Daniel B. Neill

21 West End Avenue #4009
New York, NY, U.S.A.
Work: (646) 997-0588
Cell: (412) 657-2922

E-mail: daniel.neill@nyu.edu
http://www.cs.nyu.edu/∼neill
Fax: (646) 997-0550
U.S. Citizen

Employment

2018-present: Associate Professor of Computer Science and Public Service (with tenure) and Associate Professor of Urban Analytics, Courant Institute Department of Computer Science, Wagner School of Public Service, and Center for Urban Science and Progress, New York University.

2020-present: Affiliated Faculty, Center for Data Science, New York University.

2022-present: Affiliated Faculty, Department of Computer Science and Engineering, Tandon School of Engineering, New York University.

2012-2018: Associate Professor of Information Systems, H. John Heinz III College of Information Systems and Public Policy, Carnegie Mellon University. Director, Event and Pattern Detection Laboratory. Dean's Career Development Professorship (endowed chair), 2012-2015. Appointment with tenure, 2015-2018.

2016-2018: Visiting Professor of Urban Analytics, Center for Urban Science and Progress, New York University.

2006-2012: Assistant Professor of Information Systems, H. John Heinz III College, Carnegie Mellon University.

2006-2018: Courtesy Faculty, Machine Learning Department, School of Computer Science, Carnegie Mellon University.

2006-2018: Courtesy Faculty, Robotics Institute, School of Computer Science, Carnegie Mellon University.

2007-2018: Adjunct Faculty, Department of Biomedical Informatics, School of Medicine, University of Pittsburgh.

Education

Ph.D. (Computer Science), **Carnegie Mellon University**, 2006.
Thesis: "Detection of Spatial and Spatio-Temporal Clusters"
Advisor: Andrew W. Moore

M.S. (Computer Science), **Carnegie Mellon University**, 2004.
Advisor: Andrew W. Moore

M.Phil. (Computer Speech), **Cambridge University**, 2002.
Thesis: "Fully Automatic Word Sense Induction by Semantic Clustering"
Advisor: Karen Sparck-Jones

B.S.E. (Electrical Engineering/Computer Science), **Duke University**, 2001.

Graduated summa cum laude, 1st in class, with highest distinction.
Thesis: "Optimality Under Noise"
Advisor: David Kraines

Publications

Book Chapters

1. D. Zeng, Z. Cao, and D. B. Neill. AI-enabled public health surveillance: from local detection to global epidemic monitoring and control. In L. Xing, M. L. Giger, and J. K. Min, eds., *Artificial Intelligence in Medicine*, 437-453, 2021.

2. D. B. Neill. Bayesian scan statistics. In J. Glaz and M. V. Koutras, eds., *Handbook of Scan Statistics*, 2019.

3. D. B. Neill. Subset scanning for event and pattern detection. In S. Shekhar and H. Xiong, eds., *Encyclopedia of GIS, 2nd ed.*, 2218-2228, 2017.

4. S. Speakman, S. Somanchi, E. McFowland III, and D. B. Neill. Disease surveillance, case study. In R. Alhajj and J. Rokne, eds., *Encyclopedia of Social Network Analysis and Mining*, 380-385, 2014.

5. D. B. Neill, G. F. Cooper, K. Das, X. Jiang, and J. Schneider. Bayesian network scan statistics for multivariate pattern detection. In J. Glaz, V. Pozdnyakov, and S. Wallenstein, eds., *Scan Statistics: Methods and Applications*, 221-250, 2009.

6. D. B. Neill and A. W. Moore. Methods for detecting spatial and spatio-temporal clusters. In M. Wagner, A. Moore, and R. Aryel, eds., *Handbook of Biosurveillance*, 243-254, 2006.

7. D. B. Neill and A. W. Moore. Efficient scan statistic computations. In A. Lawson and K. Kleinman, eds., *Spatial and Syndromic Surveillance for Public Health*. Chichester, UK: Wiley, 189-202, 2005.

Refereed Journal Articles

8. B. Jakubowski, S. Somanchi, E. McFowland III, and D. B. Neill. Exploiting discovered regression discontinuities to debias conditioned-on-observable estimators. *Journal of Machine Learning Research* 24(133): 1-57, 2023.

9. B. Allen, D. B. Neill, R. C. Schell, J. Ahern, B. Hallowell, M. Krieger, V. A. Jent, W. C. Goedel, A. R. Cartus, J. L. Yedinak, C. Pratty, B. D. L. Marshall, and M. Cerda. Translating predictive analytics for public health practice: a case study of overdose prevention in Rhode Island. *American Journal of Epidemiology*, 2023, in press.

10. C. A. Pehlivanian and D. B. Neill. Efficient optimization of partition scan statistics via the Consecutive Partitions Property. *Journal of Computational and Graphical Statistics* 32(2): 712-729, 2023.

11. M. Nobles, R. Lall, R. W. Mathes, and D. B. Neill. Presyndromic surveillance for improved detection of emerging public health threats. *Science Advances* 8(44): eabm4920, 2022.

12. R. C. Schell, B. Allen, W. C. Goedel, B. D. Hallowell, R. Scagos, Y. Li, M. S. Krieger, D. B. Neill, B. D. L. Marshall, M. Cerda, J. Ahern. Identifying predictors of opioid overdose death at a neighborhood level with machine learning. *American Journal of Epidemiology* 191(3): 526-533, 2022.

13. B. D. L. Marshall, N. Alexander-Scott, J. L. Yedinak, B. Hallowell, W. C. Goedel, B. Allen, R. C. Schell, Y. Li, M. S. Krieger, C. Pratty, J. Ahern, D. B. Neill, M. Cerda. Preventing overdose using information and data from the environment (PROVIDENT): Protocol for a randomised, population-based, community intervention trial. *Addiction* 117(4): 1152-1162, 2022.

14. D. J. Fitzpatrick, Y. Ni, and D. B. Neill. Support vector subset scan for spatial pattern detection. *Computational Statistics and Data Analysis* 157: 107149, 2021.

15. K. Klemmer, D. B. Neill, and S. A. Jarvis. Understanding spatial patterns in rape reporting delays. *Royal Society Open Science* 8: 201795, 2021.

16. S. A. Ibrahim, M. E. Charlson, and D. B. Neill. Big data analytics and the struggle for equity in health care: the promise and perils. *Health Equity* 4(1): 99-101, 2020.

17. W. Herlands, D. B. Neill, H. Nickisch, and A. G. Wilson. Change surfaces for expressive multidimensional changepoints and counterfactual prediction. *Journal of Machine Learning Research* 20(99): 1-51, 2019.

18. D. J. Fitzpatrick, W. L. Gorr, and D. B. Neill. Keeping score: predictive analytics in policing. *Annual Review of Criminology* 2: 473-491, 2019.

19. R.C.S.N.P. Souza, R. M. Assuncao, D. M. Oliveira, D. B. Neill, and W. Meira Jr. Where did I get dengue? Detecting spatial clusters of infection risk with social network data. *Spatial and Spatio-temporal Epidemiology* 29: 163-175, 2019.

20. Y. Sun, V. V. Murty, R. Leeman-Neill, C. Soderquist, D. Park, D. B. Neill, G. Bhagat, and B. Alobeid. Cytogenetic analysis of adult T-Cell leukemia/lymphoma: evaluation of a Caribbean cohort. *Leukemia & Lymphoma* 60: 1598-1600, 2019.

21. S. Somanchi, D. B. Neill, and A. V. Parwani. Discovering anomalous patterns in large digital pathology images. *Statistics in Medicine* 37: 3599-3615, 2018.

22. M. de Arteaga, W. Herlands, D. B. Neill, and A. Dubrawski. Machine learning for the developing world. *ACM Transactions on Management Information Systems* 9(2): 9.1-9.14, 2018.

23. D. B. Neill and W. Herlands. Machine learning for drug overdose surveillance. *Journal of Technology in Human Services* 36(1): 8-14, 2018. Presented at *Bloomberg Data for Good Exchange Conference*, 2017.

24. S. Speakman, S. Somanchi, E. McFowland III, and D. B. Neill. Penalized fast subset scanning. *Journal of Computational and Graphical Statistics*, 25(2): 382-404, 2016. **Selected for "Best of JCGS" invited session by the journal's editor in chief.**

25. B. J. Bushman, K. Newman, S. L. Calvert, G. Downey, M. Dredze, M. Gottfredson, N. G. Jablonski, A. Masten, C. Morrill, D. B. Neill, D. Romer, and D. Webster. Youth violence: what we know and what we need to know. *American Psychologist* 71(1): 17-39, 2016.

26. S. Flaxman, D. B. Neill, and A. Smola. Gaussian processes for independence tests with non-iid data in causal inference. *ACM Transactions on Intelligent Systems and Technology*, 7(2): 22:1-22:23, 2015.

27. Z. Faigen, L. Deyneka, A. Ising, D. B. Neill, M. Conway, G. Fairchild, J. Gunn, D. Swenson, I. Painter, L. Johnson, C. Kiley, L. Streichert, H. Burkom. Cross-disciplinary consultancy to bridge public health technical needs and analytic developers: asyndromic surveillance use case. *Online Journal of Public Health Informatics*, 7(3): e228, 2015.

28. S. Speakman, E. McFowland III, and D. B. Neill. Scalable detection of anomalous patterns with connectivity constraints. *Journal of Computational and Graphical Statistics*, 24(4): 1014-1033, 2015.

29. D. Gartner, R. Kolisch, D. B. Neill, and R. Padman. Machine learning approaches for early DRG classification of inpatient data in hospitals. *INFORMS Journal of Computing*, 27(4): 718-734, 2015.

30. F. Chen and D. B. Neill. Human rights event detection from heterogeneous social media graphs. *Big Data*, 3(1): 34-40, 2015.

31. E. McFowland III, S. Speakman, and D. B. Neill. Fast generalized subset scan for anomalous pattern detection. *Journal of Machine Learning Research*, 14: 1533-1561, 2013.

32. D. B. Neill, E. McFowland III, and H. Zheng. Fast subset scan for multivariate event detection. *Statistics in Medicine*, 32: 2185-2208, 2013.

33. D. B. Neill. Fast subset scan for spatial pattern detection. *Journal of the Royal Statistical Society (Series B: Statistical Methodology)*, 74(2): 337-360, 2012.

34. S. Hasan, G. T. Duncan, D. B. Neill, and R. Padman. Automatic detection of omissions in medication lists. *Journal of the American Medical Informatics Association*, 18(4): 449-458, 2011.

35. D. B. Neill. Fast Bayesian scan statistics for multivariate event detection and visualization. *Statistics in Medicine*, 30: 455-469, 2011.

36. R. J. Leeman-Neill, R. R. Seethala, S. V. Singh, M. L. Freilino, J. S. Bednash, S. M. Thomas, M. C. Panahandeh, W. E. Gooding, S. C. Joyce, M. W. Lingen, D. B. Neill, and J. R. Grandis. Inhibition of EGFR-STAT3 signaling with erlotinib prevents carcinogenesis in a chemically-induced mouse model of oral squamous cell carcinoma. *Cancer Prevention Research*, 4(2): 230-237, 2011.

37. D. Oliveira, D. B. Neill, J. H. Garrett Jr., and L. Soibelman. Detection of patterns in water distribution pipe breakage using spatial scan statistics for point events in a physical network. *Journal of Computing in Civil Engineering*, 25(1): 21-30, 2011.

38. D. B. Neill and G. F. Cooper. A multivariate Bayesian scan statistic for early event detection and characterization. *Machine Learning* 79: 261-282, 2010.

39. X. Jiang, D. B. Neill, and G. F. Cooper. A Bayesian network model for spatial event surveillance. *International Journal of Approximate Reasoning* 51: 224-239, 2010.

40. R. J. Leeman-Neill, Q. Cai, S. C. Joyce, S. M. Thomas, N. E. Bhola, D. B. Neill, J. L. Arbiser, and J. R. Grandis. Honokiol inhibits epidermal growth factor receptor signaling and enhances the antitumor effects of epidermal growth factor receptor inhibitors. *Clinical Cancer Research* 16(9): 2571-2579, 2010.

41. R. J. Leeman-Neill, S. E. Wheeler, S. V. Singh, S. M. Thomas, R. R. Seethala, D. B. Neill, M. C. Panahandeh, E.-R. Hahm, S. C. Joyce, M. Sen, Q. Cai, M. L. Freilino, C. Li, D. E. Johnson, and J. R. Grandis. Guggulsterone enhances head and neck cancer therapies via inhibition of signal transducer and activator of transcription-3. *Carcinogenesis* 30(11): 1848-1856, 2009.

42. D. B. Neill. Expectation-based scan statistics for monitoring spatial time series data. *International Journal of Forecasting* 25: 498-517, 2009.

43. D. B. Neill. An empirical comparison of spatial scan statistics for outbreak detection. *International Journal of Health Geographics* 8: 20, 2009.

44. D. B. Neill. Cascade effects in heterogeneous populations. *Rationality and Society* 17(2): 191-241, 2005.

45. D. B. Neill. Evolutionary stability for large populations. *Journal of Theoretical Biology* 227(3): 397-401, 2004.

46. M. M. Wagner, F.-C. Tsui, J. Espino, W. Hogan, J. Hutman, J. Hersh, D. Neill, A. Moore, G. Parks, C. Lewis, and R. Aller. A national retail data monitor for public health surveillance. *Morbidity and Mortality Weekly Report, Supplement on Syndromic Surveillance* 53: 40-42, 2004.

47. D. B. Neill. Optimality under noise: higher memory strategies for the Alternating Prisoner's Dilemma. *Journal of Theoretical Biology* 211(2): 159-180, 2001.

Refereed Conference Proceedings

48. K. Rosman and D. B. Neill. Detecting anomalous networks of opioid prescribers and dispensers in prescription drug data. *Proc. 37th AAAI Conf. on Artificial Intelligence*, 2023, in press.

49. P. Ravishankar, Q. Mo, E. McFowland III, and D. B. Neill. Provable detection of propagating sampling bias in prediction models. *Proc. 37th AAAI Conf. on Artificial Intelligence*, 2023, in press.

50. K. Klemmer, N. S. Safir, and D. B. Neill. Positional encoder graph neural networks for geographic data. *Proc. 26th Intl. Conf. on Artificial Intelligence and Statistics*, *PMLR* 206: 1379-1389, 2023. Also presented at *Proc. NeurIPS Workshop on Tackling Climate Change with Machine Learning*, 2022.

51. K. Klemmer, T. Xu, B. Acciaio, and D. B. Neill. SPATE-GAN: Improved generative modeling of dynamic spatio-temporal patterns with an autoregressive embedding loss. *Proc. 36th AAAI Conf. on Artificial Intelligence*, 4523-4531, 2022.

52. C. Wang, D. B. Neill, and F. Chen. Calibrated nonparametric scan statistics for anomalous pattern detection in graphs. *Proc. 36th AAAI Conf. on Artificial Intelligence*, 4201-4209, 2022.

53. K. Klemmer and D. B. Neill. Auxiliary-task learning for geographic data with autoregressive embeddings. *Proc. 29th ACM SIGSPATIAL Intl. Conf. on Advances in Geographic Information Systems*, 141-144, 2021.

54. R.C.S.N.P. Souza, R. Assuncao, D. B. Neill and W. Meira Jr. Detecting spatial clusters of disease infection risk using sparsely sampled social media mobility patterns. *Proc. 27th ACM SIGSPATIAL Intl. Conf. on Advances in Geographic Information Systems*, 359-368, 2019.

55. K. Klemmer, D. B. Neill, and S. A. Jarvis. Modeling rape reporting delays using spatial, temporal and social features. *Proc. NeurIPS Workshop on Modeling and Decision-Making in the Spatio-Temporal Domain*, 2018.

56. R.C.S.N.P. Souza, R. Assuncao, D. B. Neill, L.G.S. Silva, and W. Meira Jr. Spatial risk modeling for infectious disease surveillance using population movement data. *Proc. NeurIPS Workshop on Modeling and Decision-Making in the Spatio-Temporal Domain*, 2018.

57. W. Herlands, E. McFowland III, A. G. Wilson, and D. B. Neill. Automated local regression discontinuity design discovery. *Proc. 24th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1512-1520, 2018.

58. W. Herlands, E. McFowland III, A. G. Wilson, and D. B. Neill. Gaussian process subset scanning for anomalous pattern detection in non-iid data. *Proc. 21st International Conference on Artificial Intelligence and Statistics*, *PMLR* 84: 425-434, 2018.

59. Z. Zhang and D. B. Neill. Identifying significant predictive bias in classifiers. Presented at *NIPS Workshop on Interpretable Machine Learning for Complex Systems*, 2016, and *4th Workshop on Fairness, Accountability, and Transparency in Machine Learning*, 2017.

60. W. Herlands, A. G. Wilson, H. Nickisch, S. Flaxman, D. B. Neill, W. van Panhuis, and E. P. Xing. Scalable Gaussian processes for characterizing multidimensional change surfaces. *Proc. 19th International Conference on Artificial Intelligence and Statistics*, *PMLR* 51: 1013-1021, 2016.

61. W. Herlands, M. de Arteaga, D. B. Neill, and A. Dubrawski. Lass-0: Sparse non-convex regression by local search. *Proc. 8th NIPS Workshop on Optimization for Machine Learning*, 2015.

62. S. Flaxman, A. G. Wilson, D. B. Neill, H. Nickisch, and A. Smola. Fast Kronecker inference in Gaussian processes with non-Gaussian likelihoods. *Proc. 32nd International Conference on Machine Learning*, *PMLR* 37: 607-616, 2015.

63. F. Chen and D. B. Neill. Non-parametric scan statistics for event detection and forecasting in heterogeneous social media graphs. *Proc. 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1166-1175, 2014.

64. S. Speakman, Y. Zhang, and D. B. Neill. Dynamic pattern detection with temporal consistency and connectivity constraints. *Proc. 13th IEEE International Conference on Data Mining*, 697-706, 2013.

65. S. Somanchi and D. B. Neill. Discovering anomalous patterns in large digital pathology images. *Proc. 8th INFORMS Workshop on Data Mining and Health Informatics*, 2013.

66. K. Shao, Y. Liu, and D. B. Neill. A generalized fast subset sums framework for Bayesian event detection. *Proceedings of the 11th IEEE International Conference on Data Mining*, 617-625, 2011.

67. X. Jiang, D. B. Neill, and G. F. Cooper. Generalized AMOC curves for evaluation and improvement of event surveillance. *Proceedings of the American Medical Informatics Association Annual Symposium*, 281-285, 2009.

68. S. Hasan, G. T. Duncan, D. B. Neill, and R. Padman. Towards a collaborative filtering approach to medication reconciliation. *Proceedings of the American Medical Informatics Association Annual Symposium*, 288-292, 2008.

69. K. Das, J. Schneider, and D. B. Neill. Anomaly pattern detection in categorical datasets. *Proceedings of the 14th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 169-176, 2008.

70. C. A. Harle, D. B. Neill, and R. Padman. An information visualization approach to classification and assessment of diabetes risk in primary care. *Proceedings of the 3rd INFORMS Workshop on Data Mining and Health Informatics*, 2008.

71. M. Makatchev and D. B. Neill. Learning outbreak regions in Bayesian spatial scan statistics. *Proceedings of the ICML/UAI/COLT 2008 Workshop on Machine Learning for Health Care Applications*, 2008.

72. D. B. Neill, A. W. Moore, and G. F. Cooper. A Bayesian spatial scan statistic. In Y. Weiss, *et al.*, eds. *Advances in Neural Information Processing Systems 18*, 1003-1010, 2006.

73. M. R. Sabhnani, D. B. Neill, A. W. Moore, A. Dubrawski, and W.-K. Wong. Efficient analytics for effective monitoring of biomedical security. *Proceedings of the International Conference on Information and Automation*, 2005.

74. D. B. Neill, A. W. Moore, M. R. Sabhnani, and K. Daniel. Detection of emerging space-time clusters. *Proceedings of the 11th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 218-227, 2005.

75. D. B. Neill and A. W. Moore. Anomalous spatial cluster detection. *Proceedings of the KDD 2005 Workshop on Data Mining Methods for Anomaly Detection*, 2005.

76. M. R. Sabhnani, D. B. Neill, A. W. Moore, F.-C. Tsui, M. M. Wagner, and J. U. Espino. Detecting anomalous patterns in pharmacy retail data. *Proceedings of the KDD 2005 Workshop on Data Mining Methods for Anomaly Detection*, 2005.

77. P. Hsiung, A. Moore, D. Neill, and J. Schneider. Alias detection in link data sets. *Proceedings of the First International Conference on Intelligence Analysis*, 2005.

78. D. B. Neill, A. W. Moore, F. Pereira, and T. Mitchell. Detecting significant multidimensional spatial clusters. In L. K. Saul, *et al.*, eds., *Advances in Neural Information Processing Systems 17*, 969-976, 2005.

79. D. B. Neill and A. W. Moore. Rapid detection of significant spatial clusters. *Proceedings of the 10th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 256-265, 2004.

80. D. B. Neill and A. W. Moore. A fast multi-resolution method for detection of significant spatial disease clusters. In S. Thrun, *et al.*, eds., *Advances in Neural Information Processing Systems 16*, 651-658, 2004.

81. D. B. Neill. Cooperation and coordination in the Turn-Taking Dilemma. *Proceedings of the Ninth Conference on Theoretical Aspects of Rationality and Knowledge*, 231-244, 2003.

Refereed Conference and Journal Abstracts

82. D. B. Neill, M. Nobles, R. Lall, R. W. Mathes. Pre-syndromic disease surveillance for improved detection of emerging public health threats. Syndromic Surveillance Symposium, 2022.

83. D. B. Neill, B. Chen, A. Wei, and M. Nobles. MUSES open-source software for pre-syndromic disease surveillance. Syndromic Surveillance Symposium, 2022.

84. B. D. L. Marshall, N. Alexander-Scott, J. L. Yedinak, B. Hallowell, W. C. Goedel, B. Allen, R. C. Schell, M. S. Krieger, C. Pratty, J. Ahern, D. B. Neill, and M. Cerdá. Preventing overdose using information and data from the environment (PROVIDENT): Protocol for a randomised, population-based, community intervention trial. 9th International Conference on Health and Hepatitis Care in Substance Users, 2021.

85. M. Nobles, R. Lall, R. Mathes, and D. B. Neill. Multidimensional semantic scan for pre-syndromic disease surveillance. *Online Journal of Public Health Informatics* 11(1): e255, 2019. **Winner of the International Society for Disease Surveillance Outstanding Student or Post-Degree Abstract Award**.

86. R.C.S.N.P. Souza, D. B. Neill, R. Assuncao, and W. Meira Jr. Identifying high-risk areas for dengue infection using mobility patterns on Twitter. *Online Journal of Public Health Informatics* 11(1): e246, 2019.

87. D. B. Neill. Multidimensional tensor scan for drug overdose surveillance. *Online Journal of Public Health Informatics* 9(1): e20, 2017.

88. D. Fitzpatrick, Y. Ni, and D. B. Neill. Support vector subset scan for spatial outbreak detection. *Online Journal of Public Health Informatics* 9(1): e21, 2017.

89. S. Somanchi, D. Choi, and D. B. Neill. StarScan: a novel scan statistic for irregularly-shaped spatial clusters. *Online Journal of Public Health Informatics* 7(1): e55, 2015.

90. M. Nobles, L. Deyneka, A. Ising, and D. B. Neill. Identifying emerging novel outbreaks in textual emergency department data. *Online Journal of Public Health Informatics* 7(1): e45, 2015.

91. F. Chen and D. B. Neill. Non-parametric scan statistics for disease outbreak detection on Twitter. *Online Journal of Public Health Informatics* 6(1): e155, 2014.

92. S. Flaxman and D. B. Neill. New tests for space-time interaction in spatio-temporal point processes. *Proceedings of the 2nd Spatial Statistics Conference*, 2013.

93. D. B. Neill and T. Kumar. Fast multidimensional subset scan for outbreak detection and characterization. *Online Journal of Public Health Informatics* 5(1), 2013.

94. S. Speakman, Y. Zhang, and D. B. Neill. Tracking dynamic water-borne outbreaks with temporal consistency constraints. *Online Journal of Public Health Informatics* 5(1), 2013.

95. Y. Liu and D. B. Neill. Detecting previously unseen outbreaks with novel symptom patterns. *Emerging Health Threats Journal* 4: 11074, 2011.

96. S. Somanchi and D. B. Neill. Fast graph structure learning from unlabeled data for outbreak detection. *Emerging Health Threats Journal* 4: 11017, 2011.

97. S. Speakman, E. McFowland III, S. Somanchi, and D. B. Neill. Scalable detection of irregular disease clusters using soft compactness constraints. *Emerging Health Threats Journal* 4: 11121, 2011.

98. D. B. Neill, C. Harle, R. Padman, F. Solano, and J. Zgibor. Clinician information needs for data visualization based diabetes risk assessment and guideline compliance. *Proceedings of the American Medical Informatics Association Annual Symposium*, 2011.

99. D. B. Neill, E. McFowland III, and H. Zheng. Fast subset scan for multivariate spatial biosurveillance. *Emerging Health Threats Journal*, 4:s42, 2011.

100. D. B. Neill and Y. Liu. Generalized fast subset sums for Bayesian detection and visualization. *Emerging Health Threats Journal*, 4:s43, 2011.

101. H. Zheng, R. Padman, D. B. Neill, and S. Hasan. A comparison of collaborative filtering methods for medication reconciliation. *Proceedings of the 13th International Congress on Medical Informatics*, 2010.

102. D. Gartner, R. Kolisch, R. Padman, and D. B. Neill. Early DRG classification of inpatients in hospitals. *Proceedings of the 35th Conference on Operational Research Applied to Health Services*, 2009.

103. D. B. Neill. Fast subset sums for multivariate Bayesian scan statistics. *Proceedings of the International Society for Disease Surveillance Annual Conference*, available online at www.syndromic.org, 2009.

104. S. Speakman and D. B. Neill. Fast graph scan for scalable detection of arbitrary connected clusters. *Proceedings of the International Society for Disease Surveillance Annual Conference*, 2009.

105. K. Das, J. Schneider, and D. B. Neill. Anomaly pattern detection for biosurveillance. *Advances in Disease Surveillance* 5: 19, 2008.

106. M. Makatchev and D. B. Neill. Learning outbreak regions for Bayesian spatial biosurveillance. *Advances in Disease Surveillance* 5: 45, 2008.

107. D. B. Neill. Fast and flexible outbreak detection by linear-time subset scanning. *Advances in Disease Surveillance* 5: 48, 2008.

108. D. B. Neill and W. L. Gorr. Detecting and preventing emerging epidemics of crime. *Advances in Disease Surveillance* 4: 13, 2007.

109. D. B. Neill and J. Lingwall. A nonparametric scan statistic for multivariate disease surveillance. *Advances in Disease Surveillance* 4: 106, 2007.

110. D. B. Neill. Incorporating learning into disease surveillance systems. *Advances in Disease Surveillance* 4: 107, 2007.

111. D. B. Neill. An empirical comparison of spatial scan statistics for outbreak detection. *Advances in Disease Surveillance* 4: 259, 2007.

112. D. B. Neill, A. W. Moore, and G. F. Cooper. A multivariate Bayesian scan statistic. *Advances in Disease Surveillance* 2: 60, 2007.

113. D. B. Neill and M. R. Sabhnani. A robust expectation-based spatial scan statistic. *Advances in Disease Surveillance* 2: 61, 2007.

114. D. B. Neill, A. W. Moore, and G. F. Cooper. A Bayesian scan statistic for spatial cluster detection. *Advances in Disease Surveillance* 1: 55, 2006. **Received "Best Research Presentation" award, National Syndromic Surveillance Conference, 2005**.

115. D. B. Neill, A. W. Moore, M. R. Sabhnani, and K. Daniel. An expectation-based scan statistic for detection of space-time clusters. *Advances in Disease Surveillance* 1: 56, 2006.

116. M. R. Sabhnani, D. B. Neill, A. W. Moore, F.-C. Tsui, M. M. Wagner, and J. U. Espino. Monitoring pharmacy retail data for anomalous space-time clusters. *Advances in Disease Surveillance* 1: 62, 2006.

117. D. B. Neill, A. W. Moore, and M. R. Sabhnani. Detecting elongated disease clusters. *Morbidity and Mortality Weekly Report, Supplement on Syndromic Surveillance* 54: 197, 2005.

118. D. B. Neill and A. W. Moore. A fast grid-based scan statistic for detection of significant spatial disease clusters. *Morbidity and Mortality Weekly Report, Supplement on Syndromic Surveillance* 53: 255, 2004.

Other Journal Articles

119. A. Magdy, X. Zhou, and D. B. Neill. Guest Editorial: Special issue on analytics for local events and news. *Geoinformatica* 24: 267-268, 2020.

120. S. Somanchi and D. B. Neill. Graph structure learning from unlabeled data for early outbreak detection. *IEEE Intelligent Systems* 32(2): 80-84, 2017.

121. D. B. Neill. Using artificial intelligence to improve hospital inpatient care. *IEEE Intelligent Systems* 28(2): 92-95, 2013.

122. C. A. Harle, D. B. Neill, and R. Padman. Information visualization for chronic disease risk assessment. *IEEE Intelligent Systems* 27(6): 81-85, 2012.

123. D. B. Neill. New directions in artificial intelligence for public health surveillance. *IEEE Intelligent Systems* 27(1): 56-59, 2012.

124. S. W. Malone, C. A. Miller, and D. B. Neill. Traffic flow models and the evacuation problem. *Undergraduate Journal of Mathematics and its Applications* 22(3): 273-292, 2001. **Winner of the 2001 Mathematical Contest in Modeling**.

125. S. W. Malone, J. A. Mermin, and D. B. Neill. Air traffic control. *Undergraduate Journal of Mathematics and its Applications* 21(3): 227-241, 2000. **Winner of the 2000 Mathematical Contest in Modeling**.

Technical reports and theses

126. D. B. Neill. Detection of spatial and spatio-temporal clusters. Ph.D. thesis, Carnegie Mellon University, Department of Computer Science, Technical Report CMU-CS-06-142, 2006.

127. D. B. Neill and A. W. Moore. Detecting space-time clusters: prior work and new directions. Carnegie Mellon University, Department of Computer Science, Technical Report CMU-CS-05-115, 2005.

128. D. B. Neill. Evolutionary dynamics with large aggregate shocks. Carnegie Mellon University, Department of Computer Science, Technical Report CMU-CS-03-197, 2003.

129. D. B. Neill. An evolutionary resolution to the Finitely Repeated Prisoner's Dilemma paradox. Carnegie Mellon University, Department of Computer Science, Technical Report CMU-CS-03-155, 2003.

130. D. B. Neill. Fully automatic word sense induction by semantic clustering. Cambridge University, M.Phil. thesis, 2002.

131. D. B. Neill. Optimality under noise: higher memory strategies for the Alternating Prisoner's Dilemma. Duke University, undergraduate honors thesis, 2000.

Submitted for publication

132. K. Boxer, E. McFowland III, and D. B. Neill. "Auditing predictive models for intersectional biases."

133. B. Allen, R. C. Schell, V. A. Jent, M. Krieger, C. Pratty, B. D. Hallowell, M. Basta, W. C. Goedel, J. L. Yedinak, Y. Li, A. R. Cartus, B. D. L. Marshall, M. Cerda, J. Ahern, and D. B. Neill. "PROVIDENT: Development and validation of a machine learning model to predict neighborhood-level overdose risk in Rhode Island."

134. C. Koziatek, I. Bohart, R. Caldwell, J. Swartz, P. Rosen, S. Desai, K. Krol, D. B. Neill, and D. Lee. "Neighborhood-level risk factors for severe hyperglycemia among Emergency Department patients without a prior diabetes diagnosis."

135. I. Bohart, D. B. Neill, and D. C. Lee. "Fairness and bias of machine learning in healthcare and medicine."

136. I. C. Bohart, J. R. Caldwell, J. L. Swartz, P. Rosen, N. Genes, C. A. Koziatek, D. B. Neill*, and D. C. Lee* (co-senior authors). "Fairness and bias of machine learning approaches for diabetes screening in the emergency department."

137. N. Menghani, E. McFowland III, and D. B. Neill. "Insufficiently justified disparate impact: a new criterion for fair recommendations."

138. K. S. Boxer, B. Hong, C. E. Kontokosta, and D. B. Neill. "Estimating reporting bias in 311 complaint data."

139. W. Herlands and D. B. Neill. "Automated discovery of difference-in-differences."

140. E. McFowland III, S. Somanchi, and D. B. Neill. "Efficient discovery of heterogeneous treatment effects in randomized experiments via anomalous pattern detection."

141. D. J. Fitzpatrick, W. Gorr, and D. B. Neill. "Policing chronic and temporary hot spots of violent crime: a controlled field experiment."

142. R. F. Davies, D. B. Neill, *et al.* "Detection of the Walkerton gastroenteritis outbreak by text mining of emergency room health records."

143. S. Flaxman, A. Gelman, D. B. Neill, A. Smola, A. Vehtari, and A. G. Wilson. "Fast hierarchical Gaussian processes."

144. A. Maurya, K. Murray, Y. Liu, C. Dyer, W. W. Cohen, and D. B. Neill. "Semantic scan: detecting subtle, spatially localized events in text streams."

145. T. Kumar and D. B. Neill. "Fast tensor scan for event detection and characterization."

146. S. Flaxman and D. B. Neill. "Detecting spatially localized subsets of leading indicators for event prediction."

147. K. Das, J. Schneider, and D. B. Neill. "Detecting anomalous groups in categorical datasets."

Invited Talks and Tutorials

148. "The role of AI/ML in reaching health equity with big data," Sanofi, paid advisory board presentation, December 2022.

149. "Machine learning and event detection for urban public health," Michigan Technological University (virtual), invited colloquium with honorarium, December 2022.

150. "Machine learning and event detection for urban public health," Ludwig Maximilian University of Munich (virtual), June 2022.

151. "Use-inspired artificial intelligence and machine learning for public and population health," University of Texas at Austin, McCombs School of Business (virtual), April 2022.

152. "Machine learning and event detection for urban public health," invited keynote, Workshop on Urban Complex Systems (UCS-CCS 2021), Lyon, France (hybrid), October 2021.

153. "Machine learning for opioid and overdose surveillance," Symposium on Artificial Intelligence for Social Good (virtual), April 2021.

154. "Subset scanning for event and pattern detection," IBM Thomas J. Watson Research Center, Yorktown Heights, NY, February 2020.

155. "Machine learning and event detection for population health," invited plenary, Machine Learning for Science and Engineering Conference, Atlanta, GA, June 2019.

156. "Subset scanning for event and pattern detection," Cornell University, Department of Operations Research and Industrial Engineering, March 2019.

157. "Machine learning, automated algorithms, and risk," InsurTech Alliance, New York, NY, February 2019.

158. "Machine learning for development: challenges, opportunities, and a roadmap," NeurIPS 2018 Workshop on Machine Learning for the Developing World, Montreal, Canada, December 2018.

159. "Subset scanning for event and pattern detection," University of Connecticut, Department of Statistics, November 2018.

160. "Machine learning for population health and disease surveillance," Duke University, Department of Mathematics, November 2018.

161. "New methodological approaches for opioid and overdose surveillance," 3rd Seattle Symposium on Health Care Data Analytics, Seattle, WA, October 2018.

162. "Predictive policing in practice," Workshop on Data-Driven Criminal Justice Reform, New York, NY, October 2018.

163. "Subset scanning for event and pattern detection," invited webinar for IBM Research Africa, August 2018.

164. "Automated algorithms and risk: two sides of the coin," InsurTech Science and Engineering Innovation Expo, New York, NY, August 2018.

165. "Modeling and detecting patterns in complex urban data," Amazon, New York, NY, July 2018.

166. "Machine learning, big data, and development," International Monetary Fund, Washington, DC, May 2018.

167. "Fairness and bias in algorithmic decision-making," Big Data in Health Symposium, Cornell University, New York, NY, April 2018.

168. "Machine learning for population health and disease surveillance," Machine Learning and Medicine Seminar, Cornell University, New York, NY, March 2018.

169. "Auditing black-box algorithms for fairness and bias," Workshop on Accountable Decision Systems, New York, NY, February 2018.

170. "Novel machine learning methods for public health and disease surveillance," American Society for Microbiology Biothreats Conference, Baltimore, MD, February 2018.

171. "Event and pattern detection at the societal scale," invited keynote, *ACM SIGSPATIAL Workshop on Analytics for Local Events and News*, Redondo Beach, CA, November 2017.

172. "Detecting anomalous patterns of care using health insurance claims," *INFORMS Conference on Information Systems and Technology*, Houston, TX, October 2017.

173. "Detecting anomalous patterns of care using health insurance claims," *INFORMS Workshop on Data Science*, Houston, TX, October 2017.

174. "Multidimensional subset scanning for the public good," University of Texas at Austin, McCombs School of Business, October 2017.

175. "A pre-syndromic surveillance approach for early detection of novel and rare disease outbreaks," *Interdisciplinary Association for Population Health Science Conference*, Austin, TX, October 2017.

176. "Support vector subset scan for spatial pattern detection," *GEOMED 2017 International Conference on Spatial Statistics, Spatial Epidemiology, and Spatial Aspects of Public Health*, Porto, Portugal, September 2017.

177. "Efficient pattern detection in web-scale graphs by subcore-tree decomposition and subset scanning," *Joint Statistical Meetings*, Baltimore, MD, July 2017.

178. "Machine learning for public health and disease surveillance," New York University, Department of Population Health, January 2017.

179. "Fast subset scan for population health and disease surveillance," Weill Cornell Medical College, Department of Healthcare Policy and Research, December 2016.

180. "Anomalous and significant subgraph detection in attributed networks," half-day tutorial (with Feng Chen, Petko Bogdanov, and Ambuj Singh). *IEEE International Conference on Big Data*, Washington, D.C., December 2016.

181. "Detecting anomalous patterns of care using health insurance claims," *INFORMS Annual Meeting*, Nashville, TN, November 2016.

182. "Detecting anomalous patterns of care using health insurance claims," *Conference on Digital Experimentation*, Cambridge, MA, October 2016.

183. "Efficient discovery of heterogeneous treatment effects in randomized experiments via anomalous pattern detection," *Conference on Digital Experimentation*, Cambridge, MA, October 2016.

184. "Detecting anomalous patterns of care using health insurance claims," *Workshop on Health IT and Economics*, Washington, D.C., October 2016.

185. "Detecting anomalous patterns of care using health insurance claims," *Eighth International Workshop on Applied Probability*, Toronto, Canada, June 2016.

186. "Efficient discovery of heterogeneous treatment effects in randomized experiments via anomalous pattern detection," *Eighth International Workshop on Applied Probability*, Toronto, Canada, June 2016.

187. "Support vector subset scan for spatial pattern detection," *Eighth International Workshop on Applied Probability*, Toronto, Canada, June 2016.

188. "Fast subset scan for population health and disease surveillance," Harvard University, Department of Biostatistics, T.H. Chan School of Public Health, May 2016.

189. "Event and pattern detection for urban systems," New York University, Wagner School of Public Service, April 2016.

190. "Event and pattern detection at the societal scale," Georgia Institute of Technology, School of Computational Science and Engineering, March 2016.

191. "Event and pattern detection at the societal scale," New York University, Courant Institute, Department of Computer Science, February 2016.

192. "Event and pattern detection for urban systems," New York University, Center for Urban Science and Progress, February 2016.

193. "Novel approaches to local area spatiotemporal crime rate forecasting with Gaussian processes," *American Society of Criminology Annual Meeting*, Washington, DC, November 2015.

194. "Event and pattern detection at the societal scale," Harvard University, School of Engineering and Applied Sciences, November 2015.

195. "Event and pattern detection at the societal scale," University of Chicago, Harris School of Public Policy, October 2015.

196. "Machine learning and health: from neurons to society," *World Economic Forum: Annual Meeting of the New Champions*, Dalian, China, September 2015.

197. "Large-scale Gaussian processes for spatiotemporal modeling of disease incidence," *Joint Statistical Meetings*, Seattle, WA, August 2015.

198. "Human rights event detection from heterogeneous social media graphs," *Human Rights Media Central Workshop*, Pittsburgh, PA, July 2015.

199. "Penalized fast subset scanning," *45th Symposium on the Interface of Computing Science and Statistics* ("Best of JCGS" invited session), Morgantown, WV, June 2015.

200. "Urban predictive analytics," *INFORMS Annual Meeting*, San Francisco, CA, November 2014.

201. "Kernel space-time interaction tests for identifying leading indicators of crime," *Joint Statistical Meetings*, Boston, MA, August 2014.

202. "Discovering novel anomalous patterns in general data," *Statistical Learning and Data Mining Meeting on Data Mining in Business and Industry*, Durham, NC, June 2014.

203. "Non-parametric scan statistics for event detection and forecasting in heterogeneous social media graphs," *Seventh International Workshop on Applied Probability*, Antalya, Turkey, June 2014.

204. "A star-shaped scan statistic for detecting irregularly-shaped spatial clusters," *Seventh International Workshop on Applied Probability*, Antalya, Turkey, June 2014.

205. "Scaling up event and pattern detection to big data," NYU Stern School of Business, Information Systems Seminar, New York, NY, April 2014.

206. "Scaling up event and pattern detection to big data," *MIT Workshop on Challenges in Big Data for Data Mining, Machine Learning and Statistics*, Cambridge, MA, March 2014.

207. "Penalized fast subset scanning," *6th International Conference on Computational and Methodological Statistics*, London, UK, December 2013.

208. "Non-parametric scan statistics for event detection and forecasting in heterogeneous social media graphs," *INFORMS Annual Meeting*, Minneapolis, MN, October 2013.

209. "Machine learning and event detection for the public good," Data Science for the Social Good Summer Fellowship Program, Chicago, IL, July 2013.

210. "Fast subset scanning for scalable event and pattern detection," Stony Brook University, Stony Brook, NY, May 2013.

211. "Predicting and preventing emerging outbreaks of crime," *CMU Workshop on Machine Learning and Social Sciences*, Pittsburgh, PA, October 2012.

212. "Fast graph structure learning from unlabeled data for event detection," *INFORMS Annual Conference*, Phoenix, AZ, October 2012.

213. "Detecting spatially localized subsets of leading indicators for event prediction," *32nd International Symposium on Forecasting*, Boston, MA, June 2012.

214. "Scalable detection of anomalous patterns with connectivity constraints," *29th Quality and Productivity Research Conference*, Long Beach, CA, June 2012.

215. "Fast generalized subset scan for anomalous pattern detection," *6th International Workshop on Applied Probability*, Jerusalem, Israel, June 2012.

216. "Efficient subset scanning with soft constraints," *6th International Workshop on Applied Probability*, Jerusalem, Israel, June 2012.

217. "Analytical methods for large scale surveillance of unstructured data," *International Conference on Digital Disease Detection*, Boston, MA, February 2012.

218. "Efficient methods for anomalous pattern detection in general datasets," *INFORMS Annual Conference*, Charlotte, NC, November 2011.

219. "Fast learning of graph structure from unlabeled data for anomalous pattern detection," *INFORMS Annual Conference*, Charlotte, NC, November 2011.

220. "Dynamic pattern detection with connectivity and temporal consistency constraints," *INFORMS Annual Conference*, Charlotte, NC, November 2011.

221. "Fast multivariate subset scanning for scalable cluster detection," *Joint Statistical Meetings 2011*, Miami, FL, August 2011.

222. "Machine learning for population health and disease surveillance," *Advanced Analytics Workshop*, Washington, DC, April 2011.

223. "Spatial and subset scanning for multivariate health surveillance," *Data Fusion Research Meeting*, Ottawa, ON, March 2011.

224. "Spatial scanning tips and tricks for practical outbreak detection," invited webinar for the International Society for Disease Surveillance, January 2011.

225. "Research challenges for biosurveillance: the next ten years" (invited plenary), *International Society for Disease Surveillance Annual Conference*, Park City, UT, December 2010.

226. "Fast generalized subset scan for anomalous pattern detection," *INFORMS Annual Conference*, Austin, TX, November 2010.

227. "Scalable detection of anomalous patterns with connectivity constraints," *INFORMS Annual Conference*, Austin, TX, November 2010.

228. "Fast subset sums for scalable Bayesian detection and visualization," *Fifth International Workshop on Applied Probability*, Madrid, Spain, July 2010.

229. "Fast subset scanning for multivariate event detection," *ENAR 2010 Annual Meeting*, New Orleans, LA, March 2010.

230. "Application of spatial scan statistic methods to crime hot spot analysis," *Tenth Crime Mapping Research Conference*, New Orleans, LA, August 2009.

231. "Event detection," half-day tutorial (with Weng-Keen Wong). *15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Paris, France, June 2009.

232. "Multivariate Bayesian scan statistics for event detection and characterization," *Twelfth Biennial CDC/ATSDR Symposium on Statistical Methods*, Decatur, GA, April 2009.

233. "A nonparametric scan statistic for multivariate spatial biosurveillance," *Joint Statistical Meetings 2008*, Denver, CO, August 2008.

234. "Linear-time subset scanning," *Fourth International Workshop on Applied Probability*, Compiegne, France, July 2008.

235. "Multivariate event detection and characterization," *Washington Statistical Society Seminar*, Washington, DC, May 2008.

236. "Multivariate outbreak detection and characterization," *Donald A. B. Lindberg Lecture and Symposium*, Pittsburgh, PA, May 2008.

237. "A multivariate Bayesian method for spatial biosurveillance," *Joint Statistical Meetings 2007*, Salt Lake City, UT, July 2007.

238. "Monitoring multivariate spatial time series data for disease outbreak detection," *27th Annual International Symposium on Forecasting*, New York, NY, June 2007.

239. "Bayesian disease surveillance by detection of anomalous clusters," *Third ECADS Syndromic Surveillance Conference*, Ottawa, ON, October 2006.

240. "Bayesian disease surveillance by detection of anomalous clusters," *Third International Workshop on Applied Probability*, Storrs, CT, May 2006.

241. "Scaling up geographic disease surveillance," *Second ECADS Syndromic Surveillance Conference*, Ottawa, ON, June 2005.

Grant Funding

NSF IIS-2040898, Neill (PI), 2/1/2021-1/31/2025, supported by the NSF Program on Fairness in Artificial Intelligence in Collaboration with Amazon. "FAI: End-to-End Fairness for Algorithm-in-the-Loop Decision Making in the Public Sector". We will develop methods and tools that assist public sector organizations with fair and equitable policy interventions in areas including housing, criminal justice, and health. Total award: $1,000,000. Role: PI.

NSF IIS-1926470, Kontokosta and Neill (PIs), 10/1/2019-9/30/2021. "AI-DCL: EAGER: Bias and Discrimination in City Predictive Analytics". We will improve urban analytics based on 311 citizen complaints by developing new methods to identify systematic biases in the propensity to complain, to understand the impact of reporting bias on predictive models for allocation of city services, and to enable city agencies to account for and correct these biases. Total award: $297,652. Role: co-PI.

NIH NIDA R01-DA046620, Marshall (PI), 10/1/2019-9/30/2024. "Reducing Drug-Related Mortality Using Predictive Analytics: A Randomized, Statewide, Community Intervention Trial". This project will develop and test an opioid overdose forecasting tool which will predict areas at highest risk of future overdose deaths. It will also evaluate the impact of machine learning-based targeting of overdose prevention programs through a randomized, statewide, community-level intervention trial. Total award: $3,771,449. NYU CUSP subaward: $171,345. Role: co-I.

NYU Marron Institute, Been (PI), 7/1/2018-6/30/2019. "Making Local Governments' Housing Quality Maintenance Systems Fairer and More Efficient". We will develop and evaluate predictive analytics for targeted housing inspections by estimating the likelihood and severity of problems, accounting for biases due to differential reporting rates. Total award: $50,000. Role: co-PI.

Richard King Mellon Foundation, Neill, Gorr, Akinci, and Berges (PIs), 1/1/2016-12/31/2019. "Metro21: Knowledge-Powered Pittsburgh to Improve Urban Quality of Life." Our project, "Urban Predictive Analytics for a Safer and Cleaner Pittsburgh," will develop and deploy predictive analytics for violence prevention in Pittsburgh. By incorporating many city and county data sources, and by integrating predictive analyses at the geographic, subpopulation, and individual levels, we will provide the Pittsburgh police and the city and county leadership with situational awareness of the many inter-related factors influencing patterns of violence, assisting the development of long-term violence prevention strategies and tactical interventions. Total award: $600,000. Project award: $250,538. Role: PI.

Metro21 Initiative, Neill and Gorr (PIs), 1/1/2015-12/31/2016. "Open-Source 311 Predictive Analytics for the City of Pittsburgh." We propose to deploy, evaluate, and extend an open-source version of our CityScan predictive analytics software as part of the City of Pittsburgh's 311 call system, in collaboration with the City leadership. By predicting emerging clusters of 311 calls (non-emergency service requests) and providing support for the City's operational decisions based on the predicted clusters, we will enable the City to respond proactively and effectively to emerging challenges and citizen needs. Total award: $41,406. Role: PI.

NSF IIS-0953330, Neill (PI), 7/1/2010-6/30/2016, funded by National Science Foundation. "CAREER: Machine Learning and Event Detection for the Public Good." This project will create and explore novel methods for detection of emerging events in massive, complex, real-world datasets. This research will be integrated with a multi-pronged educational initiative to incorporate machine learning into the public policy curriculum. Total award: $529,962. Role: PI.

Disruptive Health Technology Institute, Neill (PI), unrestricted gift awarded 7/7/2014, "Discovering Anomalous Patterns of Care to Improve Health Outcomes and Reduce Costs." We plan to create a widely applicable methodological and implementation framework for using massive quantities of health insurance claims data to discover patterns of care with significant potential impacts on patient outcomes and healthcare costs. Total award: $20,000. Role: PI.

John D. and Catherine T. MacArthur Foundation, Aronson (PI), 3/1/2013-8/31/2014, awarded to CMU's Center for Human Rights Science. "Evaluating Machine Learning Methods and Tools for Use in Human Rights Work." We will develop new machine learning methods for early detection and advance prediction of conflict events, and evaluate the potential utility of these methods for enabling proactive responses to outbreaks of violence and human rights abuses. Total award: $175,000. Role: "Pattern Detection and Event Prediction" project lead.

Center for the Future of Work, Padman (PI), "Design and Implementation of a Preliminary Workbench to Support Medication Reconciliation via Machine Learning." This project will apply novel collaborative filtering methods to support the complex task of medication reconciliation at the point of care. Total award: $50,000. Role: co-PI.

UPMC Healthcare Innovation Grant, Neill (PI), unrestricted gift awarded 11/8/2010, funded by University of Pittsburgh Medical Center- Technology Development Center. "Anomalous Pattern Detection from Healthcare Data Streams." This project will apply novel pattern detection methods

to detect anomalous patterns of patient care. Total award: $121,503. Role: PI.

UPMC Healthcare Innovation Grant, Padman (PI), unrestricted gift awarded 11/8/2010, funded by University of Pittsburgh Medical Center- Technology Development Center. "Information Visualization for Cognitively Guided Decision Making for Diabetes Risk Assessment and Guideline Compliance." This project will develop and evaluate novel information visualization tools and methods for improving diabetes care. Total award: $110,120. Role: co-PI.

NSF IIS-0916345, Neill (PI), 8/1/2009-7/31/2013, funded by National Science Foundation. "III: Small: Fast Subset Scan for Anomalous Pattern Detection". This project will develop new, general subset scan methods for efficient pattern detection in massive datasets. Total award: $499,991. Role: PI.

NSF IIS-0911032, Dubrawski (PI), 9/1/2009-8/31/2014, funded by National Science Foundation. "III: Large: Discovering Complex Anomalous Patterns". This project will develop an integrated probabilistic framework for pattern discovery, incorporating detection, characterization, explanation, and learning from user feedback. Total award: $2,598,153. Role: Co-PI.

CDC 8-R01-HK000020, Dubrawski (PI), 9/30/2006-9/29/2008, funded by Centers for Disease Control and Prevention. "Efficient, Scalable, Multisource Surveillance Algorithms for BioSense". This project will develop multivariate Bayesian biosurveillance methods for inclusion in the BioSense system. Total award: $1,198,409. Role: Co-PI.

NSF IIS-0325581, Cooper (PI), 9/1/2003-8/31/2008, funded by National Science Foundation. "ITR: Bayesian Modeling for Biosurveillance". This project will develop novel Bayesian methodologies for the detection of disease outbreaks. CMU award: $1,246,800. Role: senior personnel.

CRTI-08-190RD, Davies (PI), 7/2009-6/2013, funded by CRTI. "Data Fusion Solutions for Monitoring CBRNE Threats". This project focuses on general solutions for integrating multiple data sources for public health surveillance and integrates these solutions into two specific applications, detection of severe outbreaks in hospitalized patients and surveillance of events related to illicit substance abuse. Total award: $3,000,000. Role: Technical team, expert in statistical detection methods and data mining.

CRTI-06-0234TA, Davies (PI), 7/2007-7/2010, funded by CRTI. "Advanced Syndromic Surveillance and Emergency Triage (ASSET)". This project will develop and deploy a system for syndromic surveillance of Emergency Department data in Ottawa, Ontario, for earlier detection of disease outbreaks and bioterrorist attacks. Total award: $2,000,000. CMU subcontract: $25,475. Role: Technical team, expert in statistical detection methods and data mining.

Recent Awards and Honors

Runner-up in the Department of Homeland Security's Hidden Signals Challenge, for our work on pre-syndromic disease surveillance, 2018.
Winner of the Yelp Dataset Challenge, 2016.
Dean's Career Development Professorship, Carnegie Mellon University, 2012.
Named one of the "top ten artificial intelligence researchers to watch" by *IEEE Intelligent Systems* ("AI's 10 to Watch", Jan/Feb 2011).
NSF CAREER Award, 2010.
"Best Research Presentation" award, *National Syndromic Surveillance Conference*, 2005.

NSF Graduate Research Fellowship, 2002-2005.
Winston Churchill Scholarship, 2001-2002.
Barry M. Goldwater Scholarship, 1999-2001.
Two time winner of Mathematical Contest in Modeling, 2000 and 2001.

University Awards and Honors

Carnegie Mellon University Doctoral Fellowship, 2002-2006.
Duke University, Walter J. Seeley Scholastic Award in Engineering, 2001.
Duke University, George Sherrerd Award in Electrical Engineering, 2001.
Oxford University, Lord Rothermere Scholarship, 2000.
Duke University, PRUV Research Fellowship, 2000.
Duke University, Angier B. Duke Scholarship, 1997-2001.
Honor Societies: Phi Beta Kappa, Tau Beta Pi, Eta Kappa Nu.

Teaching

Summer 2018-present: Capstone instructor for Center for Urban Science and Progress, New York University. Mentored capstone projects "Predictive modeling of opioid overdose risk for targeted public health interventions," "Accuracy and equity in predictive hot-spot policing," and "Property valuation and tax mapping from imagery data".

Spring 2017-present: Instructor for CUSP-GX-5003, Machine Learning for Cities, Center for Urban Science and Progress, New York University.
Spring 2017-present: Instructor for PADM-GP-4147/4148, Large Scale Data Analysis I and II, Wagner School of Public Service, New York University.
Spring 2010-Spring 2014: Instructor for 90-921/10-831, Special Topics in Machine Learning and Policy, Carnegie Mellon University.
Spring 2009-Spring 2016: Instructor for 90-904/10-830, Research Seminar in Machine Learning and Policy, Carnegie Mellon University.
Spring 2008-Fall 2015: Instructor for 90-866, Large Scale Data Analysis for Public Policy, Carnegie Mellon University.
Fall 2006-Fall 2015: Instructor for 95-796, Statistics for IT Managers, Carnegie Mellon University.
Spring 2005: Head teaching assistant for 15-781, Machine Learning, Carnegie Mellon University.
Spring 2004: Teaching assistant for 15-780, Advanced AI Concepts, Carnegie Mellon University.
Summer 1998: University of South Florida, Mathematics and Engineering Program. Taught undergraduate-level courses in mathematics and computer science to gifted high school students.

Students/Postdocs Advised and Thesis Committees

Primary research advisor for Martina Balestra, Smart Cities Postdoctoral Associate, Center for Urban Science and Progress, New York University.

Primary research advisor for Kate Boxer, Benjamin Jakubowski, Pavan Ravishankar, and Betty Hou, Ph.D. students, Courant Institute Department of Computer Science, New York University.

Primary research advisor for Boyuan (Jack) Chen, Ph.D. student, Department of Computer Science and Engineering, Tandon School of Engineering, New York University.

Primary research advisor for Katie Rosman, M.S. student, Department of Computer Science and Engineering, Tandon School of Engineering, New York University.

Primary research advisor for Shizhan Gong, M.S. student, Center for Data Science, New York University.

Primary research advisor for Ougni Chakraborty, M.S. student, Department of Electrical Engineering, Tandon School of Engineering, New York University.

Primary research advisor for Alexandra Lefevre, M.S. student, Department of Mathematics, Tandon School of Engineering, New York University.

Primary research advisor for Rushabh Shah, M.S. student, Department of Computer Science, Courant Institute, New York University.

Primary research advisor for Neil Menghani, M.S. student, Department of Mathematics, Courant Institute, New York University.

Primary research advisor for Devashish Khulbe, M.S. student, Center for Urban Science and Progress, New York University.

Primary research advisor for Haorui Guo, Qingyu Serene Mo, Andy Wei, and Jackson Oleson, undergraduate students, Courant Institute of Mathematical Sciences, New York University.

Primary research advisor for Ellie Haber, undergraduate student, Department of Computer Science and Engineering, Tandon School of Engineering, New York University.

Primary research advisor for Pragya Parasarathy, undergraduate student, College of Arts and Science, New York University.

Primary research advisor for Feng Chen, postdoctoral fellow, H. John Heinz III College, Carnegie Mellon University.

Primary research advisor for Skyler Speakman, Edward McFowland III, Sriram Somanchi, Abhinav Maurya, Mallory Nobles, and Zhe Zhang, Ph.D. students, H. John Heinz III College, Carnegie Mellon University.

Primary research advisor for Seth Flaxman, William Herlands, and Dylan Fitzpatrick, Ph.D. students, Joint Ph.D. Program in Machine Learning and Policy, and Adona Iosif and Rishi Chandy, Ph.D. students, Machine Learning Department, School of Computer Science, Carnegie Mellon University.

Primary research advisor for Yandong Liu, Kan Shao, Amrut Nagasunder, Kenton Murray, Xin Wu, Tarun Kumar, Kai Liu, and Dylan Fitzpatrick, M.S. students, School of Computer Science, Carnegie Mellon University.

Co-advisor for CUSP Smart Cities/Provost's Postdoctoral Fellow John Pamplin, Center for Urban Science and Progress, New York University.

Co-advisor for Ph.D. students: Bennett Allen (NYU Langone School of Medicine, Department of Population Health); Konstantin Klemmer (University of Warwick).

Co-advisor for Isaac Bohart, MD/MS student, NYU Langone School of Medicine.

Thesis committee member for Ph.D. students at Carnegie Mellon University: Kaustav Das (Machine Learning Department), Stephen Fancsali (Department of Philosophy), Jeremy Gernand (Engineering and Public Policy), Sean Green (Engineering and Public Policy), Sharique Hasan (Heinz College), Nandana Sengupta (Tepper School of Business), Yanchuan Sim (Language Technologies Institute), Ben Towne (Societal Computing), Matthew Benigni (Societal Computing).

Thesis committee member for Computer Science Ph.D. students at New York University Tandon School of Engineering and Courant Institute: Rasika Bhalerao (CSE), Nabeel Abdur Rehman (CSE), Ke Yang (CSE), Wesley Maddox (CS), Greg Benton (CS).

Thesis committee member for Xia Jiang, Ph.D. student, Department of Biomedical Informatics, School of Medicine, University of Pittsburgh.

Qualifier committee member for Jun Shi, Ph.D. student, Civil and Environmental Engineering, and Vladimir Ermakov, M.S. student, Robotics Institute, Carnegie Mellon University.

Heinz First Research Paper advisor for Chris Harle, Ph.D. student, H. John Heinz III College, Carnegie Mellon University.

Work-study and independent study advisor for Sayantan Das, Rajas Lonkar, Yun Ni, Patrick Wedgeworth, and Yating Zhang, MISM students, Heinz College, Carnegie Mellon University.

Professional Activities

Active participant in multiple large-scale efforts for development and deployment of disease surveillance systems, including the CDC BioSense project, the National Biosurveillance Integration System, the National Retail Data Monitor, three Canadian disease surveillance projects funded by CRTI, and ongoing work with North Carolina DOH and New York City DOHMH.

Established and directed a new PhD program in Machine Learning and Public Policy at Carnegie Mellon University. This joint program between the Machine Learning Department and Heinz College (School of Public Policy) is the first of its kind. The program will build bridges and encourage collaborations between researchers in machine learning, computer science, public policy, and management. It will also attract and develop a highly competitive group of students with unique skills in developing new machine learning tools and applying them to real-world policy domains.

Developed a new course, "Large Scale Data Analysis for Public Policy," for the Heinz College at Carnegie Mellon University. This master's level course enables students to tackle a wide scope of policy problems using state-of-the-art machine learning methods.

Developed a new course, "Research Seminar in Machine Learning and Policy," for the Heinz College and Machine Learning Department at Carnegie Mellon University. This Ph.D. level seminar prepares students for cutting-edge research at the intersection of machine learning and public policy through in-depth discussion of current research articles, essential topics, and ongoing projects.

Developed a new course series, "Special Topics in Machine Learning and Policy," for the Heinz College and Machine Learning Department at Carnegie Mellon University. Topics covered include Event and Pattern Detection (2010, 2014), Machine Learning for the Developing World (2011), Harnessing the Wisdom of Crowds (2012), and Mining Massive Datasets (2013).

Associate Editor for *ACM Journal on Computing and Sustainable Societies* (2022-present), *INFORMS Journal on Data Science* (2020-present), *ACM Transactions on Management Information*

Systems (2017-present), *IEEE Intelligent Systems* (2013-present), *Security Informatics* (2017-2019), and *Decision Sciences* (2017-2020). "Artificial Intelligence and Health" Department Editor for *IEEE Intelligent Systems* (2011-present). Section editor for *Handbook of Scan Statistics* (2019, Springer). Co-editor for *GeoInformatica* special issue on "Analytics of Local Events and News" (2020). Editor-in-chief selection committee member for *IEEE Intelligent Systems* (2021).

Scientific Program Chair for the *International Society for Disease Surveillance Annual Conference*, Atlanta, GA, December 2011. **Program co-chair** for *International Conference on Smart Health* (2014 and 2015). **Area chair** for *NeurIPS Workshop on AI for Social Good* (2019), *ACM Conference on Computing and Sustainable Societies* (ACM COMPASS) (2020), and *ACM Conference on Fairness, Accountability, and Transparency* (ACM FAccT) (2020). **Best paper selection committee member** for *Workshop on Information Technologies and Systems* (WITS) (2021). **Program committee member** for *International Society for Disease Surveillance Annual Conference* (2007-2009), *25th International Conference on Machine Learning* (2008), *BioSecure Workshop* (2009), *International Symposium on System Informatics and Engineering* (2011), *IEEE International Conference on Intelligence and Security Informatics* (2012-2013, 2015, 2020, 2023), *International Health Informatics Conference* (2013), *IEEE Joint Intelligence and Security Informatics Conference* (2014), *KDD Workshop on Outlier Detection and Description* (2013-2015), *International Workshop on Applied Probability* (2016), *ACM SIGSPATIAL Workshop on Learning from Events and News* (2017), *KDD Workshop on Epidemiology meets Data Mining and Knowledge Discovery* (2018), *AAAI-20 Emerging Track on AI for Social Impact* (2020), *NeurIPS Workshop on Machine Learning for Development* (2020, 2021), *AAAI "AI for Social Impact" Track* (2021), *ACM Conference on Fairness, Accountability, and Transparency* (2023), and *AAAI Conference on Artificial Intelligence* (2008, 2021-2023). **Advisory board member** for *INFORMS Workshop on Data Science*, 2017.

Served on National Science Foundation grant review panel (CCF Division), 2009. External reviewer for NIH grant review panel (NIAAA), 2011, and Swiss Data Science Center, 2017. Served on National Science Foundation CAREER review panel, 2019.

Invited panelist for MacArthur Foundation meeting on "Urban Analytics and Neighborhood Health" (Chicago, IL, May 2012), National Science Foundation subcommittee on youth violence (Washington, DC, February 2013), KDD Workshop on Data Science for the Social Good (New York, NY, August 2014), Convening on Data Science (Chicago, IL, February 2016), Workshop on Data-Driven Criminal Justice Reform (New York, NY, October 2018), and Data and Ethics Workshop (New York, NY, May 2019). **Paid advisory board member** for Sanofi, 2022.

Journal paper reviews for *Science*, *Proceedings of the National Academy of Sciences*, *Journal of Machine Learning Research*, *Machine Learning*, *Applied Mathematical Modelling*, *IEEE Transactions on Evolutionary Computation*, *IEEE Intelligent Systems*, *IEEE Transactions on Knowledge and Data Engineering*, *ACM Transactions on Knowledge Discovery in Data*, *Information Systems Research*, *PLoS Medicine*, *Management Science*, *Statistics in Medicine*, *Knowledge and Information Systems*, *Nature Scientific Reports*, *Artificial Intelligence and Law*, *PLoS Computational Biology*, *Theoretical Population Biology*, *Journal of the American Statistical Association*, *Journal of the Royal Statistical Society A*, *Environmetrics*, *Test*, *Biometrics*, *International Journal of Health Geographics*, *Geoinformatica*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *ACM Transactions on Sensor Networks*, *Computational Statistics and Data Analysis*, *Scandinavian Journal of Statistics*, *Environmental and Ecological Statistics*, *Geographical Analysis*, *Methodological*

Innovations, *Security Informatics*, *Statistical Analysis and Data Mining*, *Perspectives on Psychological Science*, and *Systems, Man, and Cybernetics B*.

Conference paper reviews for *Neural Information Processing Systems* (2006, 2017, 2019), *American Medical Informatics Association Annual Symposium* (2007), *23rd AAAI Conference on Artificial Intelligence* (2008), *25th International Conference on Machine Learning* (2008), *14th International Conference on Artificial Intelligence and Statistics* (2011), *IEEE International Conference on Intelligence and Security Informatics* (2012-2013, 2015), *KDD Workshop on Outlier Detection and Description* (2013-2015), *World Wide Web Conference* (2014), and *IEEE Joint Intelligence and Security Informatics Conference* (2014), as well as the program committees listed above.

University Service (CMU): Director, Event and Pattern Detection Laboratory. Program director, Joint Ph.D. Program in Machine Learning and Public Policy. Member of Heinz Ph.D. Committee, Heinz Scientific Computing Committee, Heinz Information Systems Management Faculty Hiring Committee. Faculty search chair, tenure-track position in societal-scale data analysis (2012) and tenure-track and research-track positions in applied statistical machine learning (2016), CMU Heinz College. Developed and implemented new procedures to reduce implicit bias and improve diversity in the College's hiring practices. Faculty search committee member, tenure-track position in operations research and analytics (2015) and tenure-track position in machine learning and economics (2016), CMU Heinz College. Co-organizer of Heinz College Faculty Research Seminars (2011-2012 and 2014-2015). Co-organizer of CMU workshop (2012) and seminar series (2013-2015) on Machine Learning and Social Sciences. Member of committee for developing the Data Analytics track of CMU's M.S. program in Public Policy and Management (2013).

University Service (NYU): Founder and Director, Machine Learning for Good Laboratory. Co-Director, NYU Urban Initiative (2019-2022). Selection committee chair, Urban Doctoral Fellowship. Member, Wagner Curriculum Committee. Member, Wagner Budget and Finance Oversight Committee. Member, Wagner PhD Committee. Co-wrote and administered Preliminary Qualifying Exam for Wagner PhD in Public Administration (2023). Search committee chair (2020) and co-chair (2019) for seven Smart Cities Postdoctoral Research Associate positions, NYU Center for Urban Science and Progress. Responsible AI faculty search committee (2023). CUSP contract faculty search committee (2023).

International Society for Disease Surveillance: Served on the ISDS Advisory Group, formed to advise the ISDS Board of Directors on the Society's mission, strategic plan, and activities, and the ISDS Analytic Solutions Advisory Group, formed to advise on public health consultancies and to facilitate the transfer of public health surveillance methodology from research to practice.

Member of Sigma Xi, Association of Computing Machinery, International Institute of Forecasters, and American Statistical Association.