Breast Cancer Surveillance Consortium Statistical Meeting (Apr 2008)
Fred Hutchinson Cancer Research Center (Apr 2008)
Computational Seminar Series, Harvard Medical School (Mar 2008)
ENAR International Biometrics Society (Mar 2008)
University of Copenhagen, Copenhagen Denmark (Jan 2008)
University of Pennsylvania (Nov 2007)

Radcliffe Institute for Advanced Studies (Sep 2007)
International Chinese Statistical Association International Conference (Jun 2007)
Harvard Schering Plough Workshop (Jun 2007)
Dept of Biostatistics, University of Rochester (May 2007)
Dana Farber Biostatistics Computational Biology Dinner Workshop (Mar 2007)
Dept of Statistics, University of Minneapolis (Nov 2006)
Joint Statistical Meeting (Aug 2006)
National Cancer Institute (Jun 2006)
Fred Hutchinson Cancer Research Center (Sep 2005)
Joint Statistical Meeting (Aug 2005)
Division of Biostatistics, Yale University (Jan 2005)
Division of Biostatistics, Boston University (Nov 2004)
Dept of Biostatistics, Columbia University (Oct 2004)
Division of Biostatistics, University of California at San Francisco (Sep 2004)
Joint Statistical Meeting (Aug 2004)
Dept of Statistics, University of Munich, Germany (Jun 2004)
National Institute of Child Health and Human Development (May 2004)
Dept of Biostatistics, Brown University (Mar 2004)