

Big Data: Implications for the Liberal Arts Curriculum AALAC Workshop Proposal

1. Big Data: Implications for the Liberal Arts Curriculum

Location: Wellesley College

Dates: A Friday and Saturday, April or June 2015

Rationale

The changing scale, nature, and structure of modern data sets is quickly transforming the skills needed for liberal arts students pursuing careers in public policy, technology, health, business, journalism, and beyond. This rise of “big data” has implications across the curriculum, as new modes of inquiry in fields from the sciences to the social sciences to the humanities increasingly depend on statistical and computational tools. We must modify the conceptual and practical learning goals in quantitative courses, create relevant modules for courses in a variety of disciplines, and develop new courses. There is also growing demand for support among faculty and research students working with big data. Meeting this demand involves both faculty-oriented training and reaching students who might not otherwise enroll in quantitative courses. This workshop will focus on how best to adapt the liberal arts curriculum for a big data world, both in the classroom and through research.

A Wellesley-hosted workshop seems particularly appropriate, as Wellesley’s new Quantitative Analysis Institute is designed to expand the role of statistics across the curriculum and forge interdisciplinary connections.

Description

This workshop targets faculty who teach and guide curricula in fields impacted by big data, including statistics, science, social science, and digital humanities. Topics include:

- What is big data?
- How does the prevalence of big data affect learning goals? What skills do students need to succeed in the workplace and graduate school?
- What role does big data play in research at liberal arts colleges? What training would best allow students to participate in this research? What training do faculty need to lead this research?

The workshop will include presentations on incorporating big data into the curriculum and opportunities to develop new resources. Workshop leaders will present approaches to teaching and supporting research currently employed at AALAC colleges, and a prominent outside speaker will discuss approaches used elsewhere. Participants will brainstorm syllabi, teaching tools, assignments, and faculty training that can be refined and shared.

Goals and Intended Impact

The workshop will:

- Establish working definition of “big data” to guide curricular changes
- Gather, share, and examine approaches to incorporating big data into the classroom
- Lead to new pedagogical resources
- Propose ways to reach student researchers who interact with big data
- Outline big data workshops designed for faculty
- Explore inter-institutional collaborations to support big data analysis

Preliminary Schedule

Day 1 - Big data in the classroom

Implications for introductory, intermediate, and advanced courses

Day 2 - Big data in research

Implications for faculty and student researchers

The schedule will include presentations, full-group discussions, breakout sessions, and time to explore tools and build teaching materials.

Follow-up

We will survey participants after the workshop and again after six months to assess preliminary impact. Findings will be described in a report and a paper for potential publication.

Pedagogical materials and recommendations from the workshop will also be disseminated.

2. Workshop leaders and planning team

Cassandra Pattanayak, Jack and Sandra Polk Guthman '65 Director,
Quantitative Analysis Institute, **Wellesley College**

Corrine Taylor, Director, Quantitative Reasoning Program, **Wellesley College**

Nicholas Horton, Professor of Statistics, **Amherst College**

Amy Wagaman, Assistant Professor of Statistics, **Amherst College**

Benjamin Baumer, Visiting Assistant Professor,
Department of Mathematics and Statistics, **Smith College**;
Director, Statistical and Data Sciences Program, **Smith College** (starting July 2014)

Ming-Wen An, Assistant Professor of Statistics, **Vassar College**

Emmanuel Kaparakis, Director, Quantitative Analysis Center and
Centers for Advanced Computing, **Wesleyan University**

3. *Workshop liaison*

Cassandra Pattanayak
Guthman Director of the Quantitative Analysis Institute
Wellesley College
106 Central Street
Wellesley, MA 02481
781-283-3435
cpattanayak@wellesley.edu

4. *Condensed CVs (attached separately)*

5. *Preliminary budget*

This budget is for a two-day workshop for 17 participants.

a. Stipend to be shared among organizers	\$1000
b. Travel and accommodations (17 participants at \$1000 total per person)	
Travel (17 at \$500 each)	\$8500
Lodging for 2 nights (17 at \$350 each)	\$5950
Meals (17 at \$150 each)	\$2550
c. Honorarium and meals for outside speaker (local)	\$1000
d. Other expenses	
Administrative expenses	Wellesley will provide
Unanticipated expenses	\$1000
Total	\$20000

CASSANDRA WOLOS PATTANAYAK

Condensed Curriculum Vitae

Wellesley College
Clapp 238, 106 Central Street, Wellesley, MA 02482
cpattanayak@wellesley.edu, ph: 781-283-3435

- EMPLOYMENT**
- Jack and Sandra Polk Guthman '65 Director, Quantitative Analysis Institute,** Wellesley College, October 2013-present.
 - Founding Director, Quantitative Analysis Institute,** Wellesley College, June 2013-October 2013.
 - Visiting Lecturer,** Wellesley College, June 2013-present.
 - College Fellow in Statistics,** Harvard University Statistics Department, July 2011-May 2013.
- EDUCATION**
- Ph.D. in Statistics, Harvard University,** November 2011.
 - A.M. in Statistics, Harvard University,** June 2008.
 - A.B. in Statistics, Harvard University, magna cum laude,** June 2006.
- HONORS**
- Certificate of Teaching Excellence for Statistics 98, Harvard University, 2012, 2013.
 - Certificate of Teaching Excellence for Economics 1127, Harvard University, 2012, 2013.
 - Best Reviewers of 2010, *Pharmacoepidemiology and Drug Safety*, 2011.
 - Winner, Promoting the Practice and Profession of Statistics Video Competition, American Statistical Association, 2011. With Paul T. Edlefsen and Xiao-Li Meng. One of three selected videos. Featured in *Significance Magazine*.
 - Certificate of Distinction in Teaching for Economics 1127, Harvard University, 2010, 2011.
 - Selected for Designing the Course of the Future, Derek Bok Center for Teaching and Learning, Harvard University, 2011.
 - David K. Pickard Award for Teaching Fellows, Harvard University Statistics Department, 2010.
 - Certificate of Distinction in Teaching for Statistics 245, Harvard University, 2010.
 - Certificate of Distinction in Teaching for Statistics 305, Harvard University, 2009.
 - Mu Sigma Rho, National Statistics Honor Society, inducted 2009.
 - Post-Qualifying Talk Award, Harvard University Statistics Department, 2009.
 - AT&T Labs Fellowship, 2006-2009.
 - Harvard College Scholar, 2006.
 - Robert C. Byrd Scholarship, 2002-2006.
 - Founders' Scholarship, Golub Foundation, 2002-2006.
- COURSES DESIGNED OR TAUGHT**
- Quantitative Analysis Institute Summer Program. Summer 2014, 2013, Wellesley College. (Designed and taught).
 - Economics 309/Quantitative Reasoning 309/Sociology 30: Causal Inference. Spring 2014, Wellesley College. (Designed and taught).
 - Quantitative Reasoning 170: Causation (First Year Seminar). Fall 2013, Wellesley College. (Designed and taught).
 - Economics 1127: Statistical Methods for Evaluating Causal Effects. (Cross-listed in statistics department). Spring 2013, 2012, Harvard University. (Taught).
 - Statistics 98: Junior Tutorial in Statistics. Spring 2013, 2012, Harvard University. (Designed and taught, with David P. Harrington).
 - Statistics 139: Statistical Sleuthing through Linear Models. Fall 2012, 2011, Harvard University. (Taught).
 - Empirical Reasoning 16/S-16: Real-Life Statistics. Fall 2012, 2011, Summer 2011, Harvard University. (Designed and taught, with Xiao-Li Meng).
 - Statistics 305: Statistical Fallacies and Paradoxes: A Cartoon Guide (Graduate Seminar in General Education). Fall 2009, Harvard University. (Designed).

SELECTED
PUBLICATIONS

- Pattanayak, C. W.** (Forthcoming). "Evaluating covariate balance." In W. Pan and H. Bai (Eds.), *Propensity Score Analysis: Fundamentals, Developments, and Extensions*. New York: Guilford Press.
- Turpie, A. G. G., Haas, S., Kreutz, R., Mantovani, L. G., **Pattanayak, C. W.**, Holberg, G., Jamal, W., Schmidt, A., van Eickels, M., & Lassen, M. R. (2014). "A non-interventional comparison of rivaroxaban with standard of care for thromboprophylaxis after major orthopaedic surgery in 17,701 patients with propensity score adjustment." *Thrombosis and Haemostasis*, *111*(1), 94-102.
- Greiner, D. J., **Pattanayak, C. W.**, & Hennessy, J. (2013). "The limits of unbundled legal assistance: A randomized study in a Massachusetts District Court and prospects for the future." *Harvard Law Review*, *126*(4), 901-989.
- Greiner, D. J., & **Pattanayak, C. W.** (2012). "Randomized evaluation in legal assistance: What difference does representation (offer and actual use) make?" *Yale Law Journal*, *121*(8), 2118-2214.
- Pattanayak, C. W.**, Rubin, D. B., & Zell, E. R. (2011). "Propensity score methods for creating covariate balance in observational studies." *Rev. Esp. Cardiol.*, *64*(10), 897-903.

INVITED
PRESENTATIONS
SINCE 2011

- "Reaching students outside the statistics classroom: Integrating statistics into the coursework and research experiences of non-statistics majors" and "Panel on Bridging the Disciplines." Electronic Conference on Teaching Statistics, Consortium for the Advancement of Undergraduate Statistics Education, May 2014.
- "Causal inference in a partially randomized preference trial with small samples." New England Statistics Symposium, Boston, MA, April 2014.
- "A potential outcomes, and typically more powerful, alternative to 'Cochran-Mantel-Haenszel.'" Department of Statistics, University of Connecticut, April 2014.
Atlantic Causal Inference Conference, Cambridge, MA, May 2013.
Applied Statistics Workshop, Harvard University Institute of Quantitative Social Science, Cambridge, MA, November 2012.
- "Inferring causation without randomization: A matched design to assess the number of embryos to transfer during in vitro fertilization." Department of Mathematics, Union College, Schenectady, NY, October 2013.
Department of Mathematics, Amherst College, Amherst, MA, September 2013.
- "Getting it right in comparative effectiveness research: Design matters!" Health Policy session at Joint Statistical Meetings, Montreal, QC, August 2013.
- "A powerful, potential outcomes method for estimating any estimand across multiple groups." Society for Research on Educational Effectiveness Spring Conference, Washington, DC, March 2013. Poster.
- "Propensity score matching to assess the effect of single versus double embryo transfer for in vitro fertilization." Center for Clinical and Translational Science and Training Grand Rounds, Cincinnati Children's Hospital, Cincinnati, OH, March 2012.
- "The limits of unbundled legal assistance: A randomized study in a Massachusetts District Court and prospects for the future." Yale Law Journal, New Haven, CT, March 2012.
- "How effective are limited legal assistance programs? A randomized experiment in housing." Conference on Empirical Legal Studies, Chicago, IL, November 2011.
- "What difference representation?" Applied Statistics Workshop, Harvard University Institute of Quantitative Social Science, Cambridge, MA, February 2011.

CORRINE HANSEN TAYLOR

Quantitative Reasoning Program
Wellesley College
106 Central Street
Wellesley, MA 02481-8203

(781) 283-2152
ctaylor1@wellesley.edu

PROFESSIONAL PREPARATION

College of William and Mary in Virginia, Economics, A.B., 1988
University of Wisconsin – Madison, Economics, M.S., 1993
University of Wisconsin – Madison, Economics, Ph.D., 1998

APPOINTMENTS

Director, Quantitative Reasoning Program, Wellesley College, 2001-

Assistant Professor, Department of Economics, Wellesley College, 1998-2001

Research Assistant to W. Lee Hansen, Department of Economics,
University of Wisconsin – Madison, 1993-1997

Instructor, Department of Economics, Upper Iowa University, Madison, WI, 1996

Project Assistant, Teacher Compensation Project, Principal Investigator Allan Odden,
Consortium for Policy Research in Education, UW – Madison, 1994-1995

Teaching Assistant, Department of Economics, UW – Madison, 1992-1994

Senior Analyst, Strategic Planning Department, MetLife Auto and Home
(Metropolitan Property and Casualty Insurance Company), Warwick, RI, 1998-1991

PUBLICATIONS

Taylor, C. (2012). Quantitative reasoning and sustainability. *Numeracy*: Vol 5: Iss. 2, Article 1. <http://dx.doi.org/10.5038/1936-4660.5.2.1>

Johnson, D.K.N., Lybecker, K.M., & Taylor, C.H. (2011). Semester, trimester, or block plan? Retention of economic principles by undergraduates on alternative curricular structures. *Journal of Education for Business*, 86(6): 332-338.

Butcher, K.F., McEwan P.J., & Taylor, C.H. (2010). The effects of quantitative skills training on college outcomes and peers. *Economics of Education Review*, Vol. 29: Iss. 2, pp. 187-199.

Taylor, C. (2009) Editorial: Assessing quantitative reasoning. *Numeracy*: Vol. 2: Iss. 2, Article 1. <http://services.bepress.com/numeracy/vol2/iss2/art1/>

Taylor, C. (2008). Preparing students for the business of the real (and highly quantitative) world. In B.L. Madison and L.A. Steen (Eds.), *Calculation vs. context: Quantitative literacy and its implications for teacher education* (pp. 109-124). Mathematics Association of America.

Campbell, J., Taylor, C., & Orr P. (2007). Preparing pathways to information literacy: Combining research, technology, and core college competencies to select first-year students. In L.L. Hardesty (Ed.), *The role of the library in the first college year* (Monograph No. 45, pp. 249-253). Columbia, SC: University of South Carolina, National Resource Center for the First-Year Experience and Students in Transition.

Taylor, C. (2006). Quantitative reasoning at Wellesley College. In R. Gillman (Ed.), *Current practices in quantitative literacy* (pp. 141-146). Mathematics Association of America.

SYNERGISTIC ACTIVITIES

NSF TUES Grant Co-PI with Ben Galluzzo (Shippensburg U). "USE Math: Undergraduate Sustainability Experiences in Mathemstics. Award #1245937.

Disciplinary Society Representative, representing the National Numeracy Network to the FIPSE-funded initiative "SISL: Sustainability Increases Student Learning," 2011-

Past-President, National Numeracy Network

Member 2004-; Board of Directors 2005-; President 2007-2011

Main contributor to the NNN Web site

Associate Editor, *Numeracy: Advancing Education in Quantitative Literacy*, 2007-
Development team member for the new on-line journal, 2006-2007

Meta-rubrics team member, QR rubrics, Valid Assessment of Learning in Undergraduate Education (VALUE), Association of American Colleges and Universities, 2008-2009

Outside reviewer of other colleges' QR programs and initiatives:
Carleton College, Trinity College, Bowdoin College

Consultant at dozens of colleges establishing QR programs over last 10 years

COURSES TAUGHT AT WELLESLEY COLLEGE

Principles of Microeconomics (ECON 101)

Education, Welfare and Taxes (ECON 226)

Public Economics (ECON 310)

Education Policy Analysis (ECON 326)

Social Science Data Analysis (QR 199/ECON 103)

Introduction to Quantitative Reasoning (QR 140)

Statistical Analysis of Education Issues (QR 180)

Statistics in the Biological Sciences (BISC 198), with Jeff Hughes (Biology)

Learning and Teaching Mathematics (EDUC 314), with Bernice Speiser and Heather Haskell (Education) and Jessica Polito (Quantitative Reasoning)

ABRIDGED CURRICULUM VITÆ

Nicholas Jon Horton
Professor of Statistics, Department of Mathematics and Statistics
Amherst College, Amherst, MA 01002

6/2014

Education

1999 Sc.D. Biostatistics, Harvard School of Public Health
1987 A.B. Psychology, Harvard University, Honors

Program and Curricular Development Related to Teaching/Learning

Chair-elect, American Statistical Association Section on Statistical Education (2015)
Chair, American Statistical Association Professional Initiative Workgroup to Revise Guidelines for Undergraduate Academic Programs in Statistics (2013-2014)
Workshop co-instructor, "Teaching Statistics using R", Electronic Conference on Teaching Statistics (5/14) also at US Conference on Teaching Statistics, Raleigh, NC (5/13 and 5/11, "Start Teaching with R", "A Compendium of Commands to Teach Statistics with R" and "Start Modeling with R" downloadable from <http://uscots2013.mosaic-web.org/Workshop/>)
Speaker (with Jenny Spohrer, Bryn Mawr College), "Blending online learning and liberal-arts classroom interaction", Teaching Arts Lunch, Smith College (11/12 and 10/11)
Developer of teaching materials (7/13, "Introduction to the Practice of Statistics using R" and "The Statistical Sleuth in R", accessible from <http://www.amherst.edu/~nhorton/ips6e>)
Building a community around modeling, statistics, computation, and calculus (Co-PI of NSF Funded Project MOSAIC, Kaplan PI, 2010-2014)

Publications Related to Teaching/Learning

Baumer BS, Cetinkaya-Rundel M, Bray A, Loi L (SC '13 undergraduate co-author) and Horton NJ. R Markdown: integrating a reproducible analysis tool into introductory statistics, *Technology Innovations in Statistics Education*, 2014; 8(1),
Stoudt S (SC '15 graduate first author), Cao Y (SC '14 graduate co-author, Udwin D (SC '14 graduate first author) and Horton NJ. What percent of the continental US is within one mile of a road?, *Statistics Education Web*, 2014,
Wild CJ, Pfannkuch M, Regan M and Horton NJ. Towards more accessible conceptions of statistical inference, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2011; 174(part 2):247-295 (the paper was read before the RSS on October 20, 2010, and the 39 written comments plus our rejoinder were published).
Horton NJ and Kleinman KP. *Using R for data management, statistical analysis and graphics (second edition)*, Chapman and Hall/CRC Press, 2015.

- Konold C, Madden S, Pollatsek A, Pfannkuch M, Wild C, Ziedins I, Finzer W, Horton NJ and Kazak S. Conceptual challenges in coordinating theoretical and data-centered estimates of probability, *Mathematical Thinking and Learning*, 2011; 13:68-86.
- Pfannkuch M, Regan M, Wild C and Horton NJ. Telling data stories: essential dialogues for comparative reasoning, *Journal of Statistics Education*, 2010; 18(1) (won award for best paper in JSE in 2010).
- Horton NJ. Clinician attitudes towards biostatistics (research letter). *Mayo Clinic Proceedings*, 2007; 82(12):1578.
- Switzer SS (SC '06 undergraduate first-author) and Horton NJ. What your doctor should know about statistics (but perhaps doesn't), *Chance*, 2007; 20(1):17-21.
- Horton NJ and Switzer SS (SC '06 undergraduate co-author). Statistical methods in the Journal (research letter). *New England Journal of Medicine*, 2005; 353(18):1977-1979.
- Horton NJ, Brown ER and Qian L (SC '05 undergraduate co-author). Use of R as a toolbox for mathematical statistics exploration. *The American Statistician*, 2004; 58(4):343-357.

Amy S. Wagaman

Abbreviated Curriculum Vitae

306 Seeley Mudd
Department of Mathematics and Statistics
Amherst College, Campus Box 2239
P.O. Box 5000
Amherst, MA 01002-5000
Office: 413-542-2423

1B Merrill House
4 Merrill Science Road
Amherst, MA 01002
Mobile: 413-531-1030

awagaman@amherst.edu

Current Position: Assistant Professor of Statistics, Amherst College
(formerly Assistant Professor of Mathematics (Statistics), Amherst College)

Education

- August 2008 Ph.D. in Statistics, University of Michigan, Ann Arbor, MI
Dissertation: Topics in High-Dimensional Inference with Applications to Raman Spectroscopy
Chair: Prof. Liza Levina, Statistics, elevina@umich.edu
- May 2005 M.S. in Statistics, University of Michigan, Ann Arbor, MI
- May 2003 B.A. in Mathematics and Anthropology, Kenyon College, Gambier, OH
High Honors in Mathematics, Summa Cum Laude, Salutatorian

Publications and Presentations Related to Teaching/Learning

Papers:

Wagaman, A.S. "Meeting Student Needs for Multivariate Data Analysis: A Case Study in Teaching a Multivariate Data Analysis Course with No Pre-requisites." Paper is arXiv-ed (ar-Xiv: 1310.7141, 2013 [stat.OT]) and is being cited by other statistics educators.

Posters:

- May 2014 *Electronic Conference on Teaching Statistics (eCOTS)*
"Visualization and Data Science with Big Data in a Multivariate Data Analysis Elective"
Poster will be available online to those participating in the conference.
- May 2013 *United States Conference on Teaching Statistics (USCOTS)*
Embassy Suites, Raleigh, NC hosted by NC State and SAS
"An Alternative Elective: Multivariate Data Analysis for Undergraduates"
- May 2007 *United States Conference on Teaching Statistics (USCOTS)*
Ohio State University, Columbus, OH
"Getting Your Students to *Click* With Statistics."

Talks/Organized Discussions:

- August 2012 *Roundtable at JSM, San Diego, CA*
"Integrating Statistics with the Sciences"

- August 2011 *Roundtable at JSM, Miami, FL*
“Using R in Introductory Statistics Courses”
- April 2011 *ASA K-12 Statistics Education Webinar (April 5) - Invited*
“An Introduction to R with RCommander for Introductory Statistics”
- June 2010 *Talk to AP Stats Teachers about using R in Intro Stats*
Loomis Chaffee School, Windsor CT

Reviews/ Editing

Associate Editor for JSE - To start Summer 2014

Other Professional Activities

- JSM 2013 Joint Statistical Meetings (JSM) - Section on Statistical Education (Stat Ed) JSM Program Chair
- 2010/2011 Attendee/participant in “Strengthening Bridges Between Statistics and the Natural Sciences” AALAC Workshop in Jan. 2011

Benjamin Strong Baumer

CONTACT INFORMATION

48 Lexington Ave. #2
Florence, MA 01062 USA

Phone: 413.218.3900 (C), 413.585.3440 (O)
Email: bbaumer@smith.edu

RESEARCH INTERESTS

data science, network science, applied statistics, sabermetrics, sports analytics, statistical modeling, analysis of algorithms, combinatorial optimization, data visualization, graph theory, combinatorics

EDUCATION

The Graduate Center of the City University of New York, New York, NY

Ph.D., Mathematics, May 2012

- Dissertation: “SENSOR STRIP COVER: Maximizing Network Lifetime on an Interval”
- Committee: Amotz Bar-Noy (chair), Joseph Malkevitch, Christina Zamfirescu

M.Phil., M.A., Mathematics, February 2011

University of California, San Diego, La Jolla, CA

M.A., Mathematics (Applied), March 2003

Wesleyan University, Middletown, CT

B.A., Economics, May 2000

Accredited Professional StatisticianTM, American Statistical Association

ACADEMIC EXPERIENCE

Smith College, Northampton, MA

Visiting Assistant Professor, Mathematics & Statistics

July 2012 - June 2018

Director, Statistical & Data Sciences Program

July 2014 - June 2018

Jointly appointed to the Smith College Statistics Program, the Five College Statistics Program, and the Spinelli Center for Quantitative Learning. Courses taught include:

- Data Science (MTH292: F13)
- Statistics: Introduction to Regression Analysis (MTH247: F12)
- Introduction to Probability and Statistics (MTH241/MTH220: S13, S14)
- Introduction to Discrete Mathematics (MTH153: S13, F13)

PROFESSIONAL EXPERIENCE

New York Mets, Flushing, NY

Statistical Analyst, Baseball Operations

January 2004 - July 2012

Sole statistical analyst for Baseball Operations. Reported directly to the General Manager and Assistant General Manager.

- Sabermetrics
 - ◇ Advised the General Manager on player acquisitions and trades
 - ◇ Conducted player evaluations and forecasts using statistical models and methodologies
 - ◇ Evaluated current-edge sabermetric research
 - ◇ Produced advance scouting reports for the uniformed coaching staff
 - ◇ Consulted with ownership on playing field dimensions
- Technology
 - ◇ Creator, architect, and lead developer for proprietary baseball information system
 - ◇ Administered and developed departmental LAMP servers
 - ◇ Managed relationships with third-party data providers

HONORS AND AWARDS

- Teaching
 - ◇ Mary P. Dolciani Project NExT Fellow, 2012-2013

- Research
 - ◊ Maven and Contributor
Project MOSAIC, 2013-present
 - ◊ Co-PI, Modeling and Design of Multi-genre Networks using Composite Graph Structures
Network Science Collaborative Technology Alliance (NS-CTA), December 2012 - present

BOOKS

- [1] *The Sabermetric Revolution: Assessing the Growth of Analytics in Baseball* (with A. Zimbalist)
University of Pennsylvania Press, Philadelphia, February 2014.
- [2] *Ten Major Questions Facing Sports Analysts* (with J. Piette and B. Macdonald)
CRC Press, under contract, expected release spring 2015.

BOOK CHAPTERS

- [1] “Social-Communication Composite Networks” (with P. Basu, A. Bar-Noy, C. Chau)
In *Opportunistic Mobile Social Networks*, accepted pending minor revisions.

PEER-REVIEWED PUBLICATIONS

- [1] “Teaching Precursors of Data Science In Introductory and Second Courses in Statistics,” (with N. Horton, H. Wickham)
ICOTS '14, accepted pending minor revisions.
- [2] “Average Case Network Lifetime on an Interval with Adjustable Sensing Ranges” (with A. Bar-Noy)
Algorithmica, preview online.
- [3] “Quantifying Market Inefficiencies in the Baseball Players’ Market” (with A. Zimbalist)
Eastern Economic Journal, special issue on sports economics, preview online.
- [4] “R Markdown: Integrating A Reproducible Analysis Tool into Introductory Statistics” (with A. Bray, M. Çetinkaya-Rundel, N. Horton, L. Loi)
Technology Innovations in Statistics Education, 8(1), 2014.
- [5] “As Strong as the Weakest Link: Mining Diverse Cliques in Weighted Graphs” (with P. Bogdanov, P. Basu, A. Singh, A. Bar-Noy)
In *ECML/PKDD '13*, pp. 525-540, 2013.
- [6] “Brief Announcement: Set It and Forget It: Tighter Approximation Bounds for ROUNDROBIN in a Restricted Lifetime Model” (with A. Bar-Noy and D. Rawitz)
In *SPAA '13*, pp. 105-107, 2013.
- [7] “Parsing the Relationship between Baserunning and Batting Abilities within Lineups” (with J. Piette and B. Null)
Journal of Quantitative Analysis in Sports: 8(2):8, 2012.
- [8] “Changing of the Guards: Strip Cover with Duty Cycling” (with A. Bar-Noy and D. Rawitz)
In *SIROCCO '12*, pp. 36-47, 2012.
- [9] “Maximizing Network Lifetime on the Line with Adjustable Sensing Ranges” (with A. Bar-Noy)
In *ALGOSENSORS '11*, pp. 28-41, 2012.
- [10] “Modeling and Analysis of Composite Network Embeddings” (with P. Basu and A. Bar-Noy)
In *MSWiM '11*, pp. 341-350, 2011.
- [11] “Using Simulation to Estimate the Impact of Baserunning Ability in Baseball”
Journal of Quantitative Analysis in Sports: 5(2):8, 2009.
- [12] “Why On-Base Percentage is a Better Indicator of Future Performance than Batting Average: An Algebraic Proof”
Journal of Quantitative Analysis in Sports: 4(2):3, 2008.

CURRENT SUBMISSIONS

- [1] “Changing of the Guards: Strip Cover with Duty Cycling” (with A. Bar-Noy and D. Rawitz)
Selected for Special Issue of *Theoretical Computer Science*, under review.
- [2] “Set It and Forget It: Tighter Approximation Bounds for ROUNDROBIN in a Restricted Lifetime Model” (with A. Bar-Noy and D. Rawitz)
Algorithmica, under review.
- [3] “OpenWAR: An Open Source System for Overall Player Performance in Major League Baseball” (with S. Jensen and G. Matthews)
Journal of the American Statistical Association, under review.

EVENTS ORGANIZED • “Five College DataFest,” (with A. Bray), *University of Massachusetts*, Mar 28-30 2014.

ABRIDGED CURRICULUM VITAE

Ming-Wen An

Assistant Professor of Statistics, Department of Mathematics | Vassar College, Poughkeepsie NY

June 2, 2014

EDUCATION

2003-2008 Ph.D. Biostatistics | [Johns Hopkins Bloomberg School of Public Health](#)

1997-2001 B.A. in Mathematics, Magna Cum Laude | [Carleton College](#)

HONORS AND AWARDS

2011 Vassar College Committee on Research Award

2007 Helen Abbey Award for Excellence in Teaching

2007 Teaching Assistant Recognition Award (via student nominations)

2006 Nominated for Teaching Assistant Recognition Award

2001 Phi Beta Kappa Society at Carleton College

SERVICE

Vassar College

2013-present Majors Co-coordinator, Department of Mathematics

2013 Committee on Fellowships (ad hoc selection committee for Churchill and Goldwater Fellowships)

2012-present Premedical Advisory Committee

2011-present Subcommittee on Course Evaluation Questionnaires

2011-2013 Center for Collaborative Approaches to Science (CCAS) Committee

2011-2012 Committee on Fellowships

2010-present Life After VAssar (LAVA) Seminar Series Coordinatos, Department of Mathematics

2010-present Stats@Vassar Faculty Conversations Group (co-founder)

2009-2010 Center for Collaborative Approaches to Science (CCAS) Committee

2009-2010 Committee on Committees

2009-2010 Mathematics Colloquium Series Co-coordinator, Department of Mathematics

Professional Societies

2014-present Member, The American Statistical Association Committee on Funded Research

2013-2014 Statistics Education Chair, The Joint Statistical Meetings 2014 Program Committee

2012-2013 Statistics Education Chair-Elect, The Joint Statistical Meetings 2013 Program Committee

Editorial Reviewer

- **Statistics Journals:** Biometrics, International Chinese Statistical Association (ICSA) Symposium Book, SpringerPlus, Statistica Sinica, Statistics in Medicine
- **Statistical Education Journals:** Journal of Statistical Education, Mathematics Magazine
- **Clinical Journals:** Archives of Physical Medicine and Rehabilitation, British Journal of Cancer, Journal of Thoracic Oncology

- **Textbooks:** *Foundations and Applications of Statistics: An Introduction Using R*, R. Pruim; *Practicing Statistics: Guided Investigations for the Second Course*, S. Kuiper and J. Sklar.

ACTIVE RESEARCH GRANTS

- Validation of New Tumor Measurement-based Phase II Clinical Trial Endpoints. National Institutes of Health (NIH) R21 (CA-167326-01), PI: **M. An** and S. Mandrekar. Role: co-PI. Funding Amount: \$382,614. Funding Period: 04/01/2012 – 03/31/2014.
- Statistical Designs and Methods for Double-Sampling for HIV/AIDS. National Institutes of Health (NIH) R01 (AI-102710-01), PI: C. Frangakis. Role: co-Investigator. Funding Period: 01/15/2013 – 12/31/2016.

SELECTED PUBLICATIONS

1. **An MW**, Mandrekar SJ, Edelman MJ, Sargent DJ. Exploring the statistical and clinical impact of two interim analyses on the phase II design with option for direct assignment. *Contemporary Clinical Trials* 2014 Apr 24; 38(2):157-262. PMID 24768938.
2. Mandrekar SJ, **An MW**, Meyers J, Grothey A, Bogaerts J, Sargent DJ. Evaluation of Alternate Categorical Tumor Metrics and Cut Points for Response Categorization Using the RECIST 1.1 Data Warehouse. *Journal of Clinical Oncology* 2014 Mar 10; 32(8):841-850. PMID 24516033.
3. **An MW**, Frangakis CF, Yiannoutsis CT. Choosing profile double-sampling designs for survival estimation with application to President's Emergency Plan for AIDS Relief evaluation. *Statistics in Medicine* 2014 Jan 10. PMID: 24408038.
4. **An MW**, Reich NG, Crawford SO, Brookmeyer R, Louis TA, Nelson KE. A Stochastic Simulator of a Blood Product Donation Environment with Demand Spikes and Supply Shock. *PLoS ONE* 2011; 6(7):e21752. PMID: 21814550.
5. Joffe A, McNeely C, Colantuoni E, **An MW**, Wang W, Scharfstein DO. Evaluation of School-Based Smoking Cessation Interventions for Self-Described Adolescent Smokers. *Pediatrics* 2009 Aug. 124(2):187-94. PMID: 19651564.

SELECTED INVITED PRESENTATIONS

1. Department of Biostatistics, Institut Curie, Paris FRANCE, Mar 2014
Alternative Tumor Measurement-based Phase II Clinical Trial Endpoints for Predicting Overall Survival (OS), using the RECIST 1.1 data warehouse
2. Department of Biostatistics and Epidemiology, Gustave Roussy, Villejuif FRANCE, Mar 2014
A two-stage phase II design with direct assignment option in stage II
3. Department of Biostatistics and Bioinformatics, University of California San Diego, CA, Feb 2014
Alternative Tumor Measurement-based Phase II Clinical Trial Endpoints for Predicting Overall Survival (OS), using the RECIST 1.1 data warehouse
4. Department of Mathematical Sciences, United States Military Academy, NY, Dec 2011
Bottleneck in the Drug Pipeline? The Need for Alternate Endpoints in Phase II Cancer Clinical Trials
5. New Student Orientation, Vassar College, NY, Aug 2011
Bottleneck in the Drug Pipeline? The Need for Alternate Endpoints in Phase II Cancer Clinical Trials
6. Department of Mathematics, Carleton College, MN, Apr 2011
Bottleneck in the Drug Pipeline? The Need for Alternate Endpoints in Phase II Cancer Clinical Trials
7. Department of Mathematics, Bard College, NY, Jul 2010
The need for double-sampling designs in survival studies: An Evaluation of PEPFAR

Biographical Sketch: Emmanuel I. Kaparakis

Professional Preparation:

University of Connecticut	Economics	Ph. D. candidate, 1985-90 (all but dissertation)
University of Bridgeport	Economics	M.A. 1984
Athens Graduate School of Economics and Business (GR)	Economics	B.A. 1984

Appointments:

2008-present	Director of Centers for Advanced Computing, Wesleyan University
2006-present	Director, Quantitative Analysis Center, Wesleyan University
2006	Visiting Assistant Professor, Wesleyan University
2003-2007	Academic Computing Manager for the Social Sciences, Wesleyan University
2002-2003	Academic Technology Coordinator, Clark University
1995-1998	Faculty Liaison for Information Technology, Providence College
1990-2002	Instructor/Lecturer of Economics, Providence College

Related Publications:

1. Dierker, L., Kaparakis, E., Rose, J., Selya, A. and Beveridge, D. Strength in Numbers: A multidisciplinary, inquiry-based approach to introductory statistics education. *Journal of Effective Teaching*, 12(2), September 2012, pp 4-14

Other Significant Publications:

1. Kaparakis E., Miller S.M., and A.G. Noulas . Short-run Cost Inefficiency of Commercial Banks: A Flexible Stochastic Frontier Approach, *Journal of Money, Credit, and Banking*, 26(4), Nov. 1994, pp. 875-93.
2. Kaparakis E., Katsimbris G.M., and S.M. Miller Inflation and Relative Price Variability: The Case of Internationally Traded Primary Commodities, *Economics Letters*, 33(1), May 1990, pp. 47-53.

Synergistic Activities:

- Currently collaborating in the development and teaching of an inquiry based applied data analysis course. The project was originally sponsored by an NSF DUE CCLI Type 1 program grant (DUE 0942246) and has now received a Phase II grant.
- Working with the academic deans and faculty across the university on the development of Wesleyan's "Digital and Computational Knowledge Initiative" which aims at facilitating curriculum development that integrates modeling and computational work across the curriculum.
- Coordinator of the Scientific Computing and Informatics Center at Wesleyan University. The Center supports computational work and the use of the University's shared computing cluster.
- Technology Fellow for statistical analysis (2008-2010), National Institute for Technology in Liberal Education (NITLE)
- Principal Investigator for a 2007 NITLE grant, *Data Sharing Systems: Supporting Data Analysis across the Curriculum* (A collaboration of Grinnell, Kenyon, Oberlin Colleges, and Wesleyan University)