

NATIONAL LIBRARY OF MEDICINE, NIH
BOARD OF SCIENTIFIC COUNSELORS
MEETING MINUTES
April 22, 2021

The Board of Scientific Counselors of the National Library of Medicine (NLM) convened by webcast on April 22, 2021, between 11:00 a.m. and 4:15 p.m. The meeting was open for viewing via NIH VideoCast.

BSC Members Participating

Kevin Johnson, MD, Vanderbilt University Medical Center (*BSC Chair*)
Hyun Min Kang, PhD, University of Michigan
Kateryna Makova, PhD, Penn State University
Lucila Ohno-Machado, MD, PhD, University of California
Ming Jack Po, MD, PhD, Google Health
Steven Salzberg, PhD, Johns Hopkins University
Donna Slonim, PhD, Tufts University
Pamela Soltis, PhD, University of Florida
Peter Tarczy-Hornoch, MD, University of Washington
Jessica Tenenbaum, PhD, Duke University
Valerie Florance, PhD, NLM (*Executive Secretary*)

NIH Staff Presenting

Olivier Bodenreider, MD, PhD, LHNCBC, NLM
Patricia Flatley Brennan, RN, PhD, NLM
Valerie Florance, PhD, NLM
Virginia Meyer, PhD, NLM
Ilene Mizrachi, PhD, NCBI, NLM
Eric Nawrocki, PhD, NCBI, NLM
Valerie Schneider, PhD, NCBI, NLM

NLM Senior Investigator Receiving Review

Clement McDonald, MD, LHNCBC, NLM

1. Welcome, Introductions, Scheduling – Kevin Johnson

Dr. Johnson welcomed participants to the meeting. He noted that this would be his last meeting as BSC Chair, and that Dr. Salzberg will be assuming the role of Chair. NLM Deputy Director Jerry Sheehan thanked Dr. Johnson for his service, noting that he was the first Chair of the merged BSCs for the National Center for Biotechnology Information (NCBI) and the Lister Hill National Center for Biomedical Communications (LHNCBC).

2. Remarks from NLM Director – Patricia Flatley Brennan

Dr. Brennan thanked the BSC members for their service and Drs. Johnson and Salzberg for serving as BSC Chairs. She focused her presentation on three areas: the NLM Strategic Plan, COVID-related activities, and NIH's COVID-19 Extramural Survey Report.

NLM Strategic Plan

NLM issued a Request for Information (RFI) in August 2020 soliciting public comment to assist and guide NLM in identifying new, and updating ongoing, efforts to implement its 2017-2027 Strategic Plan.

COVID-19-Related Activities

Dr. Brennan highlighted a number of NLM efforts and projects that support the NIH response to COVID-19, including:

- A collaboration between an NLM intramural researcher and a grantee to provide a bioinformatics framework for wastewater-based surveillance of infectious diseases
- Coordinating data ingest and distribution from 48 NIH-funded RADx Radical (RADx-rad) projects
- Funding a study on prediction of diabetic retinopathy from risk factor data and understanding why people may not be getting regular services during the pandemic
- Participation by the NLM network of member libraries, community centers and health centers, in three key NIH projects:
 - The All of US initiative
 - The Community Engagement Alliance Against COVID (CEAL) project to expand access to testing and vaccines in underserved communities
 - The Post Acute Sequelae of COVID-19 (PASC) initiative
- Accelerating access to sequence information under the Accelerating COVID-19 Therapeutic Interventions and Vaccines (ACTIV) project

NIH COVID-19 Extramural Survey Report

Dr. Brennan described the results of this survey, which focused on the impact of COVID-19 on extramural institutions and individual researchers. Findings included that about half of researchers reported a negative impact from COVID-19 on their career trajectory, particularly laboratory-based researchers. Underrepresented groups reported varying impact, with Asian scientists most negatively affected. More than 66% of the researchers cited negative mental health impact, with women and other gender identities, as well as early career investigators, most affected. Overall, 78% of researchers reported lower productivity; issues with access to labs, facilities and colleagues were the strongest predictors of lower productivity.

3. Remarks from NLM Acting Scientific Director – Valerie Florance

Dr. Florance opened her presentation by recognizing the contributions and mentorship of Dr. Milton Corn, the former NLM Acting Scientific Director, who passed away earlier this year.

Dr. Florance described the merging of NLM's two intramural research programs – the National Center for Biotechnology Information's (NCBI) Computational Biology Branch (CBB) and Lister Hill National Center for Biomedical Communications' (LHNCBC) Computational Health Research Branch – into a single Intramural Research Program (IRP). The merged IRP has one Scientific Director and one Board of Scientific Counselors. She described the merger steps that have been completed as well as those that are underway, including recruitment for a Scientific Director.

Dr. Florance also outlined the IRP's basic areas of research and displayed the new NLM IRP website.

4. Presentation and Review of Clement McDonald, Senior Investigator

Dr. McDonald's research focuses on clinical data: how to capture it, standardize it, and analyze it. He explained how he got into this area of research and noted that he is only a part-time researcher, as much of his time is spent on developing tools, serving as NLM's Chief Health Data Standards Officer, and participating on 20 standards committees. Dr. McDonald presented an overview of his research since his last BSC review in 2011, as well as a brief description of his planned future directions.

Research studies (Dr. McDonald's focus during earlier years)

- Many studies relating to standards, including measures of how well they were applied
- In collaboration with FDA, CDC, and the IVD industry helped create the LIVD specification, which defines the USCSI codes into which lab instrument codes are converted for electronic health records (EHRs)
- Imaging studies, including first set of publicly available, de-identified radiology images annotated with findings and tied to narrative reports
- Studies of a de-identification tool
- Surveys of primary care physicians related to electronic medical records use
- Report of a low-cost mobile x-ray unit used in sub-Saharan Africa
- Real-time comparison of Surescripts medical history with patient-reported history
- General informatics studies

Medical/claims database studies (Dr. McDonald's current focus)

Dr. McDonald noted that he previously focused mostly on the MIMIC database, but since 2018 he has been focusing on Medicare data. Medicare data has many advantages: it provides near complete coverage of those over age 65, carries time course data, and contains information to answer some clinical questions (vital status, outpatient medications, diagnoses and procedures). Using Medicare's Virtual Research Data Center (VRDC) database, all analyses run

inside of Medicare's computers using Medicare-supplied software, and the patient-level data never leaves the VRDC servers.

MIMIC studies

- Conducted several studies using MIMIC database, the most interesting study showing a lower death rate for obese patients than normal weight patients after intensive care admissions
- Another MIMIC study showed that among insured patients admitted to the ICU, there was a lower rate of 30-day mortality in Black and Asian patients compared to White patients

Medicare studies

- Compared the effects of metformin on survival with that of 8 other diabetes medications, statins and 6 blood pressure medications in a Cox regression analysis, finding that 2 new diabetes medications and the statins showed a reduction in mortality risk
- Studied proton pump inhibitor association with mortality risk, finding no mortality risk when lag times were employed to protect against protopathic bias
- Compared tendon rupture among people taking fluoroquinolone antibiotics to that in people taking any of 3 of the most widely used non-fluoroquinolone antibiotics, finding that there did not appear to be a class effect
- Characterized possible benefits and adverse drug effects on SARS-CoV-2 infections with use of maintenance drugs such as clopidogrel, ACE inhibitors, ARBs and hydroxychloroquine
- Performed Cox regression analysis on a large sample of Medicare women taking estrogen, finding that the relative risk of breast cancer was decreased by 20% among those taking estrogen and that there was a 22% reduction in relative risk of all mortality

Future plans

Dr. McDonald is actively recruiting for another researcher and is seeking another database to focus on, possibly IBM's medical record database of 8 million patients. He is also exploring the use of combining deep learning with Cox regression.

Following Dr. McDonald's presentation there was a brief Q&A. Discussion topics included the actionability of his observational studies and how they could be used to inform clinical trials. Following the Q&A, the BSC went into closed session with Dr. McDonald.

5. Lister Hill National Center for Biomedical Communications Reorganization – Olivier Bodenreider

Dr. Bodenreider outlined the reorganization plan for the Lister Hill National Center for Biomedical Communications (LHNCBC, or LHC), which has been approved by NIH. Motivation for the reorganization included recommendations from the Blue Ribbon Panel in 2018 for NLM

to manage its intramural research program as one program with a single scientific director. In addition, it was felt that LHC activities lacked sufficient focus, that the branch names no longer reflected their activities, and that the research activities were not easy to distinguish from the R&D activities.

As part of the first phase of the reorganization, in 2019 two of the six branches of LHC were closed – the Office of High Performance Computing, and the Audiovisual Program Development Branch – and their activities were moved elsewhere. The Office of the Director branch continued with its same name, but it took on administrative functions that were previously part of the other branches, while the three remaining branches were renamed. The three newly named branches are:

- Computational Health Research Branch – This branch is a component of the NLM IRP and consists of Principle Investigators and their groups.
- Applied Clinical Informatics Branch – This branch is not a component of the IRP and includes staff scientists and others who are focusing on application development and applied research.
- Scientific Computing Branch – Also not part of the IRP, this group provides IT services and will provide support for scientific computing such as cloud computing.

Dr. Bodenreider noted LHC will be focusing on clinical data and that the planned structure parallels that of NCBI, with branches on research (NCBI's Computational Biology Branch), R&D (NCBI's Information Engineering Branch), and IT Services (NCBI's Information Resources Branch).

The LHC Computational Health Research Branch currently has three PIs and is recruiting for two others: one in image processing (interviews for this position are starting next week) and one in clinical analytics. In addition, Dr. Michael Chiang, who was recently appointed as director of NIH's National Eye Institute, will have his research program in LHC's Computational Health Research Branch.

The new LHC organization is operational, with staff reassigned to the new branches. Still outstanding are permanent selections for several positions currently filled on an acting basis, including the LHC Director and Chiefs for the LHC branches.

Discussion

In response to a question about whether the BSC would no longer be reviewing or providing advice about NCBI or LHC public resources, Dr. Brennan explained that NLM operations fall under the NLM Board of Regents and not the BSC and that NLM is trying to make sure it properly aligns the responsibilities of its federal advisory groups. However, she noted that NLM recognizes the BSC's interest in knowing about new activities at NLM, and, as in today's agenda, can include such discussions after adjournment of the official business portion of the meetings.

Asked about the interaction between the LHC research branch and the R&D/applications branch, Dr. Bodenreider said that the Applied Clinical Informatics Branch was envisioned as

being able to develop some of the insights gained from the research branch into products or prototypes.

6. NLM Training Program – Virginia Meyer

Dr. Meyer described the NLM intramural training program and how it fits into NIH's intramural training program. NLM currently has 43 postdoctoral fellows and is in the process of hiring six more. NLM also has one graduate student, five postbaccalaureate researchers (with two more coming), and 16 summer interns. Due to COVID, the number of summer interns is about half what it usually is.

Dr. Meyer described the various programs available for trainees to come to NLM, such as the Intramural Research Training Award, the Graduate Partnerships Program, and the Medical Research Scholars Program. She also explained the role of the NIH Office of Intramural Training and Education and described how NIH and NLM support trainees, including: recruitment, hiring, orientation, activities during the appointment, and exit and alumni follow-up.

Discussion

The BSC asked about the interaction between intramural trainees and extramurally funded trainees. Responding, Dr. Florance noted that there is some interaction at the annual NLM Informatics Training Conference that is hosted by an extramural training site each year. In addition, she noted that there is a program where university trainees can spend a summer at NLM. The BSC also asked about training for non-academic jobs, such as those in industry. Dr. Meyer responded that OITE has several programs related to non-academic jobs, such as the leadership training series.

7. Plans for Future Meetings – Kevin Johnson

Dr. Landsman said he will be emailing BSC members soon with prospective dates for meetings in April and October 2022.

8. Report to NLM Acting Scientific Director and DDIR Designee (Closed session)

Adjournment of Official Business

9. New Initiatives at NLM: SARS-CoV-2 Virus Variant Project and VADR – Ilene Mizrachi and Eric Nawrocki

Dr. Mizrachi described how NCBI moved to automated processing of SARS-CoV-2 genome sequences submitted to GenBank in order to efficiently deal with the increasing volume of the submissions, which now number tens of thousands per week. Working together with CBB researcher Eric Nawrocki, GenBank built on the work previously done to develop an automated pipeline to accept dengue and norovirus sequences into GenBank and to validate and annotate

the sequences. The new automated pipeline for SARS-CoV-2 sequences was operational by June 2020 and allowed non-problematic sequences to be released within 10 or 15 minutes of submission, compared to a day or two prior to implementation of the pipeline. GenBank also built a special submission wizard for the SARS-CoV-2 sequences that directed them to the pipeline and that prompted submitters for all required information, such as extra metadata.

Dr. Mizrachi noted that GenBank's participation in the CDC-led SARS-CoV-2 Sequencing for Public Health Emergency Response, Epidemiology and Surveillance (SPHERES) network has provided feedback on the submission system that resulted in improvements being made, as well as collaboration on issues such as metadata standards. She also described other collaborations, including work with the Public Health Alliance for Genomic Epidemiology on a BioSample submission of metadata, and current talks with CDC on environmental sampling sequencing and wastewater surveillance.

Dr. Mizrachi also briefed the BSC on NCBI's involvement in the NIH ACTIV-TRACE initiative. NCBI's Virus group is taking SARS-CoV-2 submissions to the Sequence Read Archive (SRA) and assembling the genomes, tabulating variations, assigning lineages to samples, and then preparing reports of the variations and lineages to share with NIH leadership, the CDC, and others. In the future, NCBI will also submit the assemblies to GenBank so that SRA submitters don't have to submit both the reads and the assemblies.

Dr. Nawrocki described VADR, the software package originally developed for dengue and norovirus sequences and then modified for validation and annotation of SARS-CoV-2 sequences submitted to GenBank. He explained how the software calls whether a sequence passes and can go automatically into GenBank or fails. He also explained how the software was modified to significantly speed up processing of sequences; the current version of VADR can process 100,000 sequences in 30 minutes with the 20 host computers that are used for processing.

10. New Initiatives at NLM: Research Organisms Data Resources (RODR) – Valerie Schneider

Dr. Schneider described the RODR project, one of four Tier 1 initiatives at NCBI's Information Engineering Branch (IEB). RODR aims to develop a consistently annotated, cloud-based, comparative genomics resource for all eukaryotic research organisms that integrates gene and organism knowledge and that provides a foundation for reliable comparative analysis. The project was initiated at the request of NIH, which had invited NLM to submit a proposal for supporting the needs of the research community for research organisms. The RODR proposal was approved by NIH in January.

Dr. Schneider began her presentation with background on the importance of research organisms for allowing scientists to understand biological processes as well as infectious diseases. She described some of the benefits that RODR will provide, including: a central portal for information on all eukaryotic organisms; shared public tools; the ability to work in the cloud; better infrastructure for large-scale analysis; meeting research needs for data to support data

science, population genomics, and comparative genomics; and active engagement with the research/user community to augment existing data collections with expert knowledge and to create interfaces and tools that best support the community's needs.

Dr. Schneider noted that the governance of RODR will involve two groups: an NIH oversight committee that will be responsible for monitoring the budget, milestones, progress and success metrics, and that will amplify communications to NIH stakeholders; and a working group of NLM's Board of Regents, which will evaluate the alignment of RODR goals with NLM goals and research community expectations, represent the needs of different communities, and assist with outreach.

Dr. Schneider explained that the RODR development program is expected to run for five years, but that because of the Agile development approach IEB will be using, the product will be delivered in increments over time. She briefly mentioned a number of the deliverables expected during the first year, such as additional publications available in Gene records, submitter-annotated assemblies displayed in the genome browser, a phase 1 version of a cloud-based tool for contamination screening, and a new unified web portal for genome-related data. Dr. Schneider closed her presentation by showing wire frame drawings of some of the content RODR will offer.



06/18/21

Dr. Kevin Johnson, Chair (Date)
Board of Scientific Counselors

Dr. Valerie Florance (Date)
Acting Scientific Director, NLM