

National Cancer Advisory Board (NCAB)
***ad hoc* Subcommittee on Population Science, Epidemiology, and Disparities**

September 1, 2020
1:00–2:00 p.m. EDT
Virtual Meeting

SUMMARY

Subcommittee Members

Dr. Electra D. Paskett, Chair
Dr. Francis Ali-Osman
Dr. Deborah Watkins Bruner
Mr. Lawrence O. Gostin

Dr. Margaret R. Spitz
Dr. Susan Thomas Vadaparampil
Dr. Deborah Winn, Executive Secretary

Other Participants

Dr. Anna D. Barker, NCAB
Dr. L. Michelle Bennett, National Cancer
Institute (NCI)
Dr. Christine Burgess, NCI
Dr. Robert T. Croyle, NCI
Dr. Howard Fingert, NCAB
Dr. Satish Gopal, NCI
Dr. Paulette S. Gray, NCI
Dr. Douglas Lowy, NCI

Dr. Diane Palmieri, NCI
Mr. Ricardo Rawle, NCI
Dr. Norman E. Sharpless, NCI
Dr. Dinah S. Singer, NCI
Dr. Sudha Sivaram, NCI
Dr. Emmanuel A. Taylor, NCI
Ms. Joy Wyszneuckas, NCI
Dr. Robyn Engel, The Scientific Consulting
Group, Inc., Rapporteur

Welcome and Opening Remarks

Drs. Electra Paskett, Director, Division of Cancer Prevention and Control, College of Medicine, The Ohio State University, and Norman Sharpless, Director, National Cancer Institute (NCI)

Dr. Electra Paskett, Subcommittee Chair, welcomed the participants, reminding them that the NCAB *ad hoc* Subcommittee on Population Science, Epidemiology, and Disparities (Subcommittee) is charged with advising the NCI Director and NCAB on strategic approaches and opportunities to enhance the NCI's contribution to population science, epidemiology, and diversity. The Subcommittee is responsible for identifying opportunities to address populations facing disparities through multidisciplinary programs, research surveillance, patient care, primary patient education, and cancer control, and Dr. Paskett noted that this meeting would focus on next steps to address that charge.

Dr. Norman Sharpless thanked the Subcommittee members for helping the NCI strengthen its portfolio related to cancer disparities.

NCI Report on Minority Health and Health Disparities

Dr. Emmanuel Taylor, Program Director, Center to Reduce Cancer Health Disparities (CRCHD), NCI

Dr. Emmanuel Taylor explained that he would update the Subcommittee on the congressionally mandated Minority Health and Health Disparities (MHHD) Report generated by the Office of Budget and Finance for fiscal year (FY) 2019. Although it is primarily a financial report, the MHHD Report also has

programmatic implications because the budget could be used to obtain the number of grants and contracts for subsequent analysis. The intent is to monitor resource equity and inclusion in federally funded programs of racial and ethnic minority populations and those in underserved areas. The National Institutes of Health (NIH) is mandated to conduct research to refine definitions in the MHHD Report. Dr. Taylor noted that lesbian, gay, bisexual, transgender and queer (LGBTQ) people are now included as a minority population. He suggested that Subcommittee members may benefit from reading the 1998 Institute of Medicine report on the *Unequal Burden of Cancer* and the NCI response to that report.

Funding for health disparities decreased from \$471 million in FY 2018 to approximately \$426 million in FY 2019. The reported amounts are prorated based on the number and composition of projects. Although overall MHHD funding decreased in FY 2019, funding increased for CRCHD, the Center for Cancer Research (CCR), the Division of Cancer Prevention (DCP), and the Division of Cancer Biology (DCB). CRCHD funding increased for health disparities but decreased for minority health. Approximately 80 percent of MHHD funding goes to minority health.

Funding for health disparities research decreased from 7.9 percent of the total budget of the NCI in FY 2018 to 7.1 percent of the total budget in FY 2019. The Division of Cancer Control and Population Sciences (DCCPS) receives the bulk of the funding, followed by CRCHD. Funding for clinical trials related to health disparities research also decreased from FY 2018 to FY 2019. The budget increased in basic research because of allocations to CCR. Funding for training also increased.

The NCI common scientific outline (CSO) code—unlike the Research, Condition, and Disease Categorization (RCDC) system used by the NIH—codes projects and activities so that they can be compared internationally or across funding sources.

The proposed effort to automate MHHD data collection and retrieval has two steps. First the definition of an MHHD project must be determined, then qualifying projects must be prorated. Proration is based on the data reported in the human subject system (HSS). Any imperfection in the CSO coding system is driven by imperfections in data quality. Data integrity should be protected during automation by linking into appropriate data sources such as the HSS. However, only those projects that are coded to be tracked are reported in the HSS, so subjects like training and outreach would not be included. Dr. Taylor emphasized that he supported in-depth analysis of the NCI portfolio.

Dr. Taylor dedicated his presentation to the actor Chadwick Boseman who recently died from colorectal cancer at age 43, noting that early diagnosis of prostate and colorectal cancer is extremely important.

Discussion

Dr. Francis Ali-Osman requested clarification about the allocation of funds to population science research and basic and clinical research. Dr. Taylor replied that the budget was skewed significantly toward clinical research because DCCPS codes all of its budgets in the clinical category which includes population and community-based research. CCR is increasing its efforts to investigate disparities, which is creating some balance; clinical research tends to be more expensive than basic research. If the budget is converted to the number of grants, then CSO codes could indicate distribution. Most funding is in population research and cancer awareness.

Dr. Paskett emphasized that Dr. Taylor was presenting dollars spent, and because different projects cost different amounts of money, it is important to look at the number of projects across sites and populations. Dr. Taylor added that some projects were not included in the report because their funding was below the cutoff level.

Introduction of Proposed Plan to Focus on a Population Group

Drs. Electra Paskett and Deborah Winn, Senior Advisor, DCP, NCI

Dr. Paskett introduced a proposed plan to develop techniques to analyze the NCI portfolio by focusing on a single population or group. The Subcommittee would then develop a charge for the formation of a new working group to continue developing analytic techniques to be applied across the NCI.

Dr. Deborah Winn commented on the many challenges in analyzing the complete NCI portfolio, such as limitations of CSO coding and limitations of the contents of the MHHD Report. DCCPS developed a coding system that covers the broad range of activities within that division. However, it could not be used for other divisions or outside of the NCI without significant modification. In addition, the DCCPS coding system is manual, which is not sustainable on a larger scale. The Subcommittee should consider approaches for portfolio analysis that describe what the NCI is doing, that improve identification of research gaps, and help assess progress in reducing inequities. Since it is difficult to define the many dimensions involved in health disparities, it would be easier to use a single population type to develop methods for deep portfolio analysis. That small test case could be used to develop computer models that could be adapted for use on the whole NCI portfolio. Deeper analysis of the NCI portfolio is needed, given the greater national interest in social justice issues and the NCI's commitment to addressing disparities in cancer.

Dr. Paskett stated that the goal is to develop a system that will analyze the breadth of NCI's MHHD research to identify areas where more attention would have a tangible effect on health disparities across populations and across the cancer control continuum, possibly by cancer site. She conceded that much of the work would have to be done manually if only current tools were used. Senior NCI leadership suggested the subgroup idea, and this subcommittee should decide if it is a good first step and, if so, what sort of population should be used for the subgroup. Choosing either a population group for which NCI has a significant research portfolio ("large group") or a group with a population group that has had limited research funding addressing it ("small group") both have benefits. The initial analysis would help to develop artificial intelligence (AI) and machine methods that could be utilized to assess all populations of interest.

Discussion

Dr. Winn suggested that African American or Hispanic populations could be a starting point, noting that the NCI is particularly concerned about cancer-related health disparities in these groups.

In response to concern expressed by Dr. Taylor about connections between the MHHD Report and the proposed analysis, Dr. Winn stated that this analysis will complement, but was not intended to replace the MHHD Report, which is Congressionally mandated and will continue to be produced.

Dr. Susan Vadaparampil asked whether race and ethnicity were being used to define the initial population. She suggested that concurrent analysis of one small group and one large group might yield better results. Dr. Paskett replied that race or ethnicity would be used to define the initial population and agreed that analyzing two groups may be better for AI training.

Dr. Margaret Spitz also agreed that analyzing two populations at different extremes (one with more research funding and one with less) was a good suggestion. In response to a further comment by Dr. Spitz, Dr. Winn clarified that a working group would undertake whatever functional statement the subcommittee creates; the Subcommittee would draft a charge for a newly created working group.

Exploring Approaches to Examine NCI's Cancer Health Disparities Portfolio

Drs. Michelle Bennett, Director, Center for Research Strategy, NCI, and Christine Burgess, Scientific Program Analyst, NCI

Dr. Michelle Bennett stated that machine-assisted methods may be able to help identify gaps in the complex area of health disparities research, but there are benefits and challenges associated with any technique. Portfolio analysis differs from the financial analysis included in the MHHD Report by giving more information about the different dimensions being funded and various health disparity types. Dr. Bennett introduced Dr. Christine Burgess as a subject-matter expert (SME) on portfolio analysis.

Dr. Burgess explained that the first step of portfolio analysis is to choose which research to include in the portfolio. Several methods are available for choosing which research is relevant. Manual identification can be performed by SMEs, machine learning or AI methods can be employed, or a hybrid approach that relies on both machine and manual methods can be applied. The final use case and the research definition of the portfolio will guide selection of the most appropriate method to use.

Machine-assisted methods convert a document into features that are used to classify documents within the algorithm. Many different methodologies fall under the category of machine learning; the thesaurus-based approach, for example, uses terms specific to certain areas of research to classify documents. Computers also can be used to establish business rules that will make manual classification easier. Some key information—including the project's title, abstract, and goals—is available to use as classification features.

Dr. Burgess then remarked on computer model development. A gold standard data set, composed of positive and negative examples from the documents of interest, must first be developed. This step requires significant input from SMEs. The gold standard data set then is split into a training set used for model development and a test set used to determine the strength of the model. Important performance measures to consider during model development include recall, the fraction of positives identified by the chosen classifier, and precision, or the fraction of true positives that are detected. No method will have both perfect recall and perfect precision, so the chosen method will prioritize one or the other.

Dr. Burgess summarized some benefits and challenges associated with machine-based methods. The benefits include consistency and speed once the method is developed. Machine-based methods work very well for clearly defined topics, such as cancer type. Machine-based methods do not work well, however, on nuanced topics or account for language variations. Model results depend on data quality and available classification features, and developing the model is time and resource intensive.

Dr. Bennett then discussed initial steps and considerations for the Subcommittee. First, the goal and desired outcome of the portfolio analysis must be identified. Second, analysis will require a large amount of input from a working group throughout the process, because the gold standard data set will require SME input. Third, the Subcommittee should determine the desired timeline and level of effort, because a one-time analysis would be designed differently than a long-term tracking effort. Finally, the Subcommittee must determine whether the MHHD Report data are sufficient for the goals of the analysis.

Discussion

Dr. Sharpless noted that clarification regarding the importance and limitations of the RCDC (Research, Condition, and Disease Categorization) data system may be needed. Dr. Burgess explained that RCDC is a categorization scheme that the NIH uses in the public space. Biomedical terms are added to a thesaurus function, and each grant is processed to identify those terms. The vocabulary of each grant is compared to a category's definition or fingerprint to see if the grant belongs in that category. The RCDC is a machine-assisted method, but SMEs are essential in developing the fingerprints. She added that RCDC does not

have all of the descriptors needed for this project, but which is one of the reasons to undertake this new approach.

In response to a question from Dr. Winn regarding the use of MHHD Report as a gold standard, Dr. Bennett clarified that she did not intend to conflate the two. Some of the research in the MHHD Report would be used, but expert involvement would be needed to identify the gold standard data set.

Dr. Taylor suggested that the minor categorization system (MCS) might be used as a starting point. He also clarified that the MHHD Report is not strictly financial. The NCI program directors code each project for the report, so the projects have already been identified as minority health research or health disparity research.

Summary and Next Steps

Dr. Paskett stated that the final task was to choose two initial population groups. NCI staff will present a portfolio analysis on the chosen group at the December 2020 Subcommittee meeting. The formal working group charge will be discussed after that presentation. The Subcommittee also will decide what areas of expertise are needed for the working group at the December meeting. Today, the Subcommittee should establish a focus for the pilot. Dr. Winn added that the portfolio analysis presented in December should not be expected to include everything, but it should represent examples of the full distribution of research in the chosen area.

Dr. Paskett asked the group which populations they would like to include. She proposed that African Americans be the larger group. Dr. Vadaparampil suggested Hispanics or Asian/Pacific Islanders would be good candidates for the smaller group. Dr. Spitz commented that Pacific Islanders or Native Americans may be good fit for the smaller group because Hispanics cannot be considered a small population.

Dr. Ali-Osman remarked that he liked the idea of a hybrid manual and machine method. He also proposed that Caucasians be used as a contrast to African Americans. Dr. Paskett replied that the focus should remain on population groups with health disparities. Dr. Sharpless stated that the discussion and presentations had been helpful and expressed concern that analyzing two populations may not be feasible in this time frame, reminding the group that portfolio analysis is a laborious process.

In response to an observation by Dr. Spitz that there were limited opportunities for research training, Dr. Sharpless suggested a presentation on this topic at a future Subcommittee meeting.

Adjournment

Dr. Paskett thanked the participants and adjourned the Subcommittee meeting at 2:00 p.m. EDT.

<u> /s/ </u>	<u> 9/2/2020 </u>	<u> /s/ </u>	<u> 9/3/2020 </u>
Dr. Electra Paskett	Date	Dr. Deborah Winn	Date
Chair		Executive Secretary	