

Board of Scientific Counselors

Minutes of the 73rd Meeting

April 4-5, 2019

**Donald A.B. Lindberg Room
U.S. National Library of Medicine
Bethesda, Maryland**

**U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
PUBLIC HEALTH SERVICE**

BOARD OF SCIENTIFIC COUNSELORS OF THE US NATIONAL LIBRARY OF MEDICINE

Minutes of the Meeting

The Board of Scientific Counselors of the U.S. National Library of Medicine convened its 73rd meeting on Thursday, April 4, and Friday, April 5, 2019 in the Donald A.B. Lindberg Room of the U.S. National Library of Medicine, Bethesda, Maryland. The meeting was open to the public from 9:00 am to 2:30 pm on April 4 and from 9:00 am to 11:15 am on April 5. The meeting adjourned at 11:15 am on April 5. The Board Roster is enclosed.

Members of the Board of Scientific Counselors present:

Stephen M. Downs, MD
Peter J. Embi, MD, Chair
George M. Hripcsak, MD
Kevin B. Johnson, MD (via WebEx)
Susan A. Matney, PhD, RN
Shawn N. Murphy, MD, PhD
Ming Jack Po, PhD
Jessica D. Tenenbaum, PhD

**BOARD OF SCIENTIFIC COUNSELORS
LISTER HILL CENTER**

CHAIR

EMBI, Peter J., M.D. (6/30/19)
President/CEO, Regenstein Institute, Inc.
Sam Regenstein Professor of Medicine,
Associate Dean for Informatics & Health Services Research, IU School of Medicine
Associate Director of Informatics, Indiana CTSI
Vice President for Learning Health Systems, IU Health
1101 West Tenth Street
Indianapolis, IN 46202
317-274-9094 / 317-274-9302 (FAX)
pembi@regenstrief.org

DOWNS, Stephen M., M.D. (6/30/20)
Professor and Vice Chair for General
Pediatrics and Director of Children's
Health Services Research
Indiana University School of Medicine
Health Information & Translational
Science Building
410 West 10th Street, HS 1020
Indianapolis, IN 46202
317-278-0552 / 317-278-0456 (FAX)
stmdowns@iu.edu

MURPHY, Shawn N., M.D., Ph.D. (6/30/20)
Professor of Neurology
Harvard Medical School
Department of Neurology
Massachusetts General Hospital
Chief Research Information Officer
Partners HealthCare
399 Revolution Drive, Suite 725
Somerville, MA 02145-1446
857-282-3769
murphy.shawn@mgh.harvard.edu

HRIPCSAK, George M., M.D. (6/30/19)
Professor of Biomedical Informatics
Columbia University Medical Center
Columbia University
622 West 168th Street, PH20
New York, NY 10032
212-305-5712 / 212-305-5780 (FAX)
hripesak@columbia.edu

Po, Ming Jack, M.D., Ph.D. (6/30/22)
Product Manager, Google Research (Brain) and Google Cloud
1600 Amphitheatre Parkway, Mountain View, CA, 94043
718-577-2338 / 877- 991-8350
mail@jackpo.org

JOHNSON, Kevin B., M.D. (6/30/20)
Professor & Chair of Biomedical Informatics
Professor of Pediatrics
Chief Informatics Officer
Vanderbilt University Medical Center
2525 West End Avenue #1475
Nashville, TN 37203-8390
615-936-6867 / 615-936-0102 (FAX)
kevin.johnson@vanderbilt.edu

TENENBAUM, Jessica D., Ph.D. (6/30/21)
Assistant Professor
Department of Biostatistics & Bioinformatics
Division of Biomedical Informatics
Duke University
2424 Erwin Road, Box 2721
Durham, NC 27705
919-684-7308
919-451-5947 (Alternate Number)
jessie.tenenbaum@duke.edu

MATNEY, Susan A., Ph.D., RNC-OB (6/30/22)
Nurse Informaticist
Intermountain Healthcare
5171 S. Cottonwood Street, Murray, UT 84107
Salt Lake City, UT
801-447-9294 / 801-680-2161 (Cell)
susan.matney@imail.org

EXECUTIVE SECRETARY
Corn, Milton, M.D.
Deputy Director of Research and Education
National Library of Medicine
Building 38, Room 2S19
Bethesda, MD 20894
301-496-4725
cornm@mail.nih.gov

April 4, 2019

- I. The meeting was called to order by the Chair, Dr. Peter Embi, at 9:00 am.

Internal Cable Broadcast of BSC Public Sessions

Board members were informed that the open sessions of the meeting would be broadcast throughout the NLM and Lister Hill Center for other colleagues and staff to view. There were no objections.

- II. Dr. Milton Corn gave opening remarks and introductions.
- III. Dr. Dina Demner-Fushman presented *Natural language processing research at LHC*.
Dr. Sameer Antani presented *Image processing research at LHC*.
Dr. Olivier Bodenreider presented *Health information standards research at LHC*.
On behalf of Dr. Clement McDonald, Dr. Seo Baik presented *Health data-powered discovery at LHC*.
- IV. Dr. Patricia Brennan gave the NLM Director's Report.
- V. Dr. Paul Fontelo delivered his report on his report on *LHNCBC Training and Mentoring*.
- VI. Dr. Kevin Johnson reported on the NLM Blue Ribbon Panel Recommendations.
- VII. Mr. Jerry Sheehan presented the NLM Response to the Blue Ribbon Panel to Date.

April 5, 2019

- VIII. Drs. Michael Huerta and Olivier Bodenreider presented the Portfolio Analysis Report.
- IX. Board Report

The April 5, 2019 meeting of the Board of Scientific Counselors included the review of research programs at LHC.

REPORT OF THE BOARD OF SCIENTIFIC COUNSELORS - APRIL 2019

Preface

As the 50th year anniversary of The Lister Hill National Center for Biomedical Communications (LHC) concludes and the NLM plans for the evolution of its intramural programs, the mission of the LHC has never been more important. As the BSC has noted in multiple recent reports, and as the Blue Ribbon Panel has validated, the LHC's focus on foundational and applied informatics research complements the work of the NCBI and other NLM branches. The work highlighted during this meeting illustrates beautifully the unique niche being filled by the LHC as a bridge between the foundational work by Center scientists and its broad applicability to science and health.

As is the case with all foundational research in informatics, attempts to demonstrate the value of specific methods often require real-world evaluations. The BSC continues to believe that LHC investigators

should recognize when their potentially impactful science has reached a point where they should be taken to the next level of implementation and dissemination. At that time, collaborations with relevant partners might be warranted. For example, the BSC applauds the plan to create stronger connections across the NLM intramural programs and to engage with disease-focused NIH investigators from other ICs and external entities, such as other HHS agencies, universities, academic health centers, and industry partners. Examples of this kind of work and its impact are also ongoing, such as the work presented today to improve cervical cancer screening or identification of infections like malaria that could benefit from and/or inform the work of many others, both at the NIH and externally.

We recognize that this work is dependent on the efforts of intramural investigators and, in the past, has been threatened by attrition. We are heartened that the plans for evolving the intramural program include continuing the important work of LHC and hiring additional intramural investigators and staff. Given the critical importance of research advances in health informatics to enable data-driven research and improvements in the creation of a learning health system, we strongly recommend aligning several of the new positions with mission-critical components of the LHC.

The BSC also wishes to acknowledge and thank Dr. Clem McDonald for his exemplary leadership of the LHC. We also congratulate and recognize the exemplary leadership of Drs. Corn and Bodenreider in their new roles, and appreciated hearing from both of them during this meeting.

What follows is a summary of the examples of LHC efforts that were presented at today's BSC meeting, the BSC's reactions and recommendations, with an eye toward informing the NLM's plans as it continues its plan to develop a single intramural program based upon the NLM Strategic Plan and Blue Ribbon Panel report.

Investigator Presentations:

Dr. Dina Demner-Fushman - Natural Language Processing Research

Dr. Demner-Fushman presented on the extensive portfolio of projects emanating from a long history of research to improve medical "natural" language processing methods. Among the projects discussed were the CHIQA system for enabling automated answering of publicly provided questions, the Medical text indexer for automated support of NLM/PubMed indexing, the visual question answering tools, and the suite of foundational tools designed to execute NLP tasks.

Dr. Demner-Fushman mentioned that one of the most common misconceptions of her team's work is that Google or IBM Watson Health must already have operationalized the techniques upon which she is focusing. The BSC recognizes the importance of "public sector" projects that complement proprietary technologies; furthermore most of Dr. Demner-Fushman's work catalyzes all sectors involved with NLP. This work is collectively a stellar example of important work--both in methods research and applied tool improvement--that is well-suited to the NLM's mission and vision.

A few of these initiatives, such as the work to improve question-answering, while particularly valuable to users of MedLinePlus, could also be broadly usable to other search engines, the growing number of patient-facing apps and solutions, and potentially even EHRs attempting to summarize healthcare

encounters. The effort to transfer this technology into other applications does not appear to have a clear home within the investigator's group, but likely could be managed by other operationally focused components of the LHC portfolio.

Dr. Sameer Antani - Image Processing Research

The LHC's medical image analysis research program has been very productive, with over 30 papers published since just 2016. The research has two primary focus, 1) algorithms for specific clinical applications such as Glaucoma screening, Malaria screening, TB CXR screening, and Cervical Cancer Screening, and 2) limitations of algorithms and standards that current medical image analysis algorithms have.

We feel that both the applied clinical imaging analysis research AND the resulting understanding of the limitations of current algorithms are both essential research areas. The learnings gathered should be widely disseminated not only in the intramural and extramural research community, but also to the regulatory and industrial players (i.e. FDA, CMS, etc) who are actively working on bringing these technologies into clinical care. An example at NLM of this type of collaboration is the cervical screening project that is being jointly conducted with NCI, WHO, and the Gates Foundation.

We also appreciate the medical imaging infrastructure and standards work that is beginning to take shape at LHC. The ability of deep learning algorithms to create computable variables from text and images may translate into new standards. The effort to expand DICOM to accommodate characteristics of the images they annotate, for example, would be a valuable contribution.

Dr. Olivier Bodenreider: Health Information Standards Research

The LHC has a long tradition of health information standards research. This research is critical to improve their impact on health and healthcare globally. Past research includes UMLS, ontologies such as SNOMED CT, RxNorm, HL7 messaging standards such as FHIR, and common data elements (CDEs). The standards research has generally come from two perspectives, 1) assessing whether specific standards are fit for purpose (QA, interoperability, etc.) and 2) application of standards "in action" (assessing usefulness and effectiveness).

Assessing fit for purpose included QA and interoperability research such as identifying SNOMED CT hierarchical relationships using lattice graphs and mapping through pre-coordination. Application of standards "in action" included support for NLP, annotation, data integration, and mapping across terminologies. Examples include: Adverse drug event identification leveraging RxNorm, such as for drug analytics and comparisons among knowledge bases related to drug-drug interactions (which for instance identified differences that could be meaningful and need to be reconciled).

We feel that these research AND research-informed operational/developmental activities in health information standards are both essential and should not/cannot be separated given their importance to biomedical science and our larger health system. There remains a growing need for leadership to inform decisions around the best practices, choice and application of standards to enable uses of health data for care, research, learning health system development. This is a role LHC could play--either directly through expert guidance and/or as a convener of expert consensus related to standards applications and use. In

addition, LHC could lead in informing and aligning research between NLM and extramural researchers working in this area, such as in emerging needs for phenomic data standards, m-health applications, etc. Possible new areas of research could include evaluating syntactic structures, such as HL7 FHIR, C-CDA, and OpenEHR archetypes to support interoperability and NLP.

Dr. Clem McDonald (presented by Dr. Seo Baik) - Health Data Powered Discovery

Dr. Baik presented ongoing work aimed at analyzing a set of large health databases with the goal of learning from their strengths and weaknesses, improving them when possible, gaining insights from the data, and share expertise with others. Part of advancing the NLM's goals as "a Platform for Biomedical Discovery and Data-Powered Health," and to "Accelerate discovery and advance health through data-driven research."

The team considered a number of large dataset examples, but focused on MIMIC II/III, and CMS's Virtual Research Data Center (VRDC). Their goal is to learn best practices to analyze observational data by doing analyses on such datasets. Their work focuses on Cox proportional hazards modeling because of its adept handling of time dependent data and ability to account for a number of biases.

Results presented were based on the VRDC. One analysis looked at whether metformin offered a strong survival benefit ("fountain of youth") to patients with diabetes. They found that metformin did not offer a survival benefit, but diuretics, ACE-inhibitors, ARBs, and newer diabetes medications did.

A second analysis looked at a reported association between fluoroquinolones and tendon rupture. The results suggest that only some members of this drug family increased the risk, and members of other antibiotic families were also found to increase the risk.

The BSC noted that similar work is done by extramural researchers. For example, other work has already uncovered the confounding biases that may explain the erroneous metformin "fountain of youth" phenomenon.

LHC should consider related activities for which it is uniquely positioned. For example, aggregating, standardizing and optimizing such data (in collaboration with and on behalf of other agencies perhaps) for other researchers to use; providing opportunities for extramural researchers to augment datasets, taking advantage of similar efforts ongoing across our NLM-supported academic groups (and across all AHCs and general health systems for that matter); or becoming the objective, unbiased source for data and best practices to enable data-driven work that cannot be as effectively done by extramural groups nor by other HHS agencies without the expertise that NLM/LHC provides.

Some on the BSC noted that while this kind of work is usually done as part of extramural research, if the LHC evolves to address not just the research infrastructure, such as databases and ontologies, but also broader data science including causal methods, then continuing this work in a moderate scope will be important to inform the methodological goals. The group working on these studies must be deeply engaged with the rest of the intramural team, providing a demonstration of the potential value of standardized large clinical data sets, and exploring what the best practice approaches might be for analysis of inherently biased observational data. Extramural researchers have similarly found that having a very

tight connection between those carrying out clinical research and those developing methods leads to much better methods; this further confirms the value of this research within NLM.

Dr. Patricia Brennan - NLM Director Update

Dr. Brennan presented an update on activities related to the NLM Strategic Plan. The BSC was uniformly in agreement with two principles. First, we remain highly supportive of plans to enhance the existing footprint in ways that co-locate more of the staff--especially those who currently are away from the main campus. The phased approach appears to satisfy both financial constraints and some short-term needs of the NLM.

Second, while we generally applaud the changes to the organizational structure of the leadership team, we would encourage an explicit use of terms from the fields of informatics and data science in the naming of roles and organizational structures. The BSC is sympathetic to the concern about the hard-wiring of trendy vocabulary but concerned that by avoiding these terms, efforts to make the NLM's intramural activities visible to key external stakeholders could be stymied.

LHC Training Program

The LHC training program has done a spectacular job introducing trainees to informatics and moving trainees to the next level of their careers. The majority of trainees go into academics, which is considered a benchmark of success. The planned structural changes within NLM represent a unique opportunity to grow the program. As the LHC and NCBI training programs come together, the best components of each can be employed and critical mass can be achieved for addressing issues that are otherwise difficult in a very small program, including issues raised by the BSC in recent years' reports. Also, as the research components of LHC and NCBI come together, there is a great opportunity to focus on new areas of translational research informatics that have heretofore been beyond the scope of either program. Furthermore, we believe the training program will benefit from a more formal connection between the scientific and operational components of the NLM so that real problems can motivate the science.

Other considerations for the future

The mission of NLM may be considered to be a balance of service, education, and research, with much of the activities at the intersection of all three. As such, the creation and / or maintenance of data sets and tools, including the curation and annotation support of relevant datasets would be highly valuable and catalytic to researchers and educators alike. Examples of such datasets include extramural datasets and tools such as MIMIC III and, i2b2 NLP datasets, as well as NIH intramural efforts such as the CXR dataset from the NIH Clinical Center. The development of this ecosystem could enable the better positioning of NLM as a data science center and provide the public with a guaranteed level of service around the projects.

As the intramural budget expands in the future, the NLM may consider a few targeted personnel investments. One would be a more basic (theoretical) data science researcher who develops new classes of methods as well as applying them. (For example, the development of deep exponential families is the kind of work this person could have taken on.) Such a person would be a valuable additional resource to others at the NLM and throughout NIH who are applying methods in novel ways to their work. Another

would be data scientist with experience in pulling together modern empirical data science methods with mechanistic models generated throughout the NIH. This latter area will increase in importance over time as we attempt to generate personalized predictive models for patients.

The promotion of Data Science application and demonstration has often begun with a “Challenge,” where a data set for manipulation and a task to perform on the data is described and participants are “Challenged” to obtain the best results. Challenges have the potential to leverage the importance of NLM/NIH to draw contestants and promote bioinformatics and data science from across the country. Simply compiling the list of participating groups entering the contest and working on these competitive projects is of high value. The promotion and distribution of new methods could be achieved with up-front release to enable publication of the results, and an “open source” category so methods can be accessed and downloaded. At the same time, we caution not to over-rely on challenges because when they are overused, they can distract the field from innovation.

Blue Ribbon Panel Report Assessment

The Blue Ribbon Panel (BRP) was charged with reviewing NLM’s intramural research and training programs and making recommendations regarding NLM’s organizational structure, budget, staffing, partnerships, and its training program. They were instructed to make these recommendations in light of priority areas of biomedical informatics and data science research, and optimization of balance among research, development and services. In addition, the Panel was asked to suggest novel and non-traditional ways to assess outcomes and impact of research and training, and to align the NLM’s intramural research with the NLM and NIH strategic plans.

The BRP comprised nine nationally recognized experts in Informatics and related fields and was Chaired by Dr. Russ Altman from Stanford University. Two members of the Panel were current or former members of the LHC BSC. The process included 5 conference calls, review of relevant documents, and a two-day site visit with LHC, NCBI, NLM trainees, and other ICs.

Key findings from the BRP included endorsement of the vision of the Advisory Committee to the Director of NIH that NLM serves as a hub for informatics, data science, and knowledge management at NIH. The Panel also asserted the critical need for the NLM to have a vibrant and aggressive intramural research program that reinforces its unique ability to solve challenges in biomedical informatics and data science.

The Panel made 11 specific recommendations, documented [elsewhere](#). Specific significant steps taken to date to implement recommendations include: funding for, and active recruitment of, 3 new investigators; creation of and recruitment for a unified intramural Scientific Director; exploration of models and opportunities for joint appointments and collaboration with other ICs; a portfolio analysis of NLM “offerings” for the purpose of balancing research, development and service; engagement with the NIH Office of Scientific Workforce Diversity; establishment of a single NLM BSC; and designs for renovated office space to facilitate collaboration.

Conclusion

In summary, the BSC remains impressed and strongly supportive of the work of the highly talented, capable and accomplished LHC leadership, faculty investigators, scientists and staff. The future of NLM is bright and will rely on the continued strength of the work that LHC has and continues to lead in advancing science and practice of biomedical informatics for improving research, healthcare, and health. We look forward to the evolution of NLM's intramural program with great enthusiasm.