CONTENTS

HIGHLIGHTS
Clinical research informatics: a growing subspecialization of biomedical informatics
Lucila Ohno-Machado ......................................................... 223

CASE REPORT
Assessing the quality of administrative data for research: a framework from the Manitoba Centre for Health Policy
Mark Smith, Lisa M Lix, Mahmoud Azimaee, Jennifer E Enns, Justine Orr, Say Hong, and Leslie L Roos ......................................................... 224

RESEARCH AND APPLICATIONS
Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record
Jason Walonoski, Mark Kramer, Joseph Nichols, Andre Quina, Chris Moesel, Dylan Hall, Carlton Duffett, Kudakwashe Dube, Thomas Gallagher, and Scott McLachlan .............................................................. 230

The representativeness of eligible patients in type 2 diabetes trials: a case study using GIST 2.0
Anando Sen, Andrew Goldstein, Shreya Chakrabarti, Ning Shang, Tian Kang, Anil Yaman, Patrick B Ryan, and Chunhua Weng ................................... 239

A conceptual framework for evaluating data suitability for observational studies
Ning Shang, Chunhua Weng, and George Hripcsak ......................................................... 248

A computable pathology report for precision medicine: extending an observables ontology unifying SNOMED CT and LOINC
Walter S Campbell, Daniel Karlsson, Daniel J Vreeman, Audrey J Lazenby, Geoffrey A Talmon, and James R Campbell ........................................ 259

The medical science DMZ: a network design pattern for data-intensive medical science

Uncovering exposures responsible for birth season – disease effects: a global study
Mary Regina Boland, Pradipta Parhi, Li Li, Riccardo Miotto, Robert Carroll, Usman Iqbal, Phung-Anh (Alex) Nguyen, Martijn Schuemie, Seng Chan You, Donahue Smith, Sean Mooney, Patrick Ryan, Yu-Chuan (Jack) Li, Rae Woong Park, Josh Denny, Joel T Dudley, George Hripcsak, Pierre Gentine, and Nicholas P Tatonetti ........................................ 275

DataMed – an open source discovery index for finding biomedical datasets

Decentralized and reproducible geocoding and characterization of community and environmental exposures for multisite studies
Cole Brokamp, Chris Wolfe, Todd Lingren, John Harley, and Patrick Ryan ................................. 309

Assessing privacy risks in population health publications using a checklist-based approach
Christine M O’Keefe, Adrien Ickowicz, Tim Churches, Mark Westcott, Maree O’Sullivan, and Atikur Khan ......................................................... 315
Hierarchical attention networks for information extraction from cancer pathology reports
Shang Gao, Michael T Young, John X Qiu, Hong-Jun Yoon, James B Christian, Paul A Fearn,
Georgia D Tourassi, and Arvind Ramanthan ................................................................. 321

User needs analysis and usability assessment of DataMed – a biomedical data discovery index
Ram Dixit, Deevakar Rogith, Vidya Narayana, Mandana Salimi, Anupama Gururaj,
Lucila Ohno-Machado, Hua Xu, and Todd R Johnson .................................................. 337

PIE: A prior knowledge guided integrated likelihood estimation method for bias reduction in
association studies using electronic health records data
Jing Huang, Rui Duan, Rebecca A Hubbard, Yonghui Wu, Jason H Moore, Hua Xu,
and Yong Chen .............................................................................................................. 345

Clinical documentation variations and NLP system portability: a case study in asthma
birth cohorts across institutions
Sunghwan Sohn, Yanshan Wang, Chung-Ii Wi, Elizabeth A Krusemark, Euijung Ryu, Mir H Ali,
Young J. Juhn, and Hongfang Liu .................................................................................. 353

Patient preferences toward an interactive e-consent application for research using electronic
health records
Christopher A Harle, Elizabeth H Golembiewski, Kiarash P Rahmanian, Janice L Krieger,
Dorothy Hagmajer, Arch G Mainous III, and Ray E Moseley ........................................... 360

PERSPECTIVE
High-fidelity phenotyping: richness and freedom from bias
George Hripcsak and David J Albers ................................................................. 289

BRIEF COMMUNICATION
Development of a research-oriented system for collecting mechanical ventilator waveform data
Gregory B Rehm, Brooks T Kuhn, Jean-Pierre Delplanque, Edward C Guo, Monica K Lieng,
Jimmy Nguyen, Nicholas R Anderson, and Jason Y Adams ........................................... 295

CLAMP – a toolkit for efficiently building customized clinical natural language processing pipelines
Ergin Soysal, Jingqi Wang, Min Jiang, Yonghui Wu, Serguei Pakhomov, Hongfang Liu,
and Hua Xu .......................................................... 331