BDK01 – Biomedical Big Data Science

References

Adams, J and Klein, J (2011). Business Intelligence and Analytics in Health Care - A Primer. Washington, DC, The Advisory Board Company. http://www.advisory.com/Research/IT-Strategy-Council/Research-Notes/2011/Business-Intelligence-and-Analytics-in-Health-Care

Adida, B, Sanyal, A, et al. (2010). Indivo x: developing a fully substitutable personally controlled health record platform. *AMIA Annual Symposium Proceedings*, Washington, DC. American Medical Informatics Association. 6-10.

Aggarwal, CC and Zhai, C, Eds. (2012). Mining Text Data. New York, NY, Springer.

Alberts, B, Kirschner, MW, et al. (2014). Rescuing US biomedical research from its systemic flaws. *Proceedings of the National Academy of Sciences*. 111: 5773-5777.

Allen, N, Sudlow, C, et al. (2012). UK Biobank: current status and what it means for epidemiology. *Health Policy and Technology*. 1: 123-126.

Allen, NE, Sudlow, C, et al. (2014). UK Biobank data: come and get it. *Science Translational Medicine*. 6: 224ed224. http://stm.sciencemag.org/content/6/224/224ed4.short

Anonymous (2011). Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease. Washington, DC, National Academies Press.

Anonymous (2012). The value of analytics in healthcare - From insights to outcomes. Somers, NY, IBM Global Services. http://www-935.ibm.com/services/us/gbs/thoughtleadership/ibv-healthcare-analytics.html

Anonymous (2013). An assessment of demand for labour and skills, 2012-2017. London, England, SAS. http://www.e-skills.com/Documents/Research/General/BigDataAnalytics_Report_Jan2013.pdf

Anonymous (2013). National Workshop Report: Advancing the Use of Electronic Data in Patient-Centered Outcomes Research. Washington, DC, Patient-Centered Outcomes Research Institute. http://www.pcori.org/assets/2013/06/PCORI-National-Workshop-Advancing-Electronic-Data-PCOR-070312.pdf

Anonymous (2013). Report of Workshop on Enhancing Training for Biomedical Big Data, National Institutes of Health. http://bd2k.nih.gov/pdf/bd2k_training_workshop_report.pdf

Anonymous (2014). A Decade of Discovery: The NIH Roadmap and Common Fund. Bethesda, MD, National Institutes of Health. http://commonfund.nih.gov/sites/default/files/ADecadeofDiscoveryNIHRoadmapCF.pdf

Anonymous (2015). Big Data Now: 2014 Edition. Sebastopol, CA, O'Reilly Media.

Ashley, EA (2015). The Precision Medicine Initiative - a new national effort. *Journal of the American Medical Association*. 313: 2119-2120.

Bayley, KB, Belnap, T, et al. (2013). Challenges in using electronic health record data for CER: experience of four learning organizations. *Medical Care*. 51: S80-S86.

Bellazzi, R and Zupan, B (2008). Predictive data mining in clinical medicine: current issues and guidelines. *International Journal of Medical Informatics*. 77: 81-97.

Bhavnani, SK, Warden, M, et al. (2012). Researchers' needs for resource discovery and collaboration tools: a qualitative investigation of translational scientists. *Journal of Medical Internet Research*. 14: e75. http://www.jmir.org/2012/3/e75/

Bourgeois, FC, Olson, KL, et al. (2010). Patients treated at multiple acute health care facilities: quantifying information fragmentation. *Archives of Internal Medicine*. 170: 1989-1995.

Bronsert, MR, Henderson, WG, et al. (2013). Comparative effectiveness of antihypertensive therapeutic classes and treatment strategies in the initiation of therapy in primary care patients: a Distributed Ambulatory Research in Therapeutics Network (DARTNet) study. *Journal of the American Board of Family Medicine*. 26: 529-538.

Buneman, P and Davidson, SB (2010). Data provenance – the foundation of data quality. Pittsburgh, PA, Carnegie Mellon University Software Engineering Institute. http://www.sei.cmu.edu/measurement/research/upload/Davidson.pdf

Cascade, E, Marr, P, et al. (2012). Conducting research on the Internet: medical record data integration with patient-reported outcomes. *Journal of Medical Internet Research*. 14: e137. http://www.jmir.org/2012/5/e137/

Chute, CG, Pathak, J, et al. (2011). The SHARPn project on secondary use of Electronic Medical Record data: progress, plans, and possibilities. *AMIA Annual Symposium Proceedings 2011*, Washington, DC. 248-256.

Cimino, JJ, Ayres, EJ, et al. (2014). The National Institutes of Health's Biomedical Translational Research Information System (BTRIS): design, contents, functionality and experience to date. *Journal of Biomedical Informatics*. 52: 11-27.

Collins, FS, Hudson, KL, et al. (2014). PCORnet: turning a dream into reality. *Journal of the American Medical Informatics Association*. 21: 576-577.

Collins, FS and Varmus, H (2015). A new initiative on precision medicine. *New England Journal of Medicine*. 372: 793-795.

Coons, SJ, Eremenco, S, et al. (2014). Capturing patient-reported outcome (PRO) data electronically: the past, present, and promise of ePRO measurement in clinical trials. *Patient*: Epub ahead of print.

Crown, WH (2015). Potential application of machine learning in health outcomes research and some statistical cautions. *Value Health*. 18: 137-140.

Danciu, I, Cowan, JD, et al. (2014). Secondary use of clinical data: the Vanderbilt approach. *Journal of Biomedical Informatics*. 52: 28-35.

Daniels, RJ (2015). A generation at risk: young investigators and the future of the biomedical workforce. *Proceedings of the National Academy of Sciences*. 112: 313-318.

Davenport, TH and Harris, JG (2007). Competing on Analytics: The New Science of Winning. Cambridge, MA, Harvard Business School Press.

Davenport, TH and Patil, DJ (2012). Data Scientist: The Sexiest Job of the 21st Century. Harvard Business Review, October, 2012. http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/

deLusignan, S and vanWeel, C (2005). The use of routinely collected computer data for research in primary care: opportunities and challenges. *Family Practice*. 23: 253-263.

Dhar, V (2013). Data science and prediction. *Communications of the ACM*. 56(12): 64-73.

Dougherty, D and Conway, PH (2008). The "3T's" road map to transform US health care: the "how" of high-quality care. *Journal of the American Medical Association*. 299(2319-2321)

Embi, P (2013). Clinical research informatics: survey of recent advances and trends in a maturing field. IMIA Yearbook of Medical Informatics 2013. B. Séroussi, M. Jaulent and C. Lehmann. Stuttgart, Germany, Schattauer**:** 178-184.

Embi, PJ, Jain, A, et al. (2005). Effect of a clinical trial alert system on physician participation in trial recruitment. *Archives of Internal Medicine*. 165: 2272-2277.

Embi, PJ and Leonard, AC (2012). Evaluating alert fatigue over time to EHR-based clinical trial alerts: findings from a randomized controlled study. *Journal of the American Medical Informatics Association*. 19(e1): e145-e148.

Finnell, JT, Overhage, JM, et al. (2011). All health care is not local: an evaluation of the distribution of emergency department care delivered in Indiana. *AMIA Annual Symposium Proceedings*, Washington, DC. 409-416.

Flach, P (2012). Machine Learning: The Art and Science of Algorithms that Make Sense of Data. Cambridge, England, Cambridge University Press.

Fleurence, RL, Curtis, LH, et al. (2014). Launching PCORnet, a national patient-centered clinical research network. *Journal of the American Medical Informatics Association*. 21: 578-582.

Fleurence, RL, Forsythe, LP, et al. (2014). Patient-powered research networks aim to improve patient care and health research. *Health Affairs*. 33: 1212-1219.

Frankovich, J, Longhurst, CA, et al. (2011). Evidence-based medicine in the EMR era. *New England Journal of Medicine*. 365: 1758–1759.

Fraser, H, Jayadewa, C, et al. (2013). Analytics across the ecosystem - A prescription for optimizing healthcare outcomes. Somers, NY, IBM Global Services. http://www-935.ibm.com/services/us/gbs/thoughtleadership/healthcare-ecosystem/

Gardner, E (2013). The HIT Approach to Big Data. Health Data Management, March 1, 2013. http://www.healthdatamanagement.com/issues/21_3/The-HIT-Approach-to-Big-Data-Anayltics-45735-1.html

Gelijns, AC and Gabriel, SE (2012). Looking beyond translation--integrating clinical research with medical practice. *New England Journal of Medicine*. 366: 1659-1661.

Grande, D, Mitra, N, et al. (2013). Public preferences about secondary uses of electronic health information. *JAMA Internal Medicine*. 173: 1798-1806.

Grus, J (2015). Data Science From Scratch. Sebastopol, CA, O'Reilly Media.

Hamburg, MA and Collins, FS (2010). The path to personalized medicine. *New England Journal of Medicine*. 363: 301-304.

Harris, PA, Taylor, R, et al. (2009). Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*. 42: 377-381.

Hersh, WR (2014). Healthcare Data Analytics. Health Informatics: Practical Guide for Healthcare and Information Technology Professionals, Sixth Edition. R. Hoyt and A. Yoshihashi. Pensacola, FL, Lulu.com**:** 62-75.

Hillier, TA, Pedula, KL, et al. (2007). Childhood obesity and metabolic imprinting: the ongoing effects of maternal hyperglycemia. *Diabetes Care*. 30: 2287-2292.

Hornbrook, MC, Hart, G, et al. (2005). Building a virtual cancer research organization. *Journal of the National Cancer Institute Monographs*. 35: 12-25.

Hripcsak, G and Albers, DJ (2012). Next-generation phenotyping of electronic health records. *Journal of the American Medical Informatics Association*. 20: 117-121.

Huser, V and Cimino, JJ (2013). Desiderata for healthcare integrated data repositories based on architectural comparison of three public repositories. *AMIA Annual Symposium Proceedings*, Washington, DC. 648-656.

Kahn, MG, Raebel, MA, et al. (2012). A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. *Medical Care*. 50(Suppl): S21-S29.

Katzan, IL and Rudick, RA (2012). Time to integrate clinical and research informatics. *Science Translational Medicine*. 4: 162fs141. http://stm.sciencemag.org/content/4/162/162fs41.abstract

Kaushal, R, Hripcsak, G, et al. (2014). Changing the research landscape: the New York City Clinical Data Research Network. *Journal of the American Medical Informatics Association*. 21: 587-590.

Kellermann, AL and Jones, SS (2013). What will it take to achieve the as-yet-unfulfilled promises of health information technology? *Health Affairs*. 32: 63-68.

Kohane, IS, Churchill, SE, et al. (2012). A translational engine at the national scale: informatics for integrating biology and the bedside. *Journal of the American Medical Informatics Association*. 19: 181-185.

Köpcke, F and Prokosch, HU (2014). Employing computers for the recruitment into clinical trials: a comprehensive systematic review. *Journal of Medical Internet Research*. 16(7): e161. http://www.jmir.org/2014/7/e161/

Kosinski, M, Stillwell, D, et al. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*. 110: 5802-5805.

Kumar, A, Niu, F, et al. (2013). Hazy: making it easier to build and maintain big-data analytics. *Communications of the ACM*. 56(3): 40-49.

Lang, T (2011). Advancing global health research through digital technology and sharing data. *Science*. 331: 714-717.

Levy, S (2014). How the NSA Almost Killed the Internet. Wired, January 7, 2014. http://www.wired.com/threatlevel/2014/01/how-the-us-almost-killed-the-internet/

Lewis, M (2004). Moneyball: The Art of Winning an Unfair Game. New York, NY, W. W. Norton & Company.

MacKenzie, SL, Wyatt, MC, et al. (2012). Practices and perspectives on building integrated data repositories: results from a 2010 CTSA survey. *Journal of the American Medical Informatics Association*. 19(e1): e119-e124.

Macleod, MR, Michie, S, et al. (2014). Biomedical research: increasing value, reducing waste. *Lancet*. 383: 101-104.

Mandl, KD, Mandel, JC, et al. (2012). The SMART Platform: early experience enabling substitutable applications for electronic health records. *Journal of the American Medical Informatics Association*. 19: 597-603.

Mandl, KD, Simons, WW, et al. (2007). Indivo: a personally controlled health record for health information exchange and communication. *BMC Medical Informatics & Decision Making*. 7: 25. http://www.biomedcentral.com/1472-6947/7/25

Manolio, TA, Weis, BK, et al. (2012). New models for large prospective studies: is there a better way? *American Journal of Epidemiology*. 175: 859-866.

Manor-Shulman, O, Beyene, J, et al. (2008). Quantifying the volume of documented clinical information in critical illness. *Journal of Critical Care*. 23: 245-250.

Manyika, J, Chui, M, et al. (2011). Big data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute. http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Margolis, R, Derr, L, et al. (2014). The National Institutes of Health's Big Data to Knowledge (BD2K) initiative: capitalizing on biomedical big data. *Journal of the American Medical Informatics Association*. 21: 957-958.

Marsolo, K (2013). Informatics and operations--let's get integrated. *Journal of the American Medical Informatics Association*. 20: 122-124.

McGlynn, EA, Lieu, TA, et al. (2014). Developing a data infrastructure for a learning health system: the PORTAL network. *Journal of the American Medical Informatics Association*. 21: 596-601.

McMurry, AJ, Murphy, SN, et al. (2013). SHRINE: enabling nationally scalable multi-site disease studies. *PLoS ONE*. 8(3): e55811. http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0055811

Mielczarek, EV and Engler, BD (2014). Selling Pseudoscience: A Rent in the Fabric of American Medicine: A Study of Federal Funding Advancing Naturopathy, Acupuncture, Chiropractic, and Energy Healing as Acceptable Medical Protocols Finds Troubling Misuse of Taxpayer Dollars. Skeptical Inquirer, May/June, 2014

Murphy, SN, Dubey, A, et al. (2012). Current state of information technologies for the clinical research enterprise across academic medical centers. *Clinical and Translational Science*. 5: 281-284.

Murphy, SN, Weber, G, et al. (2011). Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *Journal of the American Medical Informatics Association*. 17: 124-130.

Nass, SJ, Levit, LA, et al., Eds. (2009). Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research. Washington, DC, National Academies Press.

O'Reilly, T, Loukides, M, et al. (2012). How Data Science Is Transforming Health Care - Solving the Wanamaker Dilemma. Sebastopol, CA, O'Reilly Media.

Ohno-Machado, L, Agha, Z, et al. (2014). pSCANNER: patient-centered Scalable National Network for Effectiveness Research. *Journal of the American Medical Informatics Association*. 21: 621-626.

Olson, JE, Ryu, E, et al. (2013). The Mayo Clinic Biobank: a building block for individualized medicine. *Mayo Clinic Proceedings*. 88: 952-962.

Pace, WD, Fox, CH, et al. (2015). The DARTNet Institute: seeking a sustainable support mechanism for electronic data enabled research networks. *eGEMS*. 2(2): 1063. http://repository.academyhealth.org/egems/vol2/iss2/6

Payne, PR, Johnson, SB, et al. (2005). Breaking the translational barriers: the value of integrating biomedical informatics and translational research. *Journal of Investigative Medicine*. 53: 192-200.

Payne, PR, Pressler, TR, et al. (2013). People, organizational, and leadership factors impacting informatics support for clinical and translational research. *BMC Medical Informatics & Decision Making*. 13: 20. http://www.biomedcentral.com/1472-6947/13/20

Platt, R, Carnahan, RM, et al. (2012). The U.S. Food and Drug Administration's Mini-Sentinel program: status and direction. *Pharmacoepidemiology and Drug Safety*. 21(Suppl 1): 1-8.

Prokosch, HU, Beck, A, et al. (2010). IT infrastructure components for biobanking. *Applied Clinical Informatics*. 1: 419-429.

Raebel, MA, Haynes, K, et al. (2014). Electronic clinical laboratory test results data tables: lessons from Mini-Sentinel. *Pharmacoepidemiology and Drug Safety*. 23: 609-618.

Rajput, ZA, Mbugua, S, et al. (2012). Evaluation of an Android-based mHealth system for population surveillance in developing countries. *Journal of the American Medical Informatics Association*. 19: 655-659.

Rea, S, Pathak, J, et al. (2012). Building a robust, scalable and standards-driven infrastructure for secondary use of EHR data: The SHARPn project. *Journal of Biomedical Informatics*. 45: 763-771.

Richesson, RL and Andrews, JE, Eds. (2012). Clinical Research Informatics. New York, NY, Springer.

Rogith, D, Yusuf, RA, et al. (2014). Attitudes regarding privacy of genomic information in personalized cancer therapy. *Journal of the American Medical Informatics Association*. 21(e2): e320-e325.

Rosenbloom, ST, Denny, JC, et al. (2011). Data from clinical notes: a perspective on the tension between structure and flexible documentation. *Journal of the American Medical Informatics Association*. 18: 181-186.

Ross, J, Tu, S, et al. (2010). Analysis of eligibility criteria complexity in clinical trials. *AMIA Summits on Translational Science Proceedings*, San Francisco, CA. 46-50.

Ruiz, EJ, Hristidis, V, et al. (2012). Correlating financial tme series with micro-blogging activity. *Fifth ACM International Conference on Web Search & Data Mining*, Seattle, WA http://www.cs.ucr.edu/~vagelis/publications/wsdm2012-microblog-financial.pdf

Salant, JD and Curtis, L (2012). Nate Silver-Led Statistics Men Crush Pundits in Election. Bloomberg, November 7, 2012. http://www.bloomberg.com/news/2012-11-07/nate-silver-led-statistics-men-crush-pundits-in-election.html

Savova, GK, Masanz, JJ, et al. (2010). Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *Journal of the American Medical Informatics Association*. 17: 507-513.

Scherer, M (2012). Inside the Secret World of the Data Crunchers Who Helped Obama Win. Time, November 7, 2012. http://swampland.time.com/2012/11/07/inside-the-secret-world-of-quants-and-data-crunchers-who-helped-obama-win/

Slomski, A (2011). The National Library of Medicine: 175 years of advancing biomedical knowledge. *Journal of the American Medical Association*. 305: 2158-2161.

Smith, M (2014). Targeted: How Technology Is Revolutionizing Advertising and the Way Companies Reach Consumers. Washington, DC, AMACOM.

Tannen, RL, Weiner, MG, et al. (2008). Replicated studies of two randomized trials of angiotensin-converting enzyme inhibitors: further empiric validation of the 'prior event rate ratio' to adjust for unmeasured confounding by indication. *Pharmacoepidemiology and Drug Safety*. 17: 671-685.

Tannen, RL, Weiner, MG, et al. (2009). Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings. *British Medical Journal*. 338: b81.
http://www.bmj.com/cgi/content/full/338/jan27_1/b81

Tannen, RL, Weiner, MG, et al. (2007). A simulation using data from a primary care practice database closely replicated the Women's Health Initiative trial. *Journal of Clinical Epidemiology*. 60: 686-695.

Tao, C, Jiang, G, et al. (2013). A semantic-web oriented representation of the clinical element model for secondary use of electronic health records data. *Journal of the American Medical Informatics Association*. 20: 554-562.

Vasilevsky, N, Johnson, T, et al. (2012). Research resources: curating the new eagle-i discovery system. *Database*. 2012: bar067.

Weber, GM, Murphy, SN, et al. (2009). The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories. *Journal of the American Medical Informatics Association*. 16: 624-630.

Weiskopf, NG, Hripcsak, G, et al. (2013). Defining and measuring completeness of electronic health records for secondary use. *Journal of Biomedical Informatics*. 46: 830-836.

Wilcox, AB, Gallagher, KD, et al. (2012). Research data collection methods: from paper to tablet computers. *Medical Care*. 7(Suppl 1): S68-S73.

Winslow, RL, Trayanova, N, et al. (2012). Computational medicine: translating models to clinical care. *Science Translational Medicine*. 4: 158rv111.
http://stm.sciencemag.org/content/4/158/158rv11.short

Yarborough, M and Sharp, RR (2002). Restoring and preserving trust in biomedical research. *Academic Medicine*. 77: 8-14.

Yin, HL, Gabrilove, J, et al. (2015). Sustaining the clinical and translational research workforce: training and empowering the next generation of investigators. *Academic Medicine*: Epub ahead of print.

Zaki, MJ and Meira, W (2014). Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge, England, Cambridge University Press.

Zerhouni, EA (2007). Translational research: moving discovery to practice. *Clinical Pharmacology and Therapeutics*. 81: 126-128.

Zielhuis, GA (2012). Biobanking for epidemiology. *Public Health*. 126: 214-216.

Zikopoulos, P, Eaton, C, et al. (2011). Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. New York, NY, McGraw-Hill.