Received: 1 December 2021 | Accepted: 1 December 2021

DOI: 10.1002/lrh2.10300

Learning Health Systems

ABSTRACTS

Poster abstracts from fourth annual public meeting: Mobilizing computable biomedical knowledge (MCBK 2021)

Technical posters

A heterogeneous knowledge environment for cognitive support applications

Davide Sottara[1], PhD; Jane L. Shellum[1], MS; Adam M. Bartscher[1]; Marc L. Sainvil[1], MS; Matthew D. Hegland[1]; Branden C. Hickey[1], MAS
[1]*Mayo Clinic, 200 1st Street SW, Rochester, MN 55905, USA {family.firstname}@mayo.edu*

Modern cognitive support applications are complex systems that combine background (bio)medical knowledge with knowledge about physiological processes, decision-making criteria, evidence-based guidelines, clinical workflows, human-computer interaction protocols, and semantics of data sources, including electronic health records. For each category of knowledge artifacts, the knowledge can be curated, distributed, and executed using different industry standards—usually published by organizations such as OMG, HL7, OASIS, and the W3C—leveraging commodity tools for representation and reasoning purposes. To manage the inherent complexity, we propose an architectural approach based on the notion of institution[1] first applied to ontology languages. Each artifact is categorized in terms of its knowledge content, described using metadata about its representation (language/profile, terminology, serialization, encoding), annotated semantically, and related using the notions of composition, derivation, and dependency. This architecture is an OMG standard itself (API4KP[2]) and adheres to the MCBK principles[3] of metadata fairness and trust. As a demonstration, we present the "Mayo Expert Advisor" use case: an application framework that involves 20+ types of knowledge artifacts and uses 15+ knowledge representation standards. This poster presentation is particularly relevant to knowledge engineers trying to build architectures that support different types of interoperable computable biomedical knowledge artifacts.

REFERENCES

1. Till Mossakowski, Oliver Kutz, Christoph Lange. Semantics of the Distributed Ontology Language: Institutes and Institutions. 21st International Workshop on Algebraic Development Techniques (WADT), Jun 2012, Salamanca, Spain. pp. 212-230, 10.1007/978-3-642-37635-1_13.

2. Athan, Tara & Bell, Roy & Kendall, Elisa & Paschke, Adrian & Sottara, Davide. (2015). API4KP Metamodel: A Meta-API for Heterogeneous Knowledge Platforms. 9202. 144-160. 10.1007/978-3-319-21542-6_10.

3. Alper, Brian & Flynn, Allen & Bray, Bruce & Conte, Marisa & Eldredge, Christina & Gold, Sigfried & Greenes, Robert & Haug, Peter & Jacoby, Kim & Koru, Gunes & McClay, James & Sainvil, Marc & Sottara, Davide & Tuttle, Mark & Visweswaran, Shyam & Yurk, Robin. (2021). Categorizing metadata to help mobilize computable biomedical knowledge. Learning Health Systems. 10.1002/lrh2.10271.

Use of CEDR big data for elucidating COVID-19's impact on emergency care

Dhruv B. Sharma[1], MS, ACEP, Irving, TX; Dr. Pawan Goyal[1], MD, MHA, PMP, MS, ACEP, Irving, TX; Dr. Arjun K. Venkatesh[1], MD, MBA, MHS
[1]*Yale School of Medicine, New Haven, CT, USA*

The COVID-19 pandemic generated several crises specific to the emergency department (ED) including large drops in ED patient volumes and increased burden of capturing valid pandemic data to measure nationwide effects in emergent ED conditions. Since 2015, the American College of Emergency Physicians (ACEP) has been developing and maturing its Clinical Emergency Data Registry (CEDR) into a big data resource to facilitate data reporting and democratize nationwide emergency medicine analytics. ACEP, in partnership with the Yale School of Medicine, successfully queried, normalized, and analyzed ED visits before and during the COVID-19 pandemic elucidating the effects on the occurrence of emergent ED conditions. This poster presentation is particularly relevant to front-line clinicians and hospitals facing COVID-19-related burdens, as well as public health researchers, informaticists, and specialty societies interested in big data use cases for medical specialties.

Decentralized and reproducible geocoding and characterization of community and environmental exposures for multi-site studies

Erika Rasnick[1], MS; Andrew Vancil[1], MS; Cole Brokamp[1], PhD
[1]*Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA*

DeGAUSS is a computable biomedical knowledge system for offline geocoding and assessment of place-based sociodemographic and environmental exposures in health studies. DeGAUSS alleviates common problems associated with multisite studies related to data privacy restrictions and lack of reproducibility due to differences in computational methods at different sites and is accessible for those not experienced in geospatial computation. This poster presentation is particularly relevant to health researchers interested in place-based predictors of health outcomes.

A hybrid clinical reasoning approach that includes abduction

Sabbir M. Rashid[1], Jamie P. McCusker[1], Oshani Seneviratne[1], Daniel Gruen[1], Amar K. Das[2], Deborah L. McGuinness[1]
[1]*Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180, USA;* [2]*IBM, 75 Binney Street, Cambridge, MA 02142, USA*

Computable biomedical knowledge is typically designed to support deductive forms of reasoning. Through think-aloud studies of clinical use cases in diabetes care, we have found that physicians employ additional types of reasoning, including abduction and induction. In particular, physicians interweave various forms of logic into dynamic reasoning strategies that mirror the Select-Test (S-T) model proposed by Stefanelli and Ramoni. To support physician's cognitive models, we have chosen to design a clinical knowledge base using the cyclic set of steps in the S-T model. We leverage Semantic Web technologies to encode each step of the S-T model as an AI task that uses a distinct form of reasoning, such as abduction. In order to perform abduction using semantic technologies, we trace backward through graphs containing reasoning rule paths. We then compose the AI tasks into a hybrid-reasoning architecture that can support a particular clinical reasoning strategy, such as differential diagnosis. In doing so, we are constructing various types of novel reasoners that are compatible with description logics. We plan to evaluate the reasoning system and rule representation through testing of clinical use cases that will determine whether system-generated insights align with what domain experts would conclude.

Leveraging YouTube video analytics for health literacy: An augmented-intelligence machine learning pipeline to retrieve and recommend videos on chronic and infectious diseases

Rema Padman[1], PhD; Xiao Liu[2], PhD; Anjana Susarla[3], PhD
[1]*Carnegie Mellon University, Pittsburgh, PA, USA;* [2]*Arizona State University, Tempe, AZ, USA;* [3]*Michigan State University, Lansing, MI, USA*

The easy availability of a vast repository of user-generated health information on social networks, blogs, YouTube, and Twitter presents a significant opportunity to investigate how social media channels can inform and communicate healthcare information for patient-centric health promotion and literacy improvement. YouTube hosts millions of healthcare-related videos on many medical conditions. This plethora of user-generated content can be leveraged by health systems and public health agencies to improve health literacy, self-care, and adherence to guidelines for chronic and infectious disease management. We propose a machine learning-driven augmented-intelligence approach that combines human input from domain experts with machine learning and natural language processing methods to winnow down and retrieve relevant and contextualized YouTube videos that health systems and public health experts can review and recommend to end users. Our computational approach filters videos on multiple dimensions, such as the medical information encoded in the video, its relevance and understandability so that content experts can efficiently review the reduced set for accuracy and credibility. This poster presentation is particularly relevant to researchers, clinicians, and consumers to understand how to curate and evaluate health information in rich multimedia format for prescribing at the point of care, to facilitate uptake, and improve health literacy and societal health.

Leveraging CBK to support learning health systems and their efforts to realize the Quintuple Aim

Jerome A. Osheroff[1], MD; Steve Bernstein[2], BS; Dave Carlson[3], PhD, MBA; Raj S. Ambay[4], MD
[1]*TMIT Consulting, LLC, Naples, FL, USA,* [2]*Agency for Healthcare Research and Quality, Health and Human Services, Rockville, MD, USA;* [3]*TISTA Science and Technology, Ft. Collins, CO, USA;* [4]*TISTA Science and Technology, Tampa, FL, USA*

Learning health systems (LHSs) optimize care delivery and outcomes (i.e., achieve the Quintuple Aim [ahrq.gov]) by creating a virtuous cycle from evidence to guidance to action to data and back to evidence. LHSs are rare and difficult to establish, in part because knowledge needed to drive this cycle is not FAIR (findable, accessible, interoperable, reusable) or computable. The AHRQ evidence-based Care Transformation Support (ACTS) initiative [digital.ahrq.gov] has produced a stakeholder-driven vision and roadmap for producing a knowledge ecosystem—driven by computable biomedical knowledge—to broadly realize effective LHSs. This poster illustrates this future vision coming to life through an individual patient's journey that covers managing preventive care and acute and chronic illness. Complementing this journey is an outline of the tools and infrastructure needed to mobilize computable biomedical knowledge in ways that make the vision a widespread and successful reality. The material described is an ACTS initiative work-in-progress, and meeting attendees will be invited to provide input into the vision and infrastructure needs, and engage in exploring next steps to broadly realize them. This poster presentation

is particularly relevant to those working to mobilize computable knowledge in ways that support care and LHSs, as well as those who receive or provide care.

Apervita Vital™ Platform, and Apervita Knowledge Studio

Blackford Middleton[1], MD, MPH, MSc, FACP, FACMI, FHIMSS, FIAHSI; Kenji Wong[1]
Apervita, 1 North LaSalle, Suite 1400, Chicago, IL 60602, USA

The Apervita Vital Platform supports open standards-based knowledge representation formalisms such as the HL7 Clinical Reasoning Module, FHIR Foundation, Computable Practice Guideline on FHIR, and the HL7 Clinical Quality Language (CQL), and can also encode other nonstandard logic in our clinical computation engine. The Apervita Platform was demonstrated at MCBK 2019. It has not been demonstrated in any other conference since that time. The Apervita Knowledge Studio would be a first-time demonstration. Features include: (1) general healthcare quality measurement and delivery across all health domains and (2) quality measurement and cognitive support across multiple care settings and disparate EHRs—across the patient's continuum of care—to enable the Learning Healthcare System. The primary users are healthcare delivery organizations, measure developers and implementers, and accreditation agencies measuring healthcare quality, and care delivery teams receiving decision support at the point of care.

Biomed News – A biomedical literature expertise sharing system based on machine learning and expert curation

Gavin P. McStay, PhD[1]; Thomas Krichel[2], PhD
[1]Department of Biological Sciences, Staffordshire University, Stoke-on-Trent ST4 2DF, UK; [2]Open Library Society, New York, NY, USA

Approximately 30,000 biomedical research abstracts are published each week on PubMed. This provides an enormous resource for new information that needs to be retrieved by individuals who desire access to it—researchers, industry, funders, policy makers, clinicians, patients, and the public. This enormous amount of information is retrieved using searches or tables of contents. However, these methods are overwhelming, broad, and irrelevant. Biomed News is a platform that provides a ranked abstract list on a weekly basis. Topic experts select abstracts relevant to their area of expertise. These selections are used by supervised machine learning to identify relevant papers in subsequent weekly issues. This reduces the amount of time required to identify new abstracts. Biomed News also provides an opportunity for selectors to share their weekly topic issues with other interested parties. Biomed News currently has around 75 experts from countries all over the world who are at different stages of their career. We aim for Biomed News to be a user-friendly platform in all areas of medicine that allows new abstracts to be identified and shared with interested parties. We will demonstrate the user interface of Biomed News (previously demonstrated at the CILIP HLG Annual Meeting in 2018).

A research-protocol object to generate biomedical knowledge that is auditable and reproducible

Harold P. Lehmann[1], MD, PhD; Amin A. Manna[2], MEng; Kenneth J. Wilkins[3], PhD; Katie R. Bradwell[2], PhD; Benjamin Amor[2], PhD; Johanna J. Loomba[4], ME, CCRC; Eli B. Levitt[5], MS; Andrew Williams[6], PhD
[1]Johns Hopkins, Baltimore, MD, USA; [2]Palantir Technologies, Denver, CO, USA; [3]National Institutes of Health, National Institute of Diabetes & Digestive & Kidney Diseases, Bethesda, MD, USA; [4]University of Virginia, Charlottesville, VA, USA; [5]Translational Medicine, Florida International University Herbert Wertheim College of Medicine, Miami, FL, USA; [6]Tufts Medical Center, Boston, MA, USA, For the Applicable Data Methods and Standards Workgroup, National COVID Cohort Collab

Guidelines and best practices for performing research are proliferating, whether in the clinical-epidemiological community (e.g., equator-network) or informatics (e.g., OHDSI). However, these imperatives require translation from a document into practice. A computable guideline would be able to drive and orchestrate research at multiple phases, accelerating and jump-starting analyses, without the need for re-translation at each step: Definition of population, intervention, comparator, outcome, and time horizon in a protocol; using the same definitions for eligibility criteria; then again for recruitment queries or database extraction, and then again for defining variables and performing analysis. We have created a Protocol Object for the National COVID Cohort Collaborative (N3C) that enables these functions, embedded in the computational environment of the N3C Enclave. The Object is structured along the lines of an OHDSI analytic plan and the STaRT-RWE checklist. As the user fills out the template, or links it to data/analytic resources in the Enclave, the Protocol Object is linked to objects that then serve as input to other processes in the Enclave. The result is a self-documented protocol that can be audited and re-executed, while at the same time providing analysts with decision support, using best practices.

AHRQ CEPI evidence discovery and retrieval (CEDAR)

Peter Krautscheid[1], BS; Marc Hadley[1], PhD; Mario Teran[2], MD, Edwin Lomotan[2], MD
[1]The MITRE Corporation, McLean, VA, USA; [2]Agency for Healthcare Research and Quality, Rockville, MD, USA

CEDAR focuses on the domain of patient-centered outcomes research (PCOR) and is a standards-based application programming

interface (API) that supports search, access, and use of PCOR findings across multiple repositories in AHRQ's Center for Evidence and Practice Improvement (CEPI). In addition, CEDAR aligns the AHRQ repositories with the FAIR (findable, accessible, interoperable, and reusable) data principles. CEDAR allows developers and EHR implementers to integrate AHRQ PCOR findings directly into their existing systems, where the findings can then be accessed and used by researchers, clinicians, policy makers, patients, and others. The CEDAR demonstration will show how the API facilitates search of and access to AHRQ PCOR findings from directly within an EHR system via a SMART on FHIR interface (visualized in the Concept of Operations appended below). This is the first time CEDAR has been demonstrated at a scientific conference. CEDAR impacts MCBK as it includes computable biomedical knowledge from multiple repositories, including data underlying systematic reviews from the AHRQ Systematic Review Data Repository, evidence reports from the AHRQ Effective Health Care Program, and clinical decision support artifacts from AHRQ CDS Connect.

Computable case reporting for multicenter clinical trials and registries

Andrew J. King[1], PhD; Salim Malakouti[1], MS; Edvin Music[1]; Kyle Kalchthaler[2]; Jonathan Holton[1]; Kevin Quinn[2], MBA; Gilles Clermont[1,2], MD, MSc; Oscar Marroquin[2], MD; Derek C. Angus[1,2], MD, MPH, FRCP; Christopher Horvat[1,2], MD, MHA
[1]University of Pittsburgh, Pittsburgh, PA, USA; [2]University of Pittsburgh Medical Center (UPMC), Pittsburgh, PA, USA

Both clinical trials and multicenter patient registries rely on case report forms (CRFs) with accompanying completion guidelines to direct standardized data collection across participating sites. Conventional CRFs are completed by manual data abstraction, using either structured paper forms or web-based electronic data capture portals. Automated CRF data abstraction from electronic records using programmed queries is a promising approach to improve the efficiency of collection, particularly for high enrolling trials and large registries; however, in many countries, disparate electronic record systems with non-interoperable reference data and workflow-dependent collection schemas present challenges to the uniform collection of data across centers. Additionally, trial and registry collection guidelines and database architects infrequently consider electronic data harmonization during the design phase, often generating avoidable challenges. We present a set of recommendations for individuals seeking to leverage automated, programmed data abstraction for CRF completion. We account for evolving interoperability requirements in the United States, which are anticipated to facilitate this type of work. This poster presentation is particularly relevant for informaticists, analysts, trialists, registry owners, database architects, and clinicians who are seeking a pragmatic approach to bridge the gap between the current and ideal state of electronic health data exchange for CRF completion.

A public repository to mobilize computable biomedical knowledge artifacts

Gunes Koru[1], Sai Yerram[1], Abir Rahman[1]
University of Maryland, Baltimore County, Baltimore, MD, USA

The Standards Workgroup of the MCBK initiative has developed and published metadata categories to make CBK artifacts (CBKs) findable, accessible, interoperable, reusable, and trustable (FAIR+T) [1]. The workgroup recommended further development and refinement of the categories as an important next step for mobilizing CBKs. Toward this goal, we developed an online repository that allows health professionals to create public CBKs (PCBKs) with metadata fields in 13 categories. The repository presents a documented application programming interface (API) consistent with the OpenAPI Version 3 specifications. Our team created API libraries in 18 different programming languages to facilitate mobilizing PCBKs conveniently. Discussion threads in the repository provide users with a forum for exchanging information about PCBKs. The repository site also includes user surveys that assess the perceived usefulness and complexity of using the metadata categories in producing and using PCBKs. The repository is operational and can be visited at https://cbk.pub. The site serves both health professionals interested in publishing their CBKs widely and finding other CBKs, and our study team in investigating effective and efficient metadata standards to make public CBKs FAIR+T.

Reference

1. Alper et al., Categorizing metadata to help mobilize computable biomedical knowledge, Learning Health Systems, May 2021. doi: 10.1002/lrh2.10271.

Omicsfold: An end-to-end advanced analytics and inference pipeline to accelerate drug discovery in the era of multi-omics biology

Vancheswaran Gopalakrishnan[1], Taylor Cohen[1], Antonio DiGiandomenico[1], Benjamin C. Sparklin[1], Elliot Young[2], Chiyun Lee[2], Carlos Roca[2], Philip Tedder[2], Igor Filipiuk[2], Stuart McHattie[2,3], Jonathan Ayling[2,3], Jonathan Mangion[2], Mishal Patel[4], Mathew Woodwark[2], Jim Weatherall[5], Mark Esser[1], Faisal M. Khan[6], Bret Sellman[2,*] and Khader Shameer[6,*]
[1]Microbial Science, AstraZeneca, Gaithersburg, MD, USA; [2]Data Science & Artificial Intelligence, Biopharma R&D, AstraZeneca, Cambridge, UK; [3]Tessella, Abingdon, UK; [4]Clinical Pharmacology and Safety Sciences, AstraZeneca, UK; [5]Data Science & Artificial Intelligence, R&D, AstraZeneca, Macclesfield, UK; [6]Data Science & Artificial Intelligence, Biopharma R&D, AstraZeneca, Gaithersburg, MD, USA

[*]Correspondence: bret.sellman@astrazeneca.com; shameer.khader@astrazeneca.com.

The modern drug discovery process routinely uses a plethora of molecular profiling technologies to investigate underlying biological relationships. Such profiling technologies include metagenomics, transcriptomics, metabolomics, immune profiling, etc. Understanding the relationships inferred from biological data types could aid in the discovery of targets, mechanistic pathways, de novo drug discovery, and drug repositioning.[1,2] The lack of efficient translational bioinformatics workflows combining data from omics platforms, clinical data, non-omics experimental platforms, and machine intelligence approaches leads to a data generation-to-interpretation gap. Multi-omics analyses are imminently warranted to understand complex putative mechanisms by considering measurements provided by multiple platforms in an unbiased manner. Herein, we present "Omicsfold," an advanced, general machine learning pipeline to automate the integration of heterogeneous omics datasets built on top of mixOmics[3] with additional features including novel data preprocessing and model interpretation. Capabilities: The Omicsfold framework has critical advantages as it combines statistical approaches, machine intelligence methods, and biological interpretability in a single framework. Our objective is to develop a framework that condenses large data sets into generalizable representatives across different biological data types and can eventually help understand novel underlying systems. To achieve this, we leverage linear multivariate techniques including regularized generalized canonical correlation analysis to project multiple, independent lines of evidence onto a common latent component (LC) space. As such, our approach applies selection to identify discriminant features that load onto LCs and explain an outcome in a biological context. Furthermore, it generates an optimized ranking of features across blocks for prioritization using a novel approach called BlockRank. Importantly, our workflow provides a multidimensional correlational network of selected features across all biological data layers to elucidate new targets, combinatorial biomarkers, and pathways of interest. The utility of Omicsfold has been demonstrated on several datasets and has shed light on novel biological mechanisms and potential targets.

References

1. K. Shameer, M. A. Badgeley, R. Miotto, B. S. Glicksberg, J. W. Morgan, and J. T. Dudley, "Translational bioinformatics in the era of real-time biomedical, health care and wellness data streams," *Brief Bioinform*, vol. 18, no. 1, pp. 105-124, Jan 2017, doi:10.1093/bib/bbv118.

2. K. Shameer *et al*., "Systematic analyses of drugs and disease indications in RepurposeDB reveal pharmacological, biological and epidemiological factors influencing drug repositioning," *Brief Bioinform*, vol. 19, no. 4, pp. 656-678, Jul 20 2018, doi: 10.1093/bib/bbw136.

3. F. Rohart, B. Gautier, A. Singh, and K. A. Le Cao, "mixOmics: An R package for 'omics feature selection and multiple data integration," *PLoS Comput Biol*, vol. 13, no. 11, p. e1005752, Nov 2017, doi: 10.1371/journal.pcbi.1005752.

Publets: Towards a process for creating and publishing executable models of best clinical practice

John Fox[1], PhD, Matt South[2], PhD, Peter Ashby[3], Krishnarajah Nirantharakumar[4], MD, PhD, Katie Fox[5], VetMB, MA, Omar Khan[6], MEng

[1]OpenClinical, Oxford, United Kingdom; [2]Dept of Psychiatry, Oxford University; [3]Director of Publishing, OpenClinical, Oxford, United Kingdom; [4]University of Birmingham College of Medical and Dental Sciences, Birmingham, United Kingdom; [5]Larkmead Veterinary Group, Oxfordshire, United Kingdom; [6]University of Warwick, Faculty of Science Engineering and Medicine, Coventry, United Kingdom

This poster is relevant to clinicians and others interested in improving quality, safety, and economics of patient care, updating presentations about the OpenClinical project at MCBK in 2019. The focus here is on the development of executable knowledge models called Publets. This methodology uses the PRO*forma* AI language, which has proved to be highly versatile, with many demonstrable publets in the OpenClinical repository, and 16+ successful trials and other evaluations published in peer-reviewed medical journals (review in BMJ Health Informatics). Publets are very similar to "Computable Knowledge publications" as defined in Learning Health Systems, specifically the combination of conventional human readable articles and executable models online. Recent publets include an executable clinical guideline for COVID-19 care (Learning Health Systems 2020) and an application and evaluation in veterinary medicine (JSAP 2021). The presentation will include a short demonstration of *Patient Journey*, a service in development by OpenClinical for creating publets in an open access, open-source repository (a login will be available for MCBK delegates). Current work is focused on refining the usability of our development tools and methods, implementing safe and ethical publishing procedures, and an updated knowledge engineering language PRO*forma* 2020.

Towards providing clinical context for a diabetes risk-prediction use case via user-centered explainability

Shruthi Chari[1], MS; Olivia Zhang[2], Prasant Acharya[1], Fernando Suarez Saiz[3], MD; Mohamed Ghalwash[2], PhD; Elif Eyigoz[2], PhD; Oshani Seneviratne[4], PhD; Daniel Gruen[4], PhD; Pablo Meyer Rojas[2], PhD; Prithwish Chakraborty[2], PhD; Deborah L. McGuinness[1], PhD

[1]Tetherless World Constellation, RPI, Troy, NY, USA; [2]Center for Computational Health, IBM Research, Yorktown Heights, NY, USA; [3]IBM Watson Health, Cambridge, MA, USA; [4]Institute for Data Exploration and Application, RPI, Troy, NY, USA

This poster presentation is aimed at researchers and clinicians looking for context-aware patient information and explanations related to poly-chronic disease prediction. In real-world healthcare settings, high-precision prediction models often have inadequate interpretability and explainability of model output. We address this challenge with a question-answer approach that identifies actionable domain

knowledge to contextualize model predictions. We choose a chronic disease management setting—type-2 diabetes (T2D) comorbidities risk prediction—because the monitoring and assessing of comorbidities is an example where clinicians consider multiple data points during decision-making. In a proof-of-concept (POC) implementation focusing on chronic kidney disease as T2D comorbidity, we use machine learning models to predict risk, post hoc explainer models to understand prediction dependencies, and a question-answering module to operationalize T2D clinical practice guidelines. We address a pre-canned set of questions spanning population health and patient-level treatment planning scenarios. We also consulted a medical expert during the model design and development process. The POC output is translated into clinician-facing insights on a user-centered dashboard, which we will use as a probe in user interviews to identify improvements for clinical relevance and usability. We hope such feedback will help us generate relevant, easy to understand and inspect explanations to better contextualize patients' comorbidities.

Tracking daily mood ratings, activity, sleep, and physiologic factors prior to a manic episode in a patient with type I bipolar disorder

Ramona Bledea[1] (presenter), BA; Michael Yee[2], MD; Helen J. Burgess[2], PhD; Amy Cochran[3]; Melvin G. McInnis[2], MD
[1]University of Michigan Medical School, Ann Arbor, MI, USA; [2]Heinz C. Prechter Bipolar Research Program, Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA; [3]Department of Population Health Sciences and Mathematics, University of Wisconsin-Madison, Madison, WI, USA

Recent studies have described synchronicity between mood, sleep duration, and the lunar cycle, which influences daily light exposure via moonlight. To our knowledge, this is the first report to track synchronicity between mood, activity, heart rate, and sleep onset and duration across the lunar cycle immediately prior to a hospitalization event for a manic episode in a patient with rapid cycling type I bipolar disorder. For 8 weeks prior to hospitalization, a FitBit tracked sleep duration and onset timing, intradaily resting heart rate, and intradaily activity intensity. The DigiBP app surveys patients with bipolar disorder for mood changes and was completed once in the morning and once in the evening. Lunar cycling data were collected from TimeAndDate. Mean values of each variable and timeframe were found and data were fitted to curves using previously established models from Casiraghi et al. D scores peaked 2 weeks prior to hospitalization. M scores rose gradually until 3 weeks prior, and were higher in the morning, whereas d scores were higher at night. Sleep duration varied by as much as 400 minutes between nights. Two weeks prior to hospitalization, average sleep duration was shortened by 14 minutes. Full moon dates corresponded to nights with shorter sleep duration. New moon dates corresponded to nights with longer sleep duration. A drop in average resting heart rate was also observed two weeks prior to hospitalization. “Very active” and “fairly active” activity levels demonstrated similar patterns of spiking at this time, as did d scores. Mood scores and sleep onset and duration changes near full and new moon dates appeared to synchronize, further supporting recent studies. The counterintuitive drop in average resting heart rate corresponding to the rise in d scores and activity intensity levels also suggest synchronicity and requires further exploration. These variation patterns, taken together, could be the basis for guiding future studies toward building a comprehensive picture of clinical signs that can be used in practice to predict oncoming manic and depressive episodes.

Sorting and presenting clinical trial results for public health practitioners

Baskar, R.[1]; Balas, E. A.[2]; Dhantu S.[3]; Bussi S.[4]; Amour C.[5]; Kamugisha H.[6]
[1]Medical College of Georgia, Augusta, GA, USA; [2]Biomedical Research Innovation Laboratory, Augusta University, Augusta, GA, USA; [3]University of Alabama at Birmingham, Birmingham, AL, USA; [4]Tanzania People's Defence Force, Tanzania; [5]Kilimanjaro Christian Medical University College, Tanzania; [6]Military College of Medical Sciences, Tanzania

Importance: As the number of clinical trials is growing, automation in literature searches is becoming desirable but largely unresolved. Current automation techniques like artificial intelligence cannot retrieve information that is not in the publication. Objective: The purpose of this study was to assess the readiness of clinical trials and their reports for supporting practical implementation and automated retrieval. Design: We searched the Medline database for a random sample of clinical trials of HIV/AIDS management with potential relevance to public health in Africa. Five authors assessed trial reports for inclusion, extracted data, and assessed trial quality. Subsequently, we categorized various aspects of the reported trial results in terms of outcomes, process of care, or structure of health care. Finally, recommendations were developed for more effective trials and reporting. Findings: In our sample, 97 randomized controlled trial reports were selected. Of the chosen articles, 20.8% of the articles identified as cluster RCT, and 54.9% of articles were multicenter studies. Information essential for practical implementation was largely missing: personnel resources needed 32.3% (.95 CI: 22.9-41.6); material/supply changes needed 33.3% (.95 CI: 23.9-42.8); major equipment/building investment 42.8% (CI: 33.8-53.7); methods of educating providers 53.1% (CI: 43.1-63.4); methods of educating the community 27.1% (CI: 18.2-36.0). A total number of 56 studies (57.7%) measured health/biologic outcomes, and among them only 41.1% showed positive effect. Many reports provide complex multivariate analyses or only graphic illustrations of the results that are provided without clearly stating that the differences were significant or nonsignificant. To sort and interpret clinical trial results faster and easier, a new structured reporting format is presented. Conclusions: A large percentage of clinical trials is either ineffective by design or report inconsequential messages. It is important to improve trial quality and reporting at

the time of production rather than afterwards. To facilitate automated searches and also public health improvement, impact-oriented clinical trial reporting is recommended.

Beyond safe harbor: Risk of exposing location in de-identified clinical data

Alfred Jerrod Anzalone[1], MS; Carol R. Geary[2], PhD, MBA, RN, Department of Pathology and Microbiology; James C. McClay[3], MD, MS

[1]Department of Neurological Sciences, College of Medicine, University of Nebraska Medical Center, Omaha, NE, USA; [2]Department of Pathology and Microbiology, College of Medicine, University of Nebraska Medical Center, Omaha, NE, USA; [3]Department of Emergency Medicine, College of Medicine, University of Nebraska Medical Center, Omaha, NE, USA

De-identified clinical research data warehouses (DI-CRDW) are commonly derived from electronic health records with identifiers removed. The Safe Harbor rule (1) requires strict exclusion of the 18 HIPAA identifiers (Table 1), including non-aggregated location units smaller than the state. (2) It is commonplace to enhance DI-CRDW data through linkage with external data sources to create a more comprehensive and accurate patient profile. These external data sources include location-based indices, such as the rural-urban commuting area (RUCA) codes (Table 2). (3) Location-based indices are typically not identifying as they capture large, geographically dispersed groups. However, their deployment within a DI-CRDW or an individual research project using multiple location-dependent elements presents inherent privacy concerns. This poster presentation is particularly relevant to informaticians and researchers looking to understand the privacy challenges for working with location-based indices in a DI-CRDW. We provide an example of where such practices lead to inadvertent PHI exposure using RUCA to highlight the real risk of reidentification (Figure 1). While research is important, the preservation of patient privacy needs to be built into institutional models for expanding research using real-world data. While researchers have a vested interest in preserving patient privacy, system architecture needs to ensure the system reduces the risk of reidentification.

Centre for addiction and mental health (CAMH) brainhealth databank knowledge graph

David Rotenberg[1], Operations Director & Data Strategy and Business Intelligence Director, MSc; Adeel Ansari[1] (presenter), Knowledge Engineer; Nikola Bogetic[1], Software Engineer; Joanna Yu[1], BrainHealth Databank Senior Portfolio Manager, PhD; Sean Hill[1,2], Scientific Director, PhD, Professor Psychiatry

[1]Krembil Centre for Neuroinformatics, CAMH; [2]University of Toronto, Canada

The CAMH BrainHealth Databank Knowledge Graph (BHDB-KG) is a collection of clinical and research data and knowledge that is used to inform care, accelerate discovery, and drive innovation. The BHDB-KG is currently being used to drive clinical decision support dashboards, natural language processing pipelines, and generation of data cuts for quality improvement and research analyses. The resulting products are consumed by stakeholders across the hospital including clinicians, researchers, administrators, and operations. The BHDB-KG data infrastructure consists of a number of open-source tools that integrate multiscale data from disparate sources including electronic health records, surveys, imaging, genomics, and wearables into a unified searchable knowledge graph. The BHDB-KG uses Blue Brain Nexus (bluebrainnexus.io), an open-source data management tool that enables the flexible creation of semantic knowledge graphs using virtually any data model and ontology. For the BHDB-KG, in order to ensure robust adoption of the FAIR—findable, accessible, interoperable, reusable—data principles, we implement provenance tracking, data quality measures, and standardized ontologies including FHIR/HL7.

A FHIR framework to ignite biomedical knowledge management

Muhammad Afzal[1]; Brian S. Alper[2]; Joanne Dehnbostel[3]; Andrey Soares[4]

[1]Innovation Center 422, 209 Neungdong-ro, Gwangjin-gu, Seoul 05006, South Korea; [2]Computable Publishing LLC, Ipswich, MA, USA; [3]Computable Publishing LLC, Ipswich, MA, USA; [4]University of Colorado, Aurora, CO, USA

Management of biomedical knowledge is essential to support scientific, educational, and healthcare processes and decisions. The knowledge management system (KMS) framework for current open biomedical repositories (such as PubMed Central) is document-based. Therefore, scientific communication relies on large sharable value units such as publications, reports, news articles, or books. New models are needed to manage the creation, retrieval, transmission, and storage for smaller sharable value units to go beyond the limits of document-based systems. The smaller sharable value units are called resources in the fast healthcare interoperability resources (FHIR) model. This poster presentation is particularly relevant to reframe the conceptual KMS framework for mobilizing computable biomedical knowledge (CBK). Through this effort, we aim to design a framework including eight accepted KMS processes—discovery, modification, translation, dissemination, creation, representation, storage, and retrieval. The proposed framework identifies roles and involvement of three principal elements of KMS–people, technology, and content, within each of the eight processes. An FHIR-based KMS will significantly increase mobilization of CBK because many smaller units (resources) can be used for many more things. Imagine when you can find the precise data for predicting your benefits and harms regarding your healthcare decision.