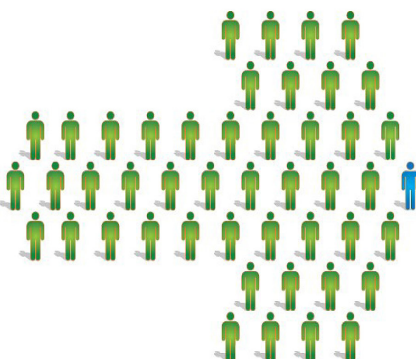National Cancer Institute

TRENDS IN 21ST CENTURY EPIDEMIOLOGY

From Scientific Discoveries to Population Health Impact

DECEMBER 12-13, 2012- NIH CAMPUS, BETHESDA, MD

WORKSHOP SUMMARY

Table of Contents

Abbreviations and Acronyms

AACR	American Association for Cancer Research
ACA	Affordable Care Act
BD2K	big data to knowledge
BMI	body mass index
CEBP	Cancer Epidemiology, Biomarkers & Prevention
CER	Comparative Effectiveness Research
CISNET	Cancer Intervention and Surveillance Modeling Network
CLIA	Clinical Laboratory Improvement Act
CMS	Centers for Medicare & Medicaid Services
CV	curriculum vitae
dbGaP	database of Genotypes and Phenotypes
DCCPS	Division of Cancer Control and Population Sciences
DCEG	Division of Cancer Epidemiology and Genetics
DNA	deoxyribonucleic acid
EGFR	epidermal growth factor receptor
EGRP	Epidemiology and Genomics Research Program
EHR	electronic health record
eMERGE	Electronic Medical Records and Genomics Network
EMR	electronic medical record
EPA	U.S. Environmental Protection Agency
FFPE	formalin-fixed, paraffin-embedded
GRIPS	Genetic Risk Prediction Studies
GWAS	genome-wide association study
HIPAA	Health Insurance Portability and Accountability Act
HapMap	Haplotype Map
HMO	health maintenance organization
HT	hormone therapy
iPOP	Integrative Personal "Omics" Profiling
IRB	Institutional Review Board
IT	informatics or information technology
MI	myocardial infarction
NCI	National Cancer Institute
NHLBI	National Heart Lung and Blood Institute

NIH	National Institutes of Health
PCORI	Patient-Centered Outcome Research Institute
PCR	polymerase chain reaction
PDQ	Physician Data Query
PLCO	Prostate, Lung, Colorectal and Ovarian
RCT	randomized controlled trial (or randomized comparative trial)
RNA	ribonucleic acid
STROBE	STrengthening the Reporting of OBservational studies in Epidemiology
TCGA	The Cancer Genome Atlas
VOI	Value-of-information

> "Yet like many other scientists, every now and then epidemiologists engage in boundary-making endeavors and disciplinary demarcation. And then, as usual in other disciplines, epidemiologists assert or reclaim contested epistemic authority and may claim jurisdiction over areas of public health, medicine, statistics, or science. These efforts evolved in the course of the twentieth century while epidemiology developed as a very diverse, eclectic—and foremost, *integrative*—field of practice and academic discipline. And so will they evolve as the societies of the twenty-first century continue developing. There is nothing wrong with that, it is the *natural* thing."
>
> Miquel Porta, January 2008, Barcelona
> In: Preface to "Dictionary of Epidemiology, Fifth Edition"
> Oxford University Press, 2008

Executive Summary

INTRODUCTION

The Epidemiology and Genomics Research Program (EGRP) of the National Cancer Institute (NCI) convened this workshop on December 12-13, 2012, with the purpose of bringing together thought leaders in epidemiology and adjacent fields to discuss future challenges and think creatively and provocatively about how epidemiology needs to evolve in a changing landscape. While this workshop played an important role in the initial process of drafting future research priorities, the organizers emphasized the ongoing thought process and need for future evolution and refinement of the ideas developed herein. A detailed summary of all presentations and discussions is provided in the main section of this report.

Appendix I provides the workshop agenda, and Appendix 2 provides the workshop participants list.

CRITICAL ISSUES

The participants discussed many facets of and current issues in epidemiology with regard to new methodologies, study designs, data sharing, the future role of cohorts, general scope, relevance for public health, and leadership. They scrutinized examples from past research to derive lessons for the future and made recommendations regarding priorities, best practices, and long-term strategies.

The following section contains a summary of the main critical issues discussed during the workshop, lively discussions of which were fostered by the participants' diverse backgrounds and areas of expertise. Additionally, ideas, questions, and challenges posed by the online community before and during the workshop were considered and discussed during the meeting.

New Methodologies

From genome to exposome. The recent success of genome-wide association studies (GWAS) for the identification of many new risk factors has created very high expectations for the field of epidemiology. The public expects epidemiologists to repeat this success in other areas and identify a wide variety of environmental and societal risk factors.

The participants called for cautious enthusiasm while keeping expectations at a realistic level. At the same time, with a plethora of newly available technologies at their fingertips, the thought leaders expressed a strong sense of responsibility to assume a leading role in this process. In the absence of foresight, proper infrastructure, and smart study designs, there is a considerable threat that the field might generate an unmanageable amount of random findings and damage its credibility. To avoid this, the participants sought to establish clear criteria to determine which technologies are fit for which purpose. Managing, understanding, and translating the expected deluge of data into public health efforts is expected to constitute a far bigger challenge than collecting the data.

Precision, accuracy, repeatability, and replication. The participants noted that these concepts are often used incorrectly, leading to confusion in the field. Genomics has recently enabled researchers to measure risk factors with unprecedented precision. Several participants acknowledged that less than a decade ago, few would have believed that robustly detecting relative risks of 1.1 and below would be possible. This great precision, however, does not mean that the results are necessarily accurate, because confounders might exist that lead to precise results of no relevance whatsoever for public health.

The participants agreed that it was unlikely that any other "–omic" technology would be able to achieve a similar level of precision in the foreseeable future, and that very large datasets would be necessary to overcome the considerable noise. To meet this challenge, new studies must be repeatable, which means that different investigators when studying the same datasets should obtain identical results. This requires much broader data sharing than is currently practiced. Finally, and very importantly, the results must be replicable. The participants discussed the importance of standardized datasets and honest brokers in this process and demanded a shift away from the current reward system based on publication volume. Instead of rewarding publication of studies whether or not they can be replicated, funders should create stronger incentives to generate replicable data. Participants suggested that data sharing and data replication plans should become standard elements of all new studies already at the design stage.

Incremental advances vs. disruptive technologies. The recent discovery that expression studies and other molecular analyses can be conducted on formalin-fixed, paraffin-embedded (FFPE) samples has led to a paradigm shift. It is now possible to revisit and obtain samples from many of the existing collections for new study designs. Participants noted the necessity for the field to remain flexible and able to swiftly adapt to these disruptive technological advances. For the most part, however, technologies will develop in incremental advances, and participants emphasized the need for researchers to constantly evaluate these technologies, assess their validity, and start early to weigh in on considerations regarding their cost-effectiveness.

Research settings vs. clinical applications. Methods that are not robust enough yet to find clinical application might still be very valuable in research settings. One great challenge in the field will be to determine the ideal time point for the transition into large-scale studies and clinical settings. The downside of the rapid development of new technologies is the fact that they can quickly become obsolete. Assays might be replaced by newer and more powerful techniques before their prices have dropped to a level that

would make them feasible for large-scale efforts. Issues of return of results to research participants also must be addressed.

Study Designs

Hypothesis-driven vs. hypothesis-free research. There was strong consensus among the participants that researchers historically have been over-optimistic regarding their ability to understand disease processes and identify candidate genes and exposures for study. Hypothesis-free research, however, requires large datasets to overcome the multiple testing problem, which automatically originates whenever scientists are allowed to conduct exploratory analyses of their datasets without any prior assumptions about the important factors.

Small vs. big studies. Given the required large sample sizes to measure small effects and overcome issues of multiple comparisons, there is no doubt that very large studies will be required in the future. There will, however, always be additional needs for smaller, targeted studies. New opportunities to access information from health care systems, social media, and other sources demand the ability to react quickly to unprecedented opportunities to study the effect of changes in society on public health. Participants also noted that most large-scale studies likely have to be tested in pilots first and that scalability should become an essential requirement for approval of any small dataset in the future.

Observational studies and randomized controlled trials (RCTs). Instead of seeing these designs as opposing entities, participants noted that great opportunities arise when the strengths of both designs are combined. Each design has its known weaknesses, and results must be interpreted with caution, as usual. In times of fiscal constraints, the combination of these methods will allow researchers to add smart new study designs to existing high-quality efforts and gain new insights at very low costs.

Epidemiology research and the fabric of society. Canada and the Scandinavian nations were mentioned as examples of countries that have made significant progress toward integrating research into their health care systems. The availability of large registries or online communities allows researchers to swiftly add studies to answer specific questions in an extremely cost-efficient manner. The participants discussed the feasibility of reaching a similar integration of research systems with electronic health records (EHRs), Census data, and other data sources in the United States and identified obstacles that must be overcome to realize this potential.

Data Sharing

Online vs. off-line data. The wide availability of the Internet and cell phones offers unprecedented prospects to reach individuals in their community with little effort and at manageable costs. Concerns that recruitment of individuals into studies and online questionnaires might lead to a strongly skewed distribution of participants with regard to social strata must be taken seriously. Early experiences with these methods suggest, however, that these problems might not be as grave as researchers have feared in the past. New generations of computer users are very open to this development, and seriously ill patients will devote a considerable amount of time online to take part in research and gain information about their condition. Interestingly, early studies also suggest that responses to online questionnaires might be as honest or even more honest than those provided during in-person interviews.

Privacy. All data-sharing plans and new study designs using online technologies raise important issues regarding privacy and data security. The workshop participants noted a currently ongoing change in society toward increased information sharing and reduced concerns about privacy. New technologies might, furthermore, allow researchers to analyze data remotely without ever having to download the data. Neither should analysis be restricted to statisticians, provided that a central resource can guard the data and provide user-friendly interfaces that allow researchers to obtain answers to their specific research questions.

Future Role of Cohorts

"The ideal" cohort. Designing a hypothetical ideal cohort can be a thought-provoking experiment. The use of mega-cohorts was discussed, but most participants focused on the importance of data harmonization between individual cohorts and the construction of large, "synthetic" cohorts across disease boundaries.

National and international cohorts. Data harmonization at the national level already constitutes a grand challenge. But national data provide only a limited window to the global heterogeneity of risk factors. Collaborations with other countries and international harmonization efforts will, therefore, be instrumental to integrating data from diverse populations. These efforts will enable researchers to conduct comprehensive analyses of risk factors across many different environments and health care systems.

Scope and Relevance for Public Health

From discovery to translation. The participants endorsed the idea that epidemiologists must become far more involved in translational efforts. At the same time, they should never feel pressured to move forward with findings that have not undergone sufficient validation yet.

Collaboration and education. Translational efforts will require the training of a new cadre of epidemiologists; participants suggested that this cadre contain representatives from all age groups and career stages. More extensive exchanges between public health schools and medical universities will be required in the future. Additionally, because of the extremely integrative nature of translational epidemiology, new investigators in the field also will need to be trained more extensively on issues regarding health policy, economics, and other adjacent fields.

Etiology, prevention, and treatment. There was very strong agreement among the participants that epidemiologists in the past have focused too much on etiology. New technologies, access to information, and better integration with the clinical care system will provide ample opportunities for epidemiologists in the future to address many additional issues, and will better equip them to support translational efforts and evaluate and improve interventions and treatments.

Leadership

Collaboration vs. career-management. A major barrier today to attracting the best and brightest young scientists into collaborative research efforts is the fact that publications as 20th author do not allow them to progress in their careers. The current system over-emphasizes competitive individual research and fails to promote highly collaborative research efforts. The participants offered several solutions, including

alternative reward systems and grant renewal mechanisms, but also practical means such as the introduction of annotated curriculum vitae (CVs) for career evaluation purposes.

Centralized vs. grass-roots/self-organization. One of the most controversial and difficult issues discussed by the participants was the right balance between self-organizing efforts and centralized leadership in future large-scale epidemiological efforts. Grass-root organization can be highly successful, but some participants expressed the need for centralized support to keep initiatives alive. Furthermore, increasing data collection efforts place increasingly larger demands on the infrastructure. Long-term and cross-disciplinary efforts therefore require substantial central support to remain viable.

Ending sub-par efforts. Many participants expressed a sense of urgency to formalize criteria to end unsuccessful or unreasonably expensive efforts. They also acknowledged that defining these criteria is a difficult task that has to be tackled by senior scientists who are no longer dependent on support by those individuals whose initiatives might be considered obsolete.

CONCLUSIONS

This workshop was designed to be one step in a larger process that involves an extensive interaction with other scientists via e-mail, Twitter, and a blog. At the end of the workshop, Dr. Patricia Hartge of the NCI summarized the priorities that had emerged from the online discussions prior to the workshop, which are presented in Appendix 3 of this report.

The workshop concluded with a poll among the on-site participants regarding their top-priorities for the coming 12 years. A summary of these priorities is included in the main part of this report.

The workshop organizers expressed their intent to continue the discussions online and to use the recommendations from the workshop participants for a first draft of a "12 in 12" list of priorities for epidemiology in the coming 12 years. This list will then be used to help NCI leadership, the research community at large, young talent in the field, and the participating scientists to identify the most urgent priorities for 21st century epidemiology.

# # #

Workshop Summary

INTRODUCTION

More than 10 years into the 21st century, we are at a major crossroads in our understanding of complex human disease. Tools of molecular biology, genomics, and other high-throughput "–omic" technologies are increasingly integrated into epidemiologic investigations. Along with these emerging tools come refined social, behavioral, and environmental exposure measurements at the individual, community, and health system levels and the ability to assess gene-gene and gene-environment interactions. There is an increased focus on complex "systems" approaches in understanding disease etiologies and intervening at multiple levels. All this has been influenced by tremendous advances in bioinformatics and information technology, allowing us to collect, analyze, and synthesize information from multiple disciplines at an ever increasing pace.

With these opportunities, however, comes the major challenge of dealing with the data deluge and uncovering true causal relationships from the millions of observations that are background noise. At the same time, increased consumer awareness and education have led to enhanced participation and co-ownership of research and research output. Thus, epidemiology now confronts important challenges and opportunities in the study of cancer and other diseases, and it must make choices of direction as it responds to rapid changes in the environment. To arrive at informed decisions, leaders must hear from different perspectives inside and outside of the cancer domain.

To help inform future research directions and funding opportunities, particularly in a climate of constrained resources, the NCI's Epidemiology and Genomics Research Program (EGRP) sponsored a workshop titled "Trends in 21st Century Epidemiology: From Scientific Discoveries to Population Health Impact" in Bethesda, Maryland, on December 12-13, 2012. Following plenary presentations, panels of experts with diverse perspectives offered brief assessments of the main challenges and most attractive opportunities. At the end of the meeting, workshop participants held an open discussion to help clarify and prioritize recommendations for enhancing the contribution of epidemiology in the next decade. The participants were chosen to represent a wide spectrum of experience ranging from population science to basic and clinical research and global health. The workshop agenda is included as Appendix 1, and a participants list is included as Appendix 2.

The workshop organizers made extensive use of traditional and new media to foster an ongoing and continuing discussion of the critical issues discussed during this 2-day workshop and to involve a broad and diverse audience in the discussions. Prior to the workshop, the EGRP invited the research community to comment on several relevant topics via the "Cancer Epidemiology Matters Blog," a platform that has been available online since June 2012.[1] These comments were used to shape the workshop discussion. Companion papers have been published in *Cancer Epidemiology, Biomarkers and Prevention* (CEBP) in connection with the workshop, with several more to follow. A

---

[1] http://blog-epi.grants.cancer.gov/.

commentary has been published in the July issue of CEBP to request input from the scientific community regarding the main foci for the future of epidemiological research.[2] Finally, the workshop was videocast live (and now archived) on the Web, and organizers encouraged online participants to assume an active role in the discussions via e-mails, blog posts, and tweets.[3]

In his welcome remarks, Dr. Robert T. Croyle, Director, Division of Cancer Control and Population Sciences (DCCPS), underscored the National Cancer Institute's (NCI's) commitment to encouraging a conversation across disciplines, institutes, and seniority in the field, as well as the view that the workshop should be considered part of a process rather than an isolated event.

As further context, Dr. Croyle informed the participants that during the past several months, the NCI has moved forward on a number of important large-scale efforts and signature projects, such as the multi-ethnic cohort. Many of these efforts are carried out collaboratively with other institutes and constitute global enterprises involving a wide array of funding mechanisms. He further emphasized the relevance of the December 2012 National Institutes of Health (NIH) advisory council meeting for the current discussions. During that meeting, NIH leaders announced the launch of the BD2K (big data to knowledge) initiative. This is a major commitment by the NIH in excess of $100 million per year with many different components, including the establishment of about 20 centers of excellence. This enhancement of infrastructure related to big data covers a wide range of activities from data security to scientific data and medical records and will be a key for all future trans-NIH efforts in the field of big data. More details about this initiative will become available during the coming months.

CHARGE TO PARTICIPANTS
Muin J. Khoury, M.D., Ph.D.
Epidemiology and Genomics Research Program (EGRP), DCCPS, NCI

In his introduction, Dr. Khoury reiterated that this meeting is very much part of a larger conversation. Despite its ambitious scope, the workshop can only provide a cross-sectional view of a much more complex and longitudinal discussion. Nevertheless, epidemiology will continue to represent an exciting research field in modern science and serve as an important driver of progress in medicine and public health. Although Dr. Khoury expressed confidence that epidemiology is here to stay, he also cautioned against ignoring the imminent difficult decisions that need to be made about prioritizing research questions, especially in times of fiscal constraint. The field simply cannot "do everything for everyone."

In a brief review of cancer epidemiology, which he based on a publication by Greenwald and Dunn (2009),[4] Dr. Khoury highlighted several successes of epidemiology during the

---

[2] Khoury, M.J., Freedman, A.N., Gillanders, E.M., Harvey, C.E., Kaefer, C., Reid, B.C., Rogers, S., Schully, S.D., Seminara, D., and Verma, M. (2012). Frontiers in cancer epidemiology: a challenge to the research community from the Epidemiology and Genomics Research Program at the National Cancer Institute. Cancer Epidemiol. Biomarkers Prev. 21, 999–1001.

[3] http://epi.grants.cancer.gov/workshops/century-trends/.

[4] Greenwald, P., and Dunn, B.K. (2009). Landmarks in the history of cancer epidemiology. Cancer Res. 69, 2151–2162.

20th century. He acknowledged the tremendous contributions of earlier epidemiologists, especially Dr. Joseph Fraumeni,[5] whose discovery of the Li-Fraumeni syndrome and other contributions played a substantial role in the early origins of cancer epidemiology.

The fields of genetics and genomics have recently assumed increasingly prominent roles in the field of epidemiology. The Human Genome Project and genome-wide association studies (GWAS) have shaped the field in recent years, and now individual sequencing and other technological advances offer unprecedented opportunities for molecular studies. It is, therefore, of great importance to think about the future of epidemiology.

Dr. Khoury then set the stage for the workshop by defining epidemiology as "the study of distribution and determinants of disease occurrence and outcomes in populations." He further explained that all epidemiological endeavors, albeit focusing on diverse outcomes, involving different risk factors, investigating different life stages, and using diverse methods in varying contexts, have in common the mandate to translate discoveries into population health.

The current meeting was designed to discuss epidemiology in the context of the translational research enterprise. Dr. Khoury reviewed the four drivers of epidemiology in the context of translational research: collaboration, technology, knowledge integration, and multilevel analysis.

These drivers fuel the translational life cycle, which consists of the following phases:

- T0: Discovery
- T1: Characterization
- T2: Evaluation
- T3: Implementation and Health Services
- T4: Outcome Research

Dr. Khoury emphasized the importance of including all efforts in epidemiology in the full cycle. The drivers shall provide the necessary ingredients to accelerate discoveries, which the field ought to translate into the later phases at a faster rate than currently accomplished. [6]

A review of EGRP-funded consortia and cohorts during the past two decades (1992-2011) showed that the number of funded consortia has increased almost 3-fold during the past 10 years, illustrating that team science has very much become a funding priority in epidemiology.

Dr. Khoury reviewed the concept of multilevel analysis based on a figure initially published by A. Barabási.[7] According to this model, interactions at the cellular, disease, and social level and interactions of factors between the different levels need to be taken into account to understand disease risk. To date, however, very few publications have attempted to carry out such multilevel analyses, with the exception of a few cautious efforts to test gene-environment interactions at the individual level.

---

[5] http://dceg.cancer.gov/about/staff-bios/fraumeni-joseph#biography.

[6] More details regarding this model can be found in a companion paper to the workshop titled "Drivers" of Translational Cancer Epidemiology in the 21st Century: Needs and Opportunities, by Lam, Spitz, Schully, and Khoury, forthcoming in CEBP.

[7] Barabási, A.-L. (2007). Network medicine—from obesity to the "diseasome." N. Engl. J. Med. 357, 404–407.

A review of the recent cancer literature further showed a strong increase in publications using diverse "–omics" techniques, especially in the sub-fields of methylation and micro-ribonucleic acid (micro-RNA). Aside from new molecular techniques, other methods such as the usage of accelerometers to objectively measure physical activity, have increased during the past decade.

On the topic of knowledge integration, Dr. Khoury noted that the past 5 years have seen a strong increase in publications using meta-analysis and systematic review techniques, while the classic, narrative review still dominates the literature. As part of the preparation for this workshop, Ioannidis et al. (2013) published a paper in CEBP that reviews the current landscape and future prospects of knowledge integration in cancer.[8]

In summary, epidemiologists find themselves today in a data-rich and technology-driven field. The vast amounts of available data in combination with publication bias harbor a significant risk for false-positive reports. Dr. Khoury argued that epidemiologists must make intelligent decisions today to avoid a future of epidemiology dominated by "incidentalomic" findings, i.e., random findings of no validity. They also must learn to identify the dead ends more quickly. If done right, the epidemiology of the 21st century will be called "translational epidemiology" and lead to interventions, new drugs, and better treatments.

Dr. Khoury reminded the participants that the major objective of the workshop and associated online and offline interactions is to draft 12 recommendations for action to influence the field of epidemiology within the next 12 years. He emphasized that discussions will continue after the workshop, again stressing the importance to consider this event a part of a much larger and ongoing dialogue.

SESSION 1: SETTING THE STAGE: THE EVOLUTION OF EPIDEMIOLOGY AND ITS APPLICATION TO CANCER
Moderator: Robert T. Croyle, Ph.D., DCCPS, NCI

Historical Perspectives on the Evolution of Cancer Epidemiology

Robert N. Hoover, M.D., Sc.D.
Division of Cancer Epidemiology and Genetics (DCEG), NCI

Dr. Hoover noted the importance of learning from the past to better prepare for the challenges of the future. He started his presentation with a review of the field of formal cancer epidemiology, the onset of which he assigned to the publication of two seminal publications in the early 1950s on the connection between smoking and lung cancer.[9] During the early years of the field, several other associations between cancers and carcinogens were identified. Notably, the relative disease risk for the exposed is larger

[8] Ioannidis, J.P.A., Schully, S.D., Lam, T.K., and Khoury, M.J. (2013). Knowledge integration in cancer: current landscape and future prospects. Cancer Epidemiol. Biomarkers Prev.
[9] Doll, R., and Hill, A.B. (1950). Smoking and carcinoma of the lung; preliminary report. Br Med J 2, 739–748; Wynder, E.L., and Graham, E.A. (1950). Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma; a study of 684 proved cases. J Am Med Assoc 143, 329–336.

than 4 for each of these findings, which means a considerable increase in risk for the individual.

Within the past few decades, epidemiology has been transformed from a "cottage industry" into a "big science" endeavor. The small research teams that used to collect manageable amounts of data from small groups of people using simple means have been replaced by very large and complex studies that are run by large and multidisciplinary teams. The need for depth in each discipline has led to tremendous specialization within these groups. Dr. Hoover then addressed why this transformation was necessary.

In the past, investigators set out to discover large risks by recording evident exposures. Their goal was to identify strong main effects, which they expected to show relative risks of well above 3 to be considered robust. Today, researchers have largely transitioned into the field of molecular epidemiology, where they face the challenge of measuring very small effects and exposures that are difficult to assess; they also must pay attention to interaction effects.

To face these challenges, molecular epidemiologists have recently been offered remarkable opportunities to overcome the weaknesses of the classical approaches. New tools have become available to measure exposures and outcomes, assess susceptibility, conduct mechanistic studies, and test large numbers of markers simultaneously.

The transition from the old to the new research paradigm did not occur in a structured and organized way, and Dr. Hoover used this failure as an argument to urge his audience to do better in the next big transition. As one example of unsuccessful research efforts in the past, he reminded the audience of the long-lasting failure of the field to determine whether or not hormone therapy (HT) for menopause increased cancer risk.

In 1971, in spite of fears based on anecdotal cases that HT might be cancerogenic, a systematic study found a protective effect of HT on risk for all cancers. About 5 years later, another study suggested that HT indeed might increase risk for breast cancer and that the increase was highest in those women with the longest duration of use. These two studies were followed by 20 years of what Dr. Hoover referred to as "a disaster." A very large number of studies were conducted to determine whether the treatment was harmful or protective. The results, however, were entirely inconclusive. This picture did not change until 1997, when a very large, collaborative study confirmed the carcinogenic effect.[10] Based on the large sample size, the authors of this study were further able to identify modifying variables, such as the time since last use and body mass index (BMI). Not having these data available until 20 years after the initial suspicions clearly indicates an unacceptable inability of the field to come to important conclusions swiftly.

A second important failed effort to learn from is the "lost decade" of candidate gene studies in the 1990s. During these early attempts to identify genetic risk factors, investigators with biological knowledge about disease processes would suggest candidate genes. Thousands of these genes were studied in small hypothesis-driven

---

[10] Collaborative Group on Hormonal Factors in Breast Cancer (1997). Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Lancet 350, 1047–1059.

studies, but only a tiny fraction of the reported associations ever replicated. By 2006, only six genes had been found with this methodology.

The completion of the human genome project and the annotation of variants in the Haplotype Map (HapMap) Project then caused a paradigm shift toward agnostic searches of susceptibility genes all across the genome. From 2006 onward, GWAS became the method of choice in genomics, leading to the identification of now about 265 risk loci that have been replicated. This number is expected to increase even further in the near future.

Dr. Hoover presented a summary of risk genes for breast cancer, all associated with small increases in risk for the individual (i.e., odds ratios of about 1.1). However, because all these loci are common in most analyzed populations, each of these variants is associated with a substantial population-attributable risk, which means that there is a considerable difference in the number of cases in the exposed compared to the unexposed population.

In spite of successes at the gene level, efforts to identify gene-environment interactions have not been successful so far. One example of inconclusive results is the association of cigarette smoking and acetylation phenotype (NAT2) with breast cancer. A recent, large, and well-conducted study could not replicate earlier results of a possible interaction effect.

Dr. Hoover summarized several lessons from these experiences and obstacles that must be overcome:

- Based on Dr. Hoover's review, there is no correlation between the quality of the underlying epidemiologic study and the robustness of derived genetic findings in case-control studies.
- Researchers must not overstate the importance of high-quality epidemiologic methods for successful identification of genetic risk factors in case-control designs, but they must ensure that these methods are applied whenever the environment is to be taken into account.
- The inconsistency encountered so far in gene-environment interaction studies, however, might very well be caused by insufficient quality of the underlying exposure data. The exact confounders, however, remain to be identified.
- Geneticists have provided the field with the ability to measure genetic markers with extremely high precision. This is not going to happen again for any of the other "–omics" techniques. Epidemiologists therefore must apply greater scrutiny when engaging in collaborations, and they must take an active part in assay development and validation to ensure that high-quality assays are used for all future "–omics" efforts.
- The best epidemiological methods and practices must be applied to minimize noise.
- Hypothesis-driven research has been far less successful than initially hoped. We do not know as much about disease processes as we think we do, and we must listen to the data instead in hypothesis-free approaches. Large sample sizes are required for this approach.
- Replication is absolutely vital to avoid reporting false-positive findings.

The overarching principle in these lessons for future research is "bigger, better, sooner." The field must become much faster and better at adapting new methods to meet

scientific needs and opportunities as they emerge. It is unacceptable that it took 20 years in the past to get to a straight answer regarding an important public health concern.

Dr. Hoover concluded his talk with several recommendations for overcoming current formidable, but surmountable, obstacles:

- Give appropriate "credit" for participating in team science and consortia efforts.
- Define roles for junior investigators.
- Consider the relative value and timing of individual versus pooled analyses.
- Take into account cultural differences between disciplines.
- Be ready to respond to rapid changes in state-of-the-art technologies.
- Ensure study subject participation, cooperation, and consent.
- Share data broadly and rapidly.
- Secure funding for the necessary infrastructure.
- Reform traditional grant mechanisms, which are inadequate for funding broad "discovery" efforts.

Panel and Audience Discussion

- What lessons and success stories have we learned from 20th century cancer epidemiology?

- What are the gaps and provocative questions that epidemiologic research can fill in the next 12 years?

Panelists

David Hunter, Sc.D., M.P.H., *Harvard University*
Timothy Rebbeck, Ph.D., *University of Pennsylvania*
Margaret R. Spitz, M.D., *Baylor College of Medicine*

Dr. David Hunter focused on the importance of large sample size. A well-designed big study will always be more valuable than a small study, when keeping all other variables constant. Today's accuracy in measuring small effects has only been made possible by very large sample sizes.

Using an example from the recent elections, Dr. Hunter illustrated the competitive nature of the field with regard to ownership of data. In this climate, it is very challenging to gather large enough samples to obtain representative and meaningful results.

It is desirable to end this unproductive fighting for control over data and co-authorship. Dr. Hunter suggested that establishing a U.S. national cohort would be one way to achieve large sample sizes in a structured collaboration. He argued that such a cohort actually already exists for cancer, due to the laudable efforts of the cancer epidemiology community working through the Cohort Consortium supported by the NCI.[11] For these efforts to be successful, there must be incentives for the investigators to share data from the start and to incorporate this expectation into their publication goals for the next grant renewal cycle.

---

[11] http://epi.grants.cancer.gov/Consortia/cohort.html.

Dr. Hunter then addressed the myth that it is impossible to measure exposures. The field already has established a number of best practices for such measurements, and future technological advances will help to further improve these efforts.

Another myth is that no diverse cohorts are available. Dr. Hunter noted that diverse cohorts are indeed available for study, but these have not always been included in all analyses in the past.

Dr. Hunter concluded his presentation by suggesting that the NCI convert the Cohort Consortium into the official U.S. national cancer cohort and swiftly initiate a similar process for case-control designs.

Dr. Timothy Rebbeck started his presentation by noting that, in addition to the nine papers already published in CEBP in the context of this workshop, seven more will be published before the April 2013 meeting of the American Association for Cancer Research (AACR).

He then focused on the challenges to conducting multilevel epidemiology. He reviewed the different levels that must be taken into account to understand complex cancer risk, which include biological and genetic pathways, individual risk factors such as age, education, and obesity, social relationships, neighborhoods, including access to health care, institutions, and social conditions and policies.

Although GWAS have been very successful recently, there have not been any larger attempts to connect these findings with the other levels. Dr. Rebbeck argued that one reason has been that the field of epidemiology has been too splintered to be successful in multilevel efforts. He then contended that there has to be a leader for the unification process, because the researchers in the field essentially lack the ability to organize themselves.

Dr. Rebbeck emphasized the need to include individuals' macro environment in epidemiological studies. He quoted from a recent publication from his group that used macroeconomic data to predict outcomes in prostate cancer and found strong associations between the macro environment, in this case the neighborhood in which the patient lived, and outcomes. The neighborhood might be a surrogate for access to health care, but this must be elucidated in further studies.

This finding is not directly linked to a biological pathway, and therefore might not be highly prioritized by some investigators with a strong focus on molecules. However, it might reveal important knowledge that will be relevant for public health decisions and thus earn its inclusion in large-scale and multilevel epidemiological analyses.

Referring to Dr. Hoover's urge to learn from the past, **Dr. Margaret Spitz** opened her presentation by acknowledging that she did not believe she has spent enough of her time and efforts on the translation of results, but rather has concentrated primarily on discovery efforts to better understand disease etiology. She then focused on two examples of currently neglected or provocative questions to illustrate her ideas for better and faster translation of discoveries into public health efforts.

In the area of smoking and lung cancer, Dr. Spitz noted the necessity to better understand genetic heritability of risk, to assess the functional consequences of already identified risk variants, to identify gene-environment interactions in general, and to determine the role of nicotine dependence in particular. Ultimately, these efforts should result in better and clinically valid risk predictions.

In the area of obesity as a risk factor for many cancers, although many genetic risk variants have been identified, too little effort has been devoted to understanding the underlying mechanisms. Furthermore, we do not know if the association between obesity and cancer actually will disappear for people who lose weight. The role of brown adipose tissue in obesity must be better understood, and new imaging techniques now provide unique opportunities to do so. Finally, how the colon microbiome affects body fat is an important question for future interdisciplinary research.

Dr. Spitz concluded from these examples the need for cohesive, interdisciplinary, and collaborative studies that apply new technologies in a field that she referred to as "integrative epidemiology." She further advocated the analysis of individuals with extreme phenotype (e.g., extremely thin vs. extremely obese, early onset vs. healthy), a strategy that is likely to enrich the study population for genetic risk variants. She then emphasized the need to take clinical relevance into account: reducing the mortality from lung cancer by 10 percent saves more lives, for example, than eliminating all gliomas. For future exposure measures, Dr. Spitz provided a brief review of many new technologies that have recently become available, as well as entirely new cohorts, such as individuals who have successfully lost weight and maintained their lower weight after bariatric surgery. In closing, she emphasized the urgent need to reshape the framework of epidemiology to realize the translational challenges she discussed.

General Panel Discussion

Tumor heterogeneity—have we been naïve to lump cancers together by physical location?

- Previous inconsistencies in genetic and epidemiologic studies might be caused by molecular heterogeneity, but Dr. Rebbeck noted that other types of non-biological heterogeneity have been neglected in the past and deserve attention. It is not necessary to understand the exact disease mechanism to issue a recommendation to stop smoking, and there might be other exposures that can be addressed in public health measures before the exact biological pathway has been elucidated.
- Dr. Hunter noted that relatively crude immunohistochemical classifications of tumors available for many years seem to align quite well with more recent expression profiles. Risk factors have been reported to be stronger in some of these groups than others, but the promise of personalized medicine and individual drugs appears to be over-stated at times. Dr. Hunter therefore urged the audience to maintain a realistic view of the number of different cancers that might become treatable with individual drugs in the future.
- Dr. Spitz replied that there are, however, several very promising examples of individualized treatments, such as never smokers with lung cancers and epidermal growth factor receptor (EGFR) mutations who respond to treatment with tyrosine kinase inhibitors. Even though these are remissions and not cures, these examples illustrate clear advances in understanding of individual tumor pathology.
- The panel further noted the importance of conducting prevention and treatment efforts in parallel.
- Dr. Hunter commented that efforts by The Cancer Genome Atlas (TCGA) project have so far focused on anonymized samples, with no available exposure

information. He voiced the hope that future demands on increased sample sizes will force such projects to obtain consent and gather individual data, which will then provide very valuable epidemiological data that can be analyzed in combination with the biological samples.

- Dr. Stephen Chanock (NCI) commented that there will be larger efforts to follow up on TCGA, including somatic DNA and epidemiological data. He cautioned the audience, however, not to expect too much from one single study; every study has its limits, and those limits need to be considered when interpreting results.

A "global cohort"

Dr. John Ioannidis (Stanford University) noted that nobody has a good grasp on the actual number of individuals available in already ongoing cohort studies. He estimated that this number might already exceed 100 million people worldwide and that many of these cohorts would be highly informative for urgent questions about etiology. He advocated carrying out a substantial effort to map the total availability of cohorts, for example by introducing new registries. This would render a more complete picture than the current efforts to encourage investigators to self-organize into larger consortia.

The "ideal" cohort

- Dr. Hunter noted that the ideal cohort does not exist and that too high demands on large cohorts in the past have, unfortunately, led to abandonment of the idea of establishing a national cohort: "the perfect became the enemy of the good." He then mentioned innovative ways to recruit a large number of people with the help of new technologies in mobile settings, so that people can be recruited quickly on a trip to their local shopping mall.
- Dr. Rebbeck supported the notion that "the perfect cohort" does not exist. Each cohort is a tool to answer a specific set of questions, and rather than being unrealistic about the "perfect study design," one should spend time and effort on designing "fit for purpose" studies of sufficient quality to yield the desired answers.
- Dr. Croyle provided several examples of successful ancillary studies that used existing samples to answer specific questions. He advocated doing this in a more systematic way across disease boundaries. Exposures that we think are important today might not be important tomorrow, and switching to a cohort that has been exposed might be faster and more cost-efficient than collecting a new one or trying to capture all exposures in a single cohort.
- Dr. Spitz agreed that in times of fiscal constraints, there is a great need to use existing cohorts in a more cost-effective way. Another participant highlighted the opportunity to integrate epidemiological research with clinical records, which can provide important access to exposure data.

The gap between discovery and translation

- Dr. Ioannidis noted the failure of the field in implementing public health measures in the field of smoking and cancer. He advocated for more implementation and political science research to be able to fix what we already know to be broken.
- Dr. Croyle commented that, during the past several years, health policy research and social context have gained importance at the NIH-wide level but are just barely touching the surface of what is possible in this area.

- Dr. Spitz noted that smoking cessation efforts are still "one size fits all." She used this example to emphasize the need to combine genetic findings with environmental exposures, because some of the genetic risk factors identified to date might increase addiction to smoking, while others might be directly carcinogenic. A better understanding of these processes will enable us to better tailor public health measures to the individual.

"Disruptive" technologies

Dr. Chanock noted the importance of occasional paradigm shifts for the epidemiology research field. One such shift is the recent realization by the TCGA project that genome and RNA sequencing can be done successfully from formulin-fixed, paraffin-embedded (FFPE) tissues. This has opened new opportunities to revisit banked samples to address questions about tumor subtypes. Therefore, periodically, the leaders in the field need to get together and think about examining extant resources and data based on these new technologies.

The scope of the field of epidemiology

- Dr. Robert Hiatt (University of California, San Francisco) asked the panelists how much responsibility epidemiologists have to get involved in other disciplines, such as education and politics, in order to drive necessary public health policy measures. Will epidemiologists in the 21st century just do "better epidemiology," or is there also a need to expand their role by interacting with adjacent disciplines? Epidemiologists have interacted successfully with geneticists to do molecular epidemiology, and there is no reason to believe that they cannot successfully interact with other, adjacent disciplines.
- Dr. Rebbeck commented on the importance to look beyond R01 grants.
- Dr. Khoury noted that most scientists focus on T0 (discovery) and possible T1 (characterization) epidemiology, as defined in his presentation. More researchers need to be incentivized to pursue translational efforts.

"Precision" vs. "validity"

Dr. Barnett Kramer (NCI) underscored the important distinction between the concepts of "precision" and "validity." Being able to measure effects very precisely does not mean that the results are necessarily valid, and in poor study designs, for example by ignoring the way the cases have been ascertained, there is a tremendous risk for confounding factors. Good epidemiological data are, therefore, still necessary, even when the genetic effects or other exposures can be measured with great precision.

Questions from the Online Audience

- Online participants wondered if advances in scientific methods will require advanced technologies in health care. Dr. Rebbeck replied that this is an important question that needs to be taken into account when implementing new health care measures, but that high cost is not necessarily a prohibitive factor if the new treatment makes a large impact for the individual.
- Dr. Hunter noted that in the recent past, better access to health care has been the strongest predictor of increased survival. He emphasized the need to balance prevention, screening, and treatment.
- In response to a question about decision making around funding in times of strong fiscal constraints, Dr. Rebbeck reiterated his pessimism regarding researchers' ability to self-organize, while Dr. Hoover noted that his experience

with the Cohort Consortium suggested that self-organization can be very effective and successful.

- The online audience also posed questions about best practices in longitudinal study designs, and Dr. Spitz offered [The National Children's Study](#) as an example of a longitudinal study that could be used for ancillary studies and regular sample collections.[12]
- A twitter question about the role of comparative effectiveness research (CER) for epidemiology was answered by Dr. Rebbeck, who thought that one should think about this the other way around: epidemiology should be at the center of all CER.

SESSION 2: THE IMPACT OF NEW METHODS AND TECHNOLOGIES ON EPIDEMIOLOGIC RESEARCH
Moderator: Stephen J. Chanock, M.D., DCEG, NCI

Technology-Driven Epidemiology: A Paradigm Shift

Geoffrey S. Ginsburg, M.D., Ph.D.
Institute for Genome Sciences and Policy, Duke University

Dr. Ginsburg used his presentation to provide an overview of new technologies that can be used by researchers and clinicians alike to enhance future epidemiological studies. He noted that in the spectrum from discovery to development and delivery, a shift to delivery must occur to fully realize the fruits of the new technologies. This implies education of a new transdisciplinary workforce.

Over the past 50 years, technology has changed significantly from observational to molecular and more recently toward genomic and digital science. New "–omics" technologies have provided us with a toolbox of high-dimensional data, including the DNA sequence and variants, gene expression profiles, the proteome, and the metabolome. There is now a great opportunity to combine old risk factors with new technologies to identify new risk factors and predict drug response and response to risk factors.

In addition to technological advances, social media and crowd-based efforts have grown dramatically over the past decade, including the establishment of several direct-to consumer genomics companies. Patients Like Me[13] is one example of a crowd-sourced platform that has provided new insights in disease etiology. These efforts will likely gain in importance in the future, particularly because the cost of genome sequencing has plummeted from $2.7 billion and 13 years in 2001 to $1,000 and one day in 2012. Another direct to consumer genomics company called 23andMe has efficiently replicated 180 previously reported genetic associations based on data from consumer volunteers.[14]

Cell phones have become ubiquitous. In underserved countries, text-messaging systems are now being used to adhere to and foster preventive health methods. Phones also can be used to simply remind people to take their medications.

---

[12] http://www.nationalchildrensstudy.gov.
[13] http://www.patientslikeme.com/.
[14] Tung, J.Y., Do, C.B., Hinds, D.A., Kiefer, A.K., Macpherson, J.M., Chowdry, A.B., Francke, U., Naughton, B.T., Mountain, J.L., Wojcicki, A., et al. (2011). Efficient replication of over 180 genetic associations with self-reported medical data. PLoS ONE 6, e23473.

Other participants already have pointed out the need for better integration of research efforts with health care systems. Dr. Ginsburg mentioned the Electronic Medical Records and Genomics Network (eMERGE) system as an example of an effort using high-quality biospecimens and genomic data linked to electronic medical records (EMRs) for discovery of associations. Several success stories have validated the eMERGE effort, recently summarized in a publication by the consortium.[15] He further noted the urgent need to complement the genome with an "exposome," as proposed in an editorial in CEBP.[16] Only a highly interdisciplinary effort can achieve this goal.

Dr. Ginsburg then demonstrated how new methods might allow us to develop new disease classifications and catalogs of exposures. In our daily lives, we are constantly bombarded by exposures, and the plethora of new technologies might help us understand better what these exposures are. He presented data from Michael Snyder's group,[17] which was able to mathematically quantitate exposures. Another study by Drs. Lawrence David and Eric Alm (personal communication) illustrated how dietary changes and disease cause substantial changes in individual microbiomic profiles. Dr. Ginsburg noted that these data are entirely descriptive at this point, and no inferences are made about cause and effect; yet they highlight the opportunity to explore different microbial communities in a dynamic fashion over longer periods of time.

Dr. Ginsburg then showed that lead[18] and radiation[19] exposures change blood RNA profiles in mouse models in a dose-dependent manner. The effects of radiation have been validated in humans in a radiation treatment cohort. These profiles are now being used by the U.S. government to triage individuals after exposure, for example during a nuclear disaster event.

Even smoking status[20] can be detected in gene expression profiles, and a model with four to five genes has been developed to reliably distinguish between smokers and non-smokers. Aerobic training[21] also left a detectable change in molecular profiles.

---

[15] Kho, A.N., Pacheco, J.A., Peissig, P.L., Rasmussen, L., Newton, K.M., Weston, N., Crane, P.K., Pathak, J., Chute, C.G., Bielinski, S.J., et al. (2011). Electronic medical records for genetic research: results of the eMERGE consortium. Sci Transl Med 3, 79re1.

[16] Wild, C.P. (2005). Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol. Biomarkers Prev. 14, 1847–1850.

[17] Chen, R., Mias, G.I., Li-Pook-Than, J., Jiang, L., Lam, H.Y.K., Chen, R., Miriami, E., Karczewski, K.J., Hariharan, M., Dewey, F.E., et al. (2012). Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell 148, 1293–1307.

[18] LaBreche, H.G., Meadows, S.K., Nevins, J.R., and Chute, J.P. (2011). Peripheral blood signatures of lead exposure. PLoS ONE 6, e23043.

[19] Dressman, H.K., Muramoto, G.G., Chao, N.J., Meadows, S., Marshall, D., Ginsburg, G.S., Nevins, J.R., and Chute, J.P. (2007). Gene expression signatures that predict radiation exposure in mice and humans. PLoS Med. 4, e106.

[20] Beineke, P., Fitch, K., Tao, H., Elashoff, M.R., Rosenberg, S., Kraus, W.E., and Wingrove, J.A. (2012). A whole blood gene expression-based signature for smoking status. BMC Med Genomics 5, 58.

[21] Timmons, J.A., Knudsen, S., Rankinen, T., Koch, L.G., Sarzynski, M., Jensen, T., Keller, P., Scheele, C., Vollaard, N.B.J., Nielsen, S., et al. (2010). Using molecular classification to

In his own research, Dr. Ginsburg has studied paradigms to carefully control for exposure in an experimental setting and is carrying out dense phenotyping pre- and post-exposure to identify molecular signatures. His group has concluded a study with eight individuals exposed to different viruses. When exposed to these viruses, not all individuals became sick; some got a little sick, while some appeared to remain entirely healthy.

The study provided a unique opportunity to study the course of disease from beginning to end. His group applied a very dense time sampling before and after exposure, and drew samples in intervals as close as every 6 hours. Multiple biospecimens were then profiled on multiple platforms.

Using the example of influenza exposure, Dr. Ginsburg presented a 50-gene model that could predict which individuals would show clinical symptoms.[22] In another recent study, his group has been able to show that expression changes that are easily detectible by quantitative polymerase chain reaction (PCR) precede the peak of clinical symptoms by as much as 50 hours.

In another study, Dr. Ginsburg's group has shown that molecular exposure profiles can be established not only for single viruses, but also for pathogen classes, and can, for example, be used to distinguish between bacterial and viral infections.[23] Studies from whole blood that can make this distinction have immediate treatment implications regarding the use of antibiotics.

Dr. Ginsburg's long-term goal is to establish a complete catalog of molecular classifiers of exposure for physiologic, pathogen, chemical, and pharmacologic exposures.

In the second part of his presentation, Dr. Ginsburg focused on the use of sensors for robust phenotyping. Usage of these sensors is increasing steadily. Movement measurers, for example, are being carried by an increasing number of people to measure daily physical activity. Gaming programs can be used to measure multiple traits such as cognitive function, mental health, and performance. Sensors have recently been developed to measure intra-ocular pressure on a contact lens for personalized dosage calculations. Very small, paper-based sensors are now being developed that can be integrated in simple tissues and would turn those into low-cost diagnostic devices.

All these new technologies generate a huge amount of new data, a substantial part of which is available through the Internet. Google, for example, is using this information to annotate flu trends, simply based on searches by users on flu-like symptoms.[24]

---

predict gains in maximal aerobic capacity following endurance exercise training in humans. J. Appl. Physiol. 108, 1487–1496.

[22] Chen, B., Chen, M., Paisley, J., Zaas, A., Woods, C., Ginsburg, G.S., Hero, A., 3rd, Lucas, J., Dunson, D., and Carin, L. (2010). Bayesian inference of the number of factors in gene-expression analysis: application to human virus challenge studies. BMC Bioinformatics 11, 552.

[23] Zaas, A.K., Chen, M., Varkey, J., Veldman, T., Hero, A.O., 3rd, Lucas, J., Huang, Y., Turner, R., Gilbert, A., Lambkin-Williams, R., et al. (2009). Gene expression signatures diagnose influenza and other symptomatic respiratory viral infections in humans. Cell Host Microbe 6, 207–217.

[24] http://www.google.org/flutrends/.

Dr. Ginsburg then posited that the greatest challenge of the 21st century is not the development of new technologies but the meaningful integration of the derived data.

He concluded the part of his presentation dealing with measurements, sensors, and data by showing a video that suggests that our future selves might check our health status as frequently as our e-mail.[25]

Dr. Ginsburg then summarized his notion of key questions for the field:

- What data are important? Are they reliable?
- How do we capture the data and when?
- How can we systematize exposure data?
- How do we ensure interoperability and standards?
- Which new mathematical methods and models need to be developed?
- How do we validate findings?
- How do we establish utility?
- How do we implement the results?
- How do we ensure that validation and utility come before implementation?

Questions and Answers to Dr. Ginsburg

Dr. Patricia Hartge (NCI) recalled earlier efforts in genomics, when the field suddenly realized that there was a strong sense of urgency: "If we do not do this right, someone will do it wrong, and this will lead to enormous confusion." She asked Dr. Ginsburg if he could see similar needs to push things forward quickly and avoid bad studies that can diminish trust. Dr. Ginsburg replied that high-quality epidemiology and specimens are essential for success with these new technologies. He considered it premature in the developmental stage of these technologies to commit to one platform versus another.

In response to Dr. Chanock's question about sampling time frames in cancer, Dr. Ginsburg stated that one of the group's priorities should be to find new ways to capture the temporal dimension, which still remains one of the great challenges. He wondered if, in the future, we might be able to predict tumors at a stage where they are not yet malignant. There will be a need to space out very many data-collecting events in yet to be determined intervals. Dr. Chanock followed up on the fact that prevention cannot be accomplished based on a snapshot alone. A more dynamic picture will be required to capture the different stages of cancer progression. Dr. Ginsburg suggested using either rapidly progressing cancers as a prototype or studying cellular models to get answers about these dynamic profiles.

Panel and Audience Discussion

- Which technologies do you feel are ready for "prime time" in epidemiologic research and for what purpose?

- What criteria would you use to determine when emerging technologies should be integrated into epidemiologic research?

Panelists
Zdenko Herceg, Ph.D., *International Agency for Research on Cancer*
Thomas A. Sellers, Ph.D., M.P.H., *Moffitt Cancer Center*

---

[25] http://gizmodo.com/5965143/holy-spock-the-star-trek-medical-tricorder-is-real-and-its-only-150.

Michael Snyder, Ph.D., *Stanford University*
Georgia D. Tourassi, Ph.D., *Oak Ridge National Laboratory*

Dr. Chanock introduced the panel by noting several important attitudinal changes with respect to the future of genomics in cancer that have been recognized by the NCI Director, Dr. Harold Varmus at an earlier meeting on the future of cancer genomics. These include a new eagerness to think about integration of genetic (germline and somatic), environmental, functional, and clinical data, and a shift of central concerns from the generation of data to data storage, access, use, integrity, and ownership. These key issues of data sharing and data availability will be important to consider when thinking about the 12 critical ideas during the second day of the current workshop. New, creative ways of data sharing with institutional and private players are required. Furthermore, issues of timing, recognition of moments of opportunity, better definition of the term "validity," and sharing of ideas and data must be resolved.

Dr. Zdenko Herceg summarized the technologies that he considered ready for prime time:

- Second-generation biomarkers of exposures, which he defines as markers derived from "-omics" and pathway-specific approaches.
- Epigenomics, including an understanding of a "normal" profile and dynamic changes over time (e.g., early life vs. aging).
- The "exposome" concept, which has been explained in the previous session. Personal and environmental monitors, geographic information systems, and more sophisticated questionnaires provide complementary approaches to gather these data. He noted that the European Union has recently launched a major project to study the exposome.[26]
- New bioinformatics tools and genomic databases provide the ability to integrate molecular data across different platforms and to develop comprehensive portraits of cancer subtypes. This will help to better understand etiology and discover opportunities for prevention measures.

Regarding the criteria to determine when new technologies are ready for epidemiological studies, Dr. Herceg listed the following:

- Measurements must be quantitative and sensitive. Dr. Herceg mentioned single cell "–omics" as a field that has yet to pass this threshold.
- New methods need to be compatible with currently employed high-throughput and genome-wide settings.
- The methods must be applicable to biobanks harboring materials for large prospective studies and population-based cohorts.
- Cost-effectiveness is an important determinant.

Dr. Thomas Sellers introduced the MyMoffitt "Total Cancer Care" patient portal, which is based on an Institutional Review Board (IRB)-approved protocol.[27] The protocol was approved in 2006, and the online portal was launched in 2009. All patients entering the hospital are asked to give permission to use their medical record for research. They are further asked to provide a biospecimen for research purposes and to allow the

---

[26] http://www.nature.com/news/daily-dose-of-toxics-to-be-tracked-1.11901.
[27] http://www.moffitt.org/takecharge/total-cancer-care.

researchers to follow their history and ask them for additional information and samples later during life.

The portal works well if the patients receive incentives to use it. Currently, these incentives include the ability to schedule appointments, pay bills, access their medical record, and join support groups with little effort. Furthermore, the portal provides a "smart" Web search function that is tailored to a patient's disease. The sites are pre-screened to ensure high information quality. The patients also can find clinical trials relevant for their situation through this portal.

The portal enables research on this captive population by providing the doctors with a cost-effective means to collect patient-provided data in a discrete format. All instruments are validated and standardized. The IRB has just approved online informed consent procedures that can be carried out in a very cost-effective manner. Copies of these consent sessions are available online for the patients at all times, so that they can log on and remind themselves exactly what they consented to, which they otherwise might forget over time.

The portal further provides a vehicle to administer follow-up surveys and gather data on patient-reported outcomes. Patient engagement with the portal ensures high participation.

Since the portal's launch, 29,000 accounts have been created and 12,000 monthly logins have been counted. As many as 84 percent of new patients create accounts. The team developing the portal was initially concerned that the online system would create a strong bias in the demographic, disadvantaging elder patients. Dr. Sellers reported, however, that these concerns have not become reality. There is no demographic bias with regard to patients who create an account compared to those who do not. Those individuals who cannot log on from their homes can use on-site machines for this purpose.

Dr. Sellers emphasized the importance of computer security for a system such as the patient portal. He also noted the importance of making the system accessible for a wide audience, both across age groups and computational platforms.

In his panel presentation, **Dr. Michael Snyder** noted the importance of longitudinal phenotyping. He then reviewed the Integrative Personal "Omics" Profiling (iPOP) study carried out by his lab using himself as the only research subject. During this study, his group carried out extensive phenotyping in close intervals. They measured information about his genome, epigenome, transcriptome, proteome, cytokines, autoantibody-ome, metabolome and microbiome. When Dr. Sellers was healthy, samples were taken every 2-3 months, but when he was sick (e.g., he had a cold), samples were taken every day for the initial phase and then every 4-7 days until the disease was over. Results from the 14 months of this research have been published.

By analyzing the resulting data and integrating the different datasets, Dr. Snyder then established profiles that reveal more information than any single technology would be able to achieve. Collecting all "–omics" information at once allows for the discovery of previously entirely unknown correlations, and having gathered all samples at the exact same time point is a crucial requirement for this strategy. The longitudinal profiles were important to understand the dynamic aspects of the diseases, and while the ideal sampling intervals remain to be determined, the often-employed sampling every 7

years might constitute too large intervals. Future plans for his lab include the inclusion of the exposome in the analyses.

Regarding the implementation of a study like this, the extensive efforts involved call for use of pilot studies, which can be scaled up into larger efforts if successful.

Dr. Georgia Tourassi then shared her view on the role of information technology for epidemiology and how information technologies can help epidemiological research bridge the gap between data and action. She had no doubt that advances in information technology are shaping the landscape of epidemiology.

Previous presentations have provided excellent examples of all the opportunities available today to collect new data of all different kinds. This trend is expected to continue, leading to vast amounts of data to be analyzed in the future. Dr. Tourassi then addressed whether today's information technology (IT) is ready to handle these vast amounts of data. In this process, she noted the importance of validating new information technologies before applying them broadly. She noted three current myths in the field:

- Myth 1: More data is always better.
- Myth 2: Information equals knowledge.
- Myth 3: Knowledge automatically leads to impactful actions.

Dr. Tourassi also indicated the need to establish common criteria for the evaluation of new methods for handling and analyzing data. This is a complex question, because each domain requires its own set of criteria.

Using the machine-learning field as an example, Dr. Tourassi summarized several lessons to be learned regarding the availability of open access training datasets for method validation. Benchmark datasets with very carefully curated data allow the developers to differentiate between errors in the data collection versus the analytical processes.

Cross-validation is equally important. Researchers must be able to transfer their tools from one dataset to another. Honest brokers must be available to guard the benchmark data and provide feedback to new developers regarding the performance of their new methods.

These measures will increase the quality of the analyses. To arrive at meaningful and, most importantly, actionable data will require additional steps. Dr. Tourassi emphasized the importance of conducting cost-effectiveness analyses early in the process to identify dead ends more quickly. Thus, even if a certain analysis might reveal a positive result, it might not cause enough of an impact to lead to clinical implementation and should be dropped from the analytical process at an early stage.

Dr. Tourassi concluded her presentation by recommending that funders such as the NIH invest in infrastructures that are capable, scalable, and sustainable. Technologies that are not scalable will not be able to provide sufficient payback for the investment.

General Panel Discussion

Dr. Ginsburg summarized central ideas from the panelists and noted that these are not unique to the NCI, requiring a broader discussion of shared infrastructures.

The transition from pilot to full-scale study
Dr. Snyder agreed with the speakers regarding the importance of gold standard data, but he noted that perfect methods do not exist and that it is therefore important to keep in mind the limitations of each method when using it.

Gold standard datasets
- According to Dr. Tourassi, the field needs to make a concerted effort to create gold standards and make them widely available. These data should be supplemented with published datasets from previously completed studies. Dr. Snyder mentioned that past consortia have done a very good job of developing gold standards. Dr. Tourassi noted the great promise that centralized data implies. Users can use applications tailored to their needs and technical understanding to perform a wide variety of analyses on these datasets. Naturally, such access requires strict procedures regarding access and security; the gatekeeper of these datasets remains to be determined. It is an important and urgent question because establishing the necessary infrastructure for the huge datasets of the future will require significant investments.
- The Database of Genotypes and Phenotypes (dbGaP) and the Gene Ontology Browser are minable data repositories that allow investigators to download data for new discoveries. Dr. Ginsburg urged journals to make deposition of data in public repositories an absolute requirement for publication. Dr. Chanock commented that several journals already have such policies in place, but they do not always have the resources for enforcement.

The transition from research to clinical practice
- Dr. Sellers noted that all efforts he described that are integrated in the patient portal protocol are ultimately meant to realize the promise of personalized medicine. A Clinical Laboratory Improvement Act (CLIA)-certified laboratory is available so that results can be returned to the patients and their treating physicians, who should not have to worry about which data are fit for what purpose. Important milestones to introduce markers in clinical practice are validation by external research publications and approval by internal pathologists, who need to have enough confidence in new assays to rely on them in their practice.
- Dr. Herceg acknowledged recent advances in cancer profiling by genomic methods but also noted current health disparities and the need for a more global approach to cancer treatment and prevention. While this does not mean that all cancers need to be sequenced in all countries, it does mean that assays developed in one setting must be validated in others before they can be applied in a global health setting.
- Dr. Ioannidis noted that the title of the session might be misleading, because often criteria for the transition from research to clinical practice are difficult to establish. At the research level, however, all methods can be considered "ready for prime time" now. Dr. Ginsburg responded that there have been occurrences of unfortunate blending of research and clinical practice. In some instances, patients were led to believe that they would receive results when this was not part of the research protocol. This transition must be better defined in the future.

Scalability of sample volumes
Although his self-experimentation used as much as 80ml of blood per draw, Dr. Snyder's lab is working on new methods that use less blood. The ultimate goal is to reduce the sample volume to drops of blood.

Respondent burden
With respect to the amount of time that patients are willing to dedicate to survey responses, Dr. Sellers noted that patients who are afflicted by a deadly disease such as cancer are almost invariably highly motivated to participate in these studies and will spend considerable amounts of time and efforts on getting their responses right.

Licensing the patient portal to other sites and sharing the data
Dr. Sellers explained that establishing the system at other sites would be possible but would require substantial investments in order to match the infrastructure at Moffitt. The protocol contains the requirement for a patient advisory board, which has been adamant about the fact that all data derived from this study be shared with other investigators as much as possible. Data sharing with a sponsor of the study from the pharmaceutical industry already has occurred.

Prioritization of research questions for the data generated by the patient portal
Committees review every suggested research question to ensure that studies use validated instruments and do not put excessive burden on the participating individual regarding the amount of required information or biospecimens.

Best practices regarding the types of data that cancer centers should collect
One participant noted that cancer centers have been struggling with their data collections due to the lack of unified standards. A modular approach that allows for different collection of data for different tumors would provide sites the flexibility to adjust for tumor peculiarities.

The cost of the exposome
Although costs for new technologies continue to decline, the demands keep increasing. There are therefore doubts about whether exposomes will become affordable for larger datasets in the foreseeable future.

Questions from the Online Audience

- Online participants inquired about the future of privacy if epidemiologists include online social media and other data in their studies. The panelists noted that societal attitudes regarding privacy issues are currently undergoing dramatic changes; the advent of social media has led to a new understanding of privacy; cancer patients are generally willing to share data as broadly as possible; and people who are very concerned about these issues should not agree to have their genome sequenced.
- A question from the Twitter feed centered on the requirements on the IT infrastructure to deal with the imminent tsunami of new data. Participants stressed the importance of collaborative efforts, new ways to integrate data, and the need for compromises regarding which parts of the data to keep in the long term. The scientific community is already planning not only for the required computational power, but also the downstream effect such as energy use. Collaborations with other scientists working in big data efforts will be helpful to separate the signal from the noise.

- In response to a question about incentives to create interoperable datasets, Dr. Ginsburg noted that the ability to talk to each other, exchange data, and achieve large sample numbers will be such incentives. The data do not necessarily have to be highly curated from the start, because downstream methods exist to extract information from unstructured datasets. Dr. Sellers added that his institute has invested hundreds of hours in the development of a data dictionary to ensure interoperability with other systems.
- Another online participant wanted to know the most pertinent research question that current technologies cannot yet address. Dr. Snyder's answer was "the exposome," while Dr. Herceg would prioritize single-cell genomics and tissue heterogeneity issues. Dr. Sellers provided a different perspective by suggesting that epidemiologists need to develop better ways to achieve buy-in of their findings by the general public, while Dr. Tourassi prioritized the development of infrastructures that allow the modeling of effects at the population level and include considerations regarding cost-effectiveness. Dr. Ginsburg emphasized the need to create pathways to validation that make the most use of already existing findings.

SESSION 3: THE EVOLUTION OF EPIDEMIOLOGIC COHORTS IN THE STUDY OF NATURAL HISTORY OF CANCER AND OTHER DISEASES
Moderator: Deborah M. Winn, Ph.D., DCCPS, NCI

What Have We Learned from Epidemiology Cohorts and Where Should We Be Going Next?

Julie Buring, Sc.D., M.S.
Harvard School of Public Health

Dr. Julie Buring presented lessons learned from epidemiology cohorts to date and implications for the future. She noted the timely occurrence of the current workshop after the recent 12-year anniversary of the NCI Cohort Consortium in October of 2012.

The consortium currently includes 46 cohorts spanning 15 countries, 4 million study participants, and 2 million DNA samples.

The goals of the consortium are to:

- Foster communication among cohort study investigators.
- Promote collaborative research on topics not easily addressed in single studies.
- Identify common challenges in cohort research and search for solutions.

During the recent annual meeting, the consortium members asked themselves a number of critical questions:

- What are our strengths and limitations to accomplish our mission?
- Should we expand our focus beyond the etiology of cancer during the next decade?
- Which gaps in knowledge are we best suited to address?
- What are the main remaining obstacles and how can we overcome these?

To answer the first question, the consortium needed to critically evaluate whether epidemiological studies would still be relevant in the future. The researchers concluded that future focus on complex interactions between genes and the environment, multilevel systems, networks and small increases in risk, suggests that epidemiological studies will increase rather than decrease in importance.

In this context, the unique strengths of cohort studies include:

- Availability of prospective data and large sample sizes.
- Multi-ethnic composition.
- Extensive serial phenotyping with repeated measures over time.
- Availability of samples stored in biobanks for access to genetic and biomarker information.

Dr. Buring next addressed whether cohorts should be expanded or extended in the future. When studying cancer, current gaps in knowledge suggest the need for detailed molecular characterizations of cancer subtypes. To do so, the consortium must assess its ability to obtain tumor tissue if not already collected.

The consortium also considered the merits of extending its scope beyond cancer etiology to include recurrence, second cancers, survivorship, and cancer treatment. There also is a need for a lifecourse perspective that includes children and adolescents in the cohorts. Further methodology should be incorporated in studies of the cohorts to validate, adapt, and extend assessments of exposures. Some of these, such as physical activity, can be assessed with quite simple means.

A very crucial component of all these considerations is the need to think about consent issues and IRB approval. Which processes are needed to allow researchers to revisit stored samples, and when do they have to re-contact study participants to obtain further approvals? Re-consenting tens of thousands of people at a time is not a trivial task.

Dr. Buring then posited that there is a great opportunity today to extend the cohorts beyond cancer to multiple disease endpoints. The mission of the study is not unique to cancer, and the consortium investigators believed that extending the cohort will add value while being cost-effective and achievable. Arguments for such an extension include the fact that many major risk factors for cancer are risk factors for multiple diseases. Furthermore, many cohorts are jointly funded, and multiple outcomes have been assessed with the same rigor as cancer. Finally, several cancer cohort members are already members of other non-cancer consortia.

A first step in the implementation of an extended scope beyond cancer might be a proof of principle study that would assess non-cancer outcomes, such as cardiovascular disease, in the current cohorts. If such an effort leads to feasible findings, the consortium could extend communication to other consortia, offering to contribute to solutions in other research areas.

Dr. Buring noted that there still are many obstacles to be overcome to achieve the expansion of the scope of the consortium; NIH assistance in this process is critical. Dr. Buring divided the anticipated obstacles into three categories:

- Structural
- Methodological
- Human resource related

With regard to the structural obstacles, Dr. Buring emphasized that sponsorship from multiple institutes will be needed; non-disease-specific funds will be critical for the applications to be successful in their respective study sections; and an integrated NIH management of the cohorts will be necessary. The primary concern of cohort leaders is securing funding for basic infrastructure to maintain data collection and blood

repositories, and to validate endpoints. Continuous support of the consortium is essential, and there is always a danger of underestimating the time and effort required to maintain basic functionality. The consortium cannot carry out unfunded activities and sometimes has had to decline exciting proposals for additional analysis simply because of a lack of resources.

Beyond maintaining basic functionality, the consortium needs support to incorporate new methodologies and technologies and to provide support for cross-cohort projects and data harmonization efforts.

With regard to the methodological obstacles, Dr. Buring emphasized the importance of the NIH acting as a liaison for cohorts to realize low-cost opportunities such as record linkage with EMRs. With the help of the NIH, these efforts can be carried out in a centralized manner. The consortium does not have the power or resources to drive these changes by itself and would become overwhelmed by the challenge of dealing with these issues state by state.

Dr. Buring acknowledged support from the NCI to date in efforts to harmonize data and noted several recent processes (e.g., the pilot of the National Virtual Cancer Registry) that would not have been possible if each individual investigator had had to rely on R01 grant funding. Thus, while the beginnings of the consortium were based on activities by individual investigators seeking collaborations, it would not have been able to survive without the expertise and funding provided by the NCI.

With regard to human resource obstacles, Dr. Buring noted the importance of addressing career development opportunities for young investigators. Consortia are problematic because promotion committees do not know how to recognize an individual's contribution to consortia activities. One concrete solution to overcoming this issue is the introduction of annotated CVs, in which the principal investigators are trained to specify the exact contribution of the younger investigators, who might have made a very critical contribution but are still found in the middle of a long list of co-authors. Similar issues are encountered during grant renewals, when consortium scientists have to establish special routines for writing their progress report. Setting up data sharing infrastructure is expensive and will require infrastructure grant support.

Dr. Buring then described her idea of the future "perfect" cohort study:

- It is jointly funded so that it can cross multidisciplinary lines to maximize impact.
- All data are or become harmonized.
- Methods are standardized to the extent possible while accommodating population-specific needs.
- Investigators have access to inexpensive common data sources to ascertain events and exposures.
- The researchers leverage innovative methods in the digital age.
- A reliable source of continued infrastructure funding is available.
- The investigators are allowed to focus on "better, faster, and cheaper" studies because the consortium can provide resources in a business-like manner.
- It provides flexibility and becomes a cornerstone for innovative research.
- Researchers think ahead when beginning observational studies; trials take into account from the onset considerations about future uses of samples and data and the use of ancillary studies.

Dr. Buring concluded her presentation by noting that synthetic and mega-cohorts do not preclude each other; each has an important role in future cohort efforts. She further stressed the importance of leveraging the existing cohorts while developing new ones to fill identified gaps. There is a strong sense of urgency to move forward, and it is already possible to use the cohorts from the consortium to establish a rich research portfolio on environmental, lifestyle, and genetic factors for cancer and other diseases.

Questions and Answers

Dr. Sellers inquired about the relative relevance of new discoveries versus confirmation of established or debated findings in the field. Dr. Buring responded that the goal to address questions that cannot be answered by individual studies is meant to focus on new challenges, rather than to better measure odds ratios of existing findings. Furthermore, the available large sample sizes will allow investigators to study cancer subtypes that they otherwise could not. This also holds for entirely new hypotheses, for which the consortium offers a large pool of potential participants who can be re-contacted for additional investigations.

Panel and Audience Discussion

- What developments are needed to make epidemiologic cohorts a cornerstone of the discovery to practice continuum?

- How should NCI and NIH facilitate multidisciplinary collaboration to integrate these developments into the research portfolio?

Panelists

Julie R. Palmer, Sc.D., M.P.H., *Boston University School of Public Health*
Lyle Palmer, Ph.D., *Ontario Institute for Cancer Research*
Daniela Seminara, Ph.D., M.P.H., *EGRP, DCCPS, NCI*

Dr. Julie Palmer began her panel presentation by noting that she very much agreed with Dr. Buring's priorities. She noted that about 10 years ago, cohorts usually only started with questionnaire data, and since then, collection of biological samples has been added to essentially every cohort effort. Thus, one advantage of cohort studies is the ability to expand them with additional methods as they become available. Dr. J. Palmer then offered possible improvements for existing cohorts, with the caveat that each cohort has unique goals and features and not all of her suggestions apply to every cohort:

- Online questionnaires
- Repeated measures
- Data from the Census Bureau, the Environmental Protection Agency (EPA), etc.
- Treatments, recurrence, and second cancers
- Tumor tissue samples and tissue microarrays

Dr. J. Palmer noted the importance of using external data sources to assess variables such as socioeconomic data (e.g., Census) and environmental exposures (e.g., EPA). She further emphasized the importance of including stress as a variable in as many studies as possible, because stress is increasingly recognized as a potential risk factor for a broad spectrum of diseases. Methods to validate self-reports of stress also ought to be established.

Dr. J. Palmer provided further support for Dr. Buring's notion that collaborative cohort consortia critically rely on central infrastructure reporting. She acknowledged the importance of current and future harmonization efforts of exposure data and noted that access to outside data such as cancer registries, death files, and Medicare claims records must be obtained. Investigators must further make greater efforts to collaborate with outside investigators and should develop best practices for data handling, storage, and analysis, which will make it easier for funders such as the NIH to estimate the necessary resources and make funding decisions.

Dr. Lyle Palmer opened his presentation by conducting a thought experiment: if the participants were given a budget of $50 million, what would they do? Using this hypothetical scenario, he developed several key features for new cohorts in the digital age. Many of these items have been implemented in the Ontario Health Study that Dr. Palmer supports:

- Data collection is performed online. Against popular belief, this strategy usually does not lead to a highly skewed demographic.
- There is responsive and constant evaluation of the target population.
- Follow-up occurs longitudinally over the entire life course. With an online design, these follow-ups are no longer prohibitively expensive.
- The study is large in scale.
- The study is inclusive: representative and diverse samples are collected.
- A comprehensive platform for disease and health research is established. This requires a shift in research culture: this is "our" not "my" research.
- The study is closely integrated with linked administrative health data.
- The study is closely integrated with government initiatives and clinical and public health networks. Etiology is important, but treatment response and other questions are also highly relevant and can only be addressed in data based on a large infrastructure.

Implementing a study like this in the United States would require a considerable change in culture away from the highly competitive individual funding mechanisms toward shared resources and infrastructures. Dr. L. Palmer was optimistic that these changes can be accomplished and advocated for establishing a similar system in the United States nationally.

He concluded his presentation by showing a screenshot of the Ontario Health Study platform, which provides a life-long and personalized home page for each participant.[28] Information presented on the page is stratified by patient data. The platform is also used to return results and provide specific advice. The website further allows for dynamic contact with all participants, who can provide data about their weight, diet, and other factors on a daily basis. It also enables online handling of consent, so that participants have an easily accessible record of exactly which studies are using their data, and allows the participants to compare their health status to other individuals in their region. In addition, the patients receive reports from studies that have been carried out based on their samples and data.

Dr. Seminara started her presentation by observing that the major new trend in epidemiology of the past decade has been the exponential growth of the number of

---

[28] https://www.ontariohealthstudy.ca/.

consortia. These collaborative efforts have increased in many research areas, but especially in the field of epidemiology. She then summarized lessons from the past that might guide future development of collaborative efforts.

The current Cohort Consortium has already stated its intent to grow beyond cancer etiology and include other diseases and endpoints. Dr. Seminara further emphasized the need to support an NIH-wide effort to create a large synthetic cohort. This cohort should be U.S.-based, receive funding primarily from the NIH, and study common diseases across the lifespan.

The next step would be to establish a worldwide cohort with population representations in each region and assessments of a wide range of exposures and diseases. Such a cohort would span diverse models of health care delivery.

To implement these new developments and cohorts, the following are needed:

- Cohort Consortium
  - Gather data across the lifespan.
  - Study multiple endpoints from etiology to survivorship.
  - Collect biospecimens prospectively and cross-sectionally.
  - Use innovative approaches for multilevel data collections.
  - Conduct research in diverse populations.
  - Rapidly validate and integrate new, high-throughput technologies.
- NIH synthetic cohort
  - Employ new, creative, and smart research designs.
  - Work on data harmonization issues in a systematic way.
  - Secure initial funding to establish feasibility of such efforts.
  - Organize think tanks to leverage the brainpower available within the collaborative approach.
  - Employ multilevel data integration and develop complex models for hypothesis generation and testing.
  - Make multilevel data available to investigators across the career span. Young investigators need access and mentorship regarding how to use these data.
- World-wide cohorts network
  - Establish cross-disciplinary research teams.
  - Create partnerships between academia, industry, government, and advocacy groups.
  - Integrate cohort research efforts seamlessly into the design of new randomized clinical trials.
  - Ensure rapid communication and dissemination of results to practitioners, policy-makers, patients, participants, health care organizations, and funding agencies.

Dr. Seminara concluded her presentation by summarizing the main future challenges as follows:

- Continuous yet flexible funding must be made available to support not only concrete research questions but also general infrastructure.
- Researchers must focus on unique research questions across the continuum of collaborative research efforts.

- Issues of multilevel data sharing and publications must be addressed. Current funding mechanisms are not suited to promote these efforts.
- The consortium will need help to manage complex governing structures and a substantial heterogeneity of applicable policies, especially at the international level.

General Panel Discussion

Dr. Deborah Winn remarked that the panelists have addressed all issues raised online and encouraged the onsite participants to engage in discussions.

Best practices for the transition from pilot studies to large-scale cohorts
- Dr. J. Palmer noted the importance of balancing speed and validity to avoid wasting substantial resources on research that does not work.
- Dr. L. Palmer noted that the Ontario Health Study employs expert panels to support these decisions.
- Drs. Seminara and Buring added that the Cohort Consortium would have the required expertise to provide the know-how required for such panels. The consortium is already divided into working groups. Traditionally, these have focused on individual research questions, but they could be re-focused to address other questions about new technologies from the discovery pipeline.

Opportunities and challenges of working with industry
Dr. L. Palmer noted successful collaborations with Google[29] and Facebook[30] in his country and did not see any major concerns for interactions with industry, as long as best practices are followed.

Incentivizing data sharing
How do we avoid publication bias, and how do we ensure that all data are available to all interested scientists? Should all funding be made contingent on the release of data?

- Dr. J. Palmer suggested tying funding to the release of data for new studies, but that this strategy would be difficult to pursue for existing data because of consent issues.
- Dr. Seminara noted the possibility to release the data by issuing "challenges." The data could be put into the public domain with associated challenges for interested investigators to work on.
- Dr. L. Palmer reported that his study has an open access policy and did not think that he would be able to get as much buy-in from the patients if they were not convinced that their data would be widely used to solve many research issues. In addition to sharing data, his study even shares the research platform for interested investigators to use.

Data sharing and similarities between epidemiological and genetic data

Although the Homer et al. study[31] was carried out using genetic data, the same concerns should apply to epidemiological data, so why should different standards for sharing and

---

[29] http://www.google.com/.
[30] https://www.facebook.com/.
[31] Homer, N., Szelinger, S., Redman, M., Duggan, D., Tembe, W., Muehling, J., Pearson, J.V., Stephan, D.A., Nelson, S.F., and Craig, D.W. (2008). Resolving individuals contributing

release of data apply? One member of the audience noted a data portal recently created by the Centers for Medicare & Medicaid Services (CMS) that investigators can access for their analyses without needing to download the raw data. As epidemiologists are moving to larger studies and links to EMRs, such systems might provide a means of preventing inappropriate release of personal information.

The feasibility of transitioning old cohorts into online communities on a large scale

Dr. J. Palmer considered the transition possible, as long as compliance with the Health Insurance Portability and Accountability Act (HIPAA) and other privacy concerns are taken seriously.

Inclusion of computer illiterate and socially disadvantaged groups
Dr. L. Palmer noted a slight over-representation of educated individuals in his data but that, overall, his data frequencies match those in Census data well. So far, he has not engaged in specific outreach efforts, but plans for such exist should evidence for skewed representations of the population become available.

More honest answers to sensitive questions through online surveys
In discussing whether participants are more or less truthful in online surveys, Dr. L. Palmer noted that online surveys actually yield more honest replies on sensitive issues such as drug use and sexual orientation compared to other survey modes.

Making scripts available for data cleanup that would be included with the data
Dr. L. Palmer noted that a harmonized data infrastructure should enable researchers to conduct research without any extensive data cleaning efforts.

---

trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. PLoS Genet. 4, e1000167.

Epidemiology and Evidence-Based Research along the Cancer Care Continuum

David F. Ransohoff, M.D.
University of North Carolina at Chapel Hill

Dr. Ransohoff posited that the best way for observational epidemiology to make the greatest scientific contribution is to cultivate observational cohorts. He reviewed the definition and importance of observational cohorts, and he shared examples and lessons from past observational cohort studies.

A cohort is a defined group followed over time. A cohort study design can address questions of diagnosis, prognosis, and treatment response, including measurement of biomarkers. The strength of the design depends on the absence of confounders (i.e., internal validity), the relevance of the research question (i.e., external validity), and the detail of measurements regarding the baseline state, exposure, and outcomes. A deficiency in any of these three domains can render a cohort study worthless.

"Observational" does not mean "passive." An observational study is more than "data+analysis." Each study requires a solid design, and the investigators should supply sufficient details about the methods in each publication to enable other researchers to make inferences about the study quality. They also should be able to provide a solid assessment of limitations and possible biases in the discussion section.

Although scientists might imagine an ideal design for a research study, in the real world they have to think about enhancing existing studies to reach a sufficiently strong design to answer the research questions at hand.

Dr. Ransohoff provided several examples of observational studies. For each of these, he assessed the nature of the inherent versus the added design, the level of effort it took to add to the design, and whether or not the overall design was strong enough to answer the research questions of interest. He then used these analyses to infer lessons for cultivating observational cohorts with strong research designs.

1. Paik et al. (2004) reported successful development of a prognostic assay to predict recurrence of a certain type of breast cancer. [32] The study was carried out as an addition to an existing randomized control trial and made use of the strong inherent design of the underlying study. At the time of the initial design of the study, RNA profiling from FFPE samples was not yet available, illustrating how new technologies can successfully be added to old studies.
2. Zhu et al. (2011) debunked earlier made strong claims, by others from the year 2002, regarding the identification of biomarkers in blood that would detect

---

[32] Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F.L., Walker, M.G., Watson, D., Park, T., et al. (2004). A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N. Engl. J. Med. 351, 2817–2826.

presence of ovarian cancer. [33] The study made use of samples collected under the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial.[34] Results indicated that none of the tested panels performed any better than the long-established CA125 marker. This study shows how utilization of blood samples that were taken just before a diagnosis was made can be used to design new studies on the identification of predictive biomarkers. Dr. Ransohoff noted that not many studies are available that collect blood samples in this longitudinal manner.

3. Imperiale et al. (2004) established that screening for abnormal fecal DNA provides a more sensitive test to detect colorectal cancer in average-risk populations than does the clinical practice of screening for fecal occult blood.[35] The study was successful because it was added onto a sound prospective study design. Unfortunately, samples from these efforts were not shared with the research community because of industry sponsorship, and the screening was very expensive.

4. Selby et al. (1992) determined that colonoscopy screening for colorectal cancer reduced mortality by about 60 percent.[36] This study used a cohort begun in the 1970s. The researchers added a case-control design, obtained information about the cause of death and whether the colonoscopy was carried out for screening or diagnostic purposes, and created an internal control sample. The study design was strong, and the results led to an almost immediate change in clinical practice, although they were derived from an observational study and not from a randomized trial.

Dr. Ransohoff then encouraged his audience to be opportunistic in the quest to cultivate additional observational studies. Existing randomized control studies and cohorts (e.g., Framingham Nurses Health Study) as well as clinical practice settings (e.g., health maintenance organizations [HMOs], pharmaceutical companies) provide possible bases for these cohorts.

Although the examples provided above are likely not new to this group, the emphasis today is on learning from these examples for the future. Dr. Ransohoff therefore concluded his presentation by reviewing opportunities, identifying challenges, and making several recommendations based on these examples. He urged his audience to embrace big efforts, to constantly look for opportunities to piggyback studies onto existing infrastructure, and to be aware of the opportunities that nested case-control designs can offer. His final recommendations were as follows:

---

[33] Zhu, C.S., Pinsky, P.F., Cramer, D.W., Ransohoff, D.F., Hartge, P., Pfeiffer, R.M., Urban, N., Mor, G., Bast, R.C., Jr, Moore, L.E., et al. (2011). A framework for evaluating biomarkers for early detection: validation of biomarker panels for ovarian cancer. Cancer Prev Res (Phila) 4, 375–383.

[34] http://prevention.cancer.gov/plco.

[35] Imperiale, T.F., Ransohoff, D.F., Itzkowitz, S.H., Turnbull, B.A., and Ross, M.E. (2004). Fecal DNA versus fecal occult blood for colorectal-cancer screening in an average-risk population. N. Engl. J. Med. 351, 2704–2714.

[36] Selby, J.V., Friedman, G.D., Quesenberry, C.P., Jr, and Weiss, N.S. (1992). A case-control study of screening sigmoidoscopy and mortality from colorectal cancer. N. Engl. J. Med. 326, 653–657.

- Do not solely focus on data and specimen collections, but continuously be aware of the methods and the designs that are necessary to answer relevant research questions.
- Explore many different approaches, collect preliminary data, and then scale up into larger efforts.
- To realize these opportunities, epidemiologists will have to overcome substantial obstacles with regard to supervision, logistics, and motivation.

Questions and Answers

Dr. Snyder observed that prospective collection of samples is problematic when it is not yet known how samples need to be taken and stored in order to allow for analyses with new technologies. Dr. Ransohoff agreed, but he also noted that researchers are too often not even thinking about preserving enough samples for later studies and, in standard cases such as blood collection, can make a better effort to collect for future studies.

Panel and Audience Discussion

- How do we fill evidence gaps in care and prevention using epidemiology?
- What is the role of observational epidemiology in a "data rich" environment where randomized clinical trials may not always be feasible?

Panelists

Barnett Kramer, M.D., M.P.H., *Division of Cancer Prevention, NCI*
Michael Lauer, M.D., *Division of Cardiovascular Sciences, National Heart Lung and Blood Institute (NHLBI)*
Jeffrey A. Meyerhardt, M.D., M.P.H., *Harvard Medical School*
Olufunmilayo I. Olopade, M.D. F.A.C.P., *University of Chicago*

Dr. Barnett Kramer's professional focus is on the effectiveness of interventions. A key issue in this field is that the stakes are very high when designing interventions: we do not want to risk making healthy people sick. Therefore, this field is forced to look for high levels of evidence.

Most questions in prevention research cannot be answered in randomized trials. Therefore, researchers often must rely on observational evidence. Observational studies, however, must be interpreted with caution.

In his work for the Physician Data Query (PDQ) Adult Treatment Editorial Board,[37] Dr. Kramer makes frequent use of epidemiological data. When making recommendations, he first ranks the available studies by design strength. He then assesses the internal validity of the study and the consistency and volume of evidence: Has evidence been presented in one or multiple studies? Is it derived from small or large studies, and are the directions of outcomes consistent? He then takes into account the magnitude of the effect size and evaluates external validity.

Researchers are too often misled by flawed evidence, which might become a barrier to new discoveries. They must be very certain about an effect before claiming that "the evidence is in" and should make a conscious effort not to dismiss new trials too easily based on data of questionable quality.

---

[37] http://www.cancer.gov/cancertopics/pdq/adult-treatment-board.

Dr. Michael Lauer presented his vision of the future of epidemiology. He described the approach in Scandinavian countries, where every single case of myocardial infarction (MI) is added to a registry. Between 2010 and 2012, about 10,000 patients have undergone a stenting procedure and been added to the database. About half of these patients have then been randomized into a trial. This trial, which is embedded into a registry, has been termed a "clinical registry trial." The cost of the trial is only about $50 per patient.

Dr. Lauer considered this example a good illustration of how observational studies and randomized trials should not be considered opposites: randomized trials can successfully be embedded into prospective studies. Because of the impressive sample size, results from this study will be credible and actionable.

Dr. Lauer stressed the importance for epidemiology to move further on this path and start to consider observational studies as platforms in which clinical trials can take place in a cost-efficient manner. Similar trends can now be observed in other parts of the world. Those who do not embrace this new way of conducting epidemiology put themselves at risk of becoming obsolete.

Dr. Jeffrey Meyerhardt focused his presentation on the integration of epidemiology into randomized clinical trials. He described the basic study protocols used in these studies and noted recent additions to these protocols, such as collection of biospecimens and, more recently, epidemiological data. These data include a multitude of measurements and records of comorbidities, complementary medications, adiposity, physical activity, smoking status, etc. The extended protocol is currently being used in three clinical trials. The data are obtained from the treating hospital and not directly from the patient, which distinguishes these trials from other observational cohorts.

Adding questionnaires as requirements for clinical trials from the onset is problematic, because these efforts should not interfere with accruals. When optional, about 70 percent of the patients have so far opted to complete epidemiological questionnaires before entering the treatment provided during the trial.

Dr. Olufunmilayo Olopade addressed the knowledge disparity gap in her presentation and offered possible solutions to close it.

She showed results from a study of population differences in breast cancer, which revealed that triple negative breast cancer is overrepresented in indigenous African women.[38] The classifications used in this study were not available when the current cadre of physicians was trained, and Dr. Olopade challenged the next generation of epidemiologists to think about what the future of epidemiology will look like and how it is going to be different from today.

She contended that cancer patients and the population at large are not interested in hearing cancer statistics. They care more about their own cancer and whether it is treatable. Principal investigators leading consortia, on the other hand, are very much focused on large sample sizes and might not pay enough attention to the individual. How can these two worlds be united?

---

[38] Huo, D., Ikpatt, F., Khramtsov, A., Dangou, J.-M., Nanda, R., Dignam, J., Zhang, B., Grushko, T., Zhang, C., Oluwasola, O., et al. (2009). Population differences in breast cancer: survey in indigenous African women reveals over-representation of triple-negative breast cancer. J. Clin. Oncol. 27, 4515–4521.

Dr. Olopade posited that more observational epidemiologists should work in clinical settings. A better understanding of molecular pathology provides tremendous opportunities for better treatments. She encouraged the audience to think globally, pay greater attention to diversity in the United States and elsewhere, and put the patients at the center of all translational efforts.

General Panel Discussion

The traditional separation in training of epidemiologists (schools of public health) and physicians (medical schools)

- Dr. Ransohoff commented that the division in training has contributed to the divide between research on etiology versus treatment and other outcomes.
- Dr. Kramer, who has experience from both sides, agreed that there is indeed a huge divide between the disciplines and that public health knowledge is of such great relevance for the clinical disciplines that more of it should be taught there, and vice versa. The divide often causes animosities and misunderstandings.
- Although Dr. Lauer observed a softening of boundaries between the two fields, Dr. Meyerhardt noted that a lot of work remains even today to integrate knowledge from one discipline into the other and that the formal courses that exist today might not always be successful in providing true cross-training.
- Dr. Olopade emphasized the need for more interdisciplinary training to meet the challenges of the future.

The role of the NCI and other drivers in integrating cohort studies with randomized trials

- Dr. Lauer observed that incentives must be created for researchers to interact in new and creative ways. Expensive trials should be replaced by new and more cost-effective methodologies where available.
- Dr. Olopade pointed to opportunities that public-private partnerships and empowered communities can offer to conduct research on a larger scale for less money. Epidemiologists must leverage diverse resources to accelerate progress.
- Dr. Ransohoff cautioned that researchers must ask relevant questions and not get entirely caught up in collecting data.
- Dr. Meyerhardt emphasized the need to integrate EHRs with research data. Harmonization of diverse systems is an important priority.
- Dr. Spitz acknowledged the historic role of the NCI in establishing epidemiology and enabling the establishment of infrastructure, especially during times when epidemiologists were challenged by other disciplines.
- Dr. Hiatt added that epidemiologists today head only a few clinical cancer centers and that these centers are the places where the interactions between medicine and public health should occur. There should not be any separation between clinical and public health questions.
- A workshop participant noted the importance of health economics for answering these questions and collaborating with other countries that already have established a linkage between health care and research data systems. Dr. Lauer remarked that the field of health economics already has influenced randomized trial designs. It was noted as well that the NIH might benefit from applying randomized trial designs to its own methods of operation.

Alternative sources for cohorts

With the introduction of the Affordable Care Act (ACA), many more individuals can gain access to health care. The massive entry of new people into the system provides a tremendous opportunity for studies on social determinants of health and disease. Dr. Kramer further developed this thought by outlining several new opportunities that changes such as the ACA offer to address questions about disparities. These studies must be conducted quickly or the opportunity will be lost.

Criteria for declaring studies or designs obsolete

- Dr. Lauer indicated that times of fiscal constraints actually offer an opportunity to shed studies that are very expensive and replace them with modern and more cost-effective efforts. Dr. Ransohoff echoed the need to withdraw incentives to weed out weak research.
- Dr. Olopade added that on a large scale, all methods become expensive, and researchers need to consider the scalability of research efforts and the resulting costs.
- "Ossifications" in the system sometimes lead to continuation of methods of the past. Dr. Kramer commented that while innovative methods are generally popular, it might be difficult for younger researchers to dispense with older technologies, because investigators who determine future funding might still be attached to these.
- To learn how to better deal with changes, Dr. Lauer suggested learning from individuals in "fringe" markets and organizations that have successfully implemented disruptive technologies. Central funding, for example provided by the NIH Common Fund, was seen as essential to incubate new and potentially disruptive technologies to the point where their success or failure can become evident.

Questions from the Online Audience

Must all epidemiology research be translatable, and whose job is it to carry out the translation?

- Dr. Olopade highlighted the need to diversify the portfolio, include more translational efforts, and accomplish a balance between basic science and translational efforts.
- Dr. Kramer commented that policy and intervention are not the only endpoints into which epidemiology should be translated. It also should be translated into numerous additional research questions. Epidemiology that does not get translated in either direction is not useful.
- Dr. Meyerhardt cautioned that some research findings might become translatable in the future and that, in these scenarios, researchers and public health policy developers must be patient and resist pressure from the media to translate findings prematurely.

The Role of Epidemiology in Knowledge Integration and Meta-Research

John Ioannidis, M.D., D.Sc.
Stanford Prevention Research Center

Knowledge integration consists of knowledge management, synthesis, and translation. It constitutes the challenge to integrate a plethora of information into simple answers, such as general health recommendations. To accomplish this goal, a very large number of publications must be considered. The total number of publications mentioning cancer has now exceeded 2.6 million. In addition to the published data, unpublished results also must be included to provide a comprehensive picture.

Dr. Ioannidis noted the importance of data sharing and full disclosure of all analyses that were performed in a study. He then reviewed different methods for knowledge integration such as meta-analyses, field synopses, and multilevel evidence appraisals. New methods for multiple levels of information have been developed that allow for high-level review of entire research fields (i.e., "meta-research"). The popularity of meta-analyses differs between different fields. Genomics and smoking are two fields in which very large numbers of these studies have been conducted.

Dr. Ioannidis then provided examples of knowledge integration at the meta-research level and explained how these pertained to issues of associations, predictions, and treatments.

1. A random selection of 50 ingredients from a popular cookbook was assessed to see how many of the ingredients were associated with significantly increased or decreased cancer risk in the literature. Forty out of the 50 ingredients were significantly associated with risk in at least one study. For the remaining 10 items, metabolites were often associated. While some valid associations (e.g., protection by fruit and increased risk from meat) might exist in these data, a meta-analysis of relative risks showed that most of these scattered around 1, and studies that reported very high or very low risks were likely flawed. A plot of the p-values from these studies shows a bimodal distribution with a huge gap in the middle. This is not what one would expect to see if the associations were valid and suggests that there is a very large body of unpublished negative findings.

2. About 2,000 prognostic tumor markers are published each year; only about 12-15 of these are used in clinical practice. The vast majority (90-95 percent) of the studies in the field report significant associations, and a more detailed view of the "negative" studies revealed that very few made a clear claim that there was no effect.[39] Instead, researchers offer apologies and discuss nonsignificant trends.

---

[39] Kyzas, P.A., Denaxa-Kyza, D., and Ioannidis, J.P. (2007). Almost all articles on cancer prognostic markers report statistically significant results. Eur J Cancer 43(17): 2559–2579 (Nov).

3. Ioannidis and Panagiotou (2011) compared effect sizes for cancer biomarkers from individual articles to those in subsequent meta-analyses.[40] The estimates of the effects are essentially always smaller in the larger study than in the initial reports. While this does not automatically invalidate all these findings, it indicates that the true effects are usually much smaller than initially claimed.
4. A recent review of microRNAs as clinical outcome predictors in human cancer showed that the left side of the confidence intervals of essentially all significant studies are just barely above one;[41] this calls for cautious optimism regarding this new technology with, however, a clear remaining risk that a number of these might turn out to be false positives.
5. In an example of a field synopsis, Chatzinasiou et al. (2011) reviewed findings from many different methods and platforms and assessed the findings for robustness.[42]

Researchers in the field of knowledge integration applaud the growth in meta-analyses, which promises to increase the quality of the available data. However, when looking at the numbers of such studies in the field of genetics in recent years, Dr. Ioannidis noted that they were not realistic. A breakdown by country revealed that the vast majority of these studies over the past several years came from China. Most of these studies report significant effects for candidate genes, and most of these are likely wrong. Thus, even the field of meta-analyses deserves scrutiny regarding its methods and biases.

Dr. Ioannidis then focused on treatment issues and noted that preclinical research is currently in a crisis. When Amgen scientists recently tried to replicate oncology drug targets derived from academic research, they could only replicate 6 out of 53 findings.[43] The failure to win the "War on Cancer" thus might be due at least partially to a lack of valid basic scientific findings.

On the topic of clinical trials, Dr. Ioannidis noted that about 700 randomized trials have been carried out for advanced breast cancer, and 1,200 for bevacizumab. When comparing odds ratios between different trials, the drug likely has a true protective effect on mortality of about 10 percent. This is a very different effect size from what was seen in the early trials, which showed large effects and were stopped early, further inflating the initial effect size estimates.

In a paper reporting on an empirical evaluation of very large treatment effects of medical interventions,[44] Dr. Ioannidis and colleagues found that the identified large

---

[40] Ioannidis, J.P.A., and Panagiotou, O.A. (2011). Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses. JAMA 305, 2200–2210.

[41] Nair, V.S., Maeda, L.S., and Ioannidis, J.P.A. (2012). Clinical outcome prediction by microRNAs in human cancer: a systematic review. J. Natl. Cancer Inst. 104, 528–540.

[42] Chatzinasiou, F., Lill, C.M., Kypreou, K., Stefanaki, I., Nicolaou, V., Spyrou, G., Evangelou, E., Roehr, J.T., Kodela, E., Katsambas, A., et al. (2011). Comprehensive field synopsis and systematic meta-analyses of genetic association studies in cutaneous melanoma. J. Natl. Cancer Inst. 103, 1227–1235.

[43] Begley, C.G., and Ellis, L.M. (2012). Drug development: raise standards for preclinical cancer research. Nature 483, 531–533.

[44] Pereira, T.V., Horwitz, R.I., and Ioannidis, J.P.A. (2012). Empirical evaluation of very large treatment effects of medical interventions. JAMA 308, 1676–1684.

effects invariably shrunk in subsequent analyses. He concluded that the field has to learn to live with the fact that true effects are going to be tiny.[45] Referencing several examples from the recent literature,[46] Dr. Ioannidis posited that the field will be able to successfully measure and identify small environmental effects, provided that very large sample sizes are employed to help distinguish the signal from the noise. Hundreds and possibly thousands of environmental exposures (e.g., toxins, nutritional factors) already can be included in these analyses.

Dr. Ioannidis proposed several recommendations for improving research-reporting standards. The STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) standard[47] aims at improving observational studies in epidemiology, and the Genetic Risk Prediction Studies (GRIPS) group works toward strengthening the reporting of genetic risk prediction studies.[48] Dr. Ioannidis further noted that he is a strong believer in registration[49] and that the field will not be successful if it does not widely implement registration of analyses. He reviewed different levels of registration ranging from no registration at all (which should be reserved for exploratory studies) to full open live streaming of all conducted analyses. He also reviewed the data sharing policies by high-impact journals[50] and criticized the reluctance by the field to make full raw data available.

Dr. Ioannidis concluded his presentation by addressing the issues of repeatability and validity of results. In a recent evaluation, he and his colleagues found that only 2 out of 18 microarray studies could be repeated successfully.[51] To improve validation practices, new reward systems are necessary.[52] These incentives should drive validation efforts from analytical validity via repeatability, replication, external validation, and clinical validity to clinical utility.

[45] Siontis, G.C.M., and Ioannidis, J.P.A. (2011). Risk factors and interventions with statistically significant tiny effects. Int J Epidemiol 40, 1292–1307.

[46] Patel, C.J., Cullen, M.R., Ioannidis, J.P.A., and Butte, A.J. (2012). Systematic evaluation of environmental factors: persistent pollutants and nutrients correlated with serum lipid levels. Int J Epidemiol 41, 828–843; Tzoulaki, I., Patel, C.J., Okamura, T., Chan, Q., Brown, I.J., Miura, K., Ueshima, H., Zhao, L., Van Horn, L., Daviglus, M.L., et al. (2012). A nutrient-wide association study on blood pressure. Circulation 126, 2456–2464.

[47] Gallo, V., Egger, M., McCormack, V., Farmer, P.B., Ioannidis, J.P.A., Kirsch-Volders, M., Matullo, G., Phillips, D.H., Schoket, B., Stromberg, U., et al. (2011). STrengthening the Reporting of OBservational studies in Epidemiology—Molecular Epidemiology (STROBE-ME): an extension of the STROBE Statement. PLoS Med. 8, e1001117.

[48] Janssens, A.C.J.W., Ioannidis, J.P.A., Van Duijn, C.M., Little, J., and Khoury, M.J. (2011). Strengthening the reporting of Genetic Risk Prediction Studies: the GRIPS statement. Genet. Med. 13, 453–456.

[49] Ioannidis, J.P.A. (2012). The importance of potential studies that have not existed and registration of observational data sets. JAMA 308, 575–576.

[50] Alsheikh-Ali, A.A., Qureshi, W., Al-Mallah, M.H., and Ioannidis, J.P.A. (2011). Public availability of published research data in high-impact journals. PLoS ONE 6, e24357.

[51] Ioannidis, J.P.A., Allison, D.B., Ball, C.A., Coulibaly, I., Cui, X., Culhane, A.C., Falchi, M., Furlanello, C., Game, L., Jurman, G., et al. (2009). Repeatability of published microarray gene expression analyses. Nat. Genet. 41, 149–155.

[52] Ioannidis, J.P.A., and Khoury, M.J. (2011). Improving validation practices in "omics" research. Science 334, 1230–1232.

Dr. Ioannidis summarized his recommendations for future knowledge integration as follows:

- Knowledge management
  - Methods for mining published and unpublished data.
  - Registration of observational datasets and, when appropriate, protocols.
  - Availability of raw data and analysis codes.
  - Facilitation of repeatability and reproducibility checks; replication culture.
  - Consideration of live stream information.
- Knowledge synthesis
  - Facilitation of consortia with prospective measurements.
  - Optimization of multiconsortial space, competition, and communication.
  - Prospective study networks.
- Knowledge translation
  - Anticipatory rather than post hoc brokering.

Panel and Audience Discussion

- How can epidemiology help integrate knowledge from basic, clinical, and population sciences to accelerate translation from research to practice?

Panelists

Robert A. Hiatt, M.D., Ph.D., *University of California San Francisco*
Katrina Goddard, Ph.D., *Kaiser Permanente Northwest*
Ann Zauber, Ph.D., *Memorial Sloan-Kettering Cancer Center*
Martin L. Brown, Ph.D., *Applied Research Program, DCCPS, NCI*

Dr. Robert Hiatt concentrated his presentation on how to better interact with stakeholders in adjacent fields. He agreed with previous presenters about the fact that epidemiology is a healthy and exciting research field, and he posited that the greatest challenge for the future lies in effective translation of knowledge. The public, the funders, and the legislators are often critical of the field. This can be overcome by working harder to incorporate research into practice.

In this context, Dr. Hiatt provided an alternative definition of the term knowledge integration as "the effective incorporation of knowledge into the decisions practices and policies of organizations and systems." This issue relates back to the question from the audience whether all epidemiology research should be translated into other efforts. Dr. Hiatt noted that the field, indeed, has a responsibility to achieve this goal. Most of the current efforts are still restricted to the T0 (discovery) and T1 (characterization) fields, and epidemiologists in the 21st century must make much greater efforts to translate findings into practice.

Dr. Hiatt further noted the importance of big data. Epidemiologists are not engaged enough in these developments and risk being left behind by adjacent fields if they do not become more engaged in the near future.

On the definition of the most relevant elements of translation, Dr. Hiatt noted that translational science is best accomplished in teams using a transdisciplinary approach. Epidemiologists must emerge as central figures in these teams. No other discipline is as integrative, leaving this field in a unique position to make an impact on society.

Epidemiologists also are poised to play an important and needed role in educating and training health professionals in population health and global perspectives.

Dr. Ann Zauber presented results from the Cancer Intervention and Surveillance Modeling Network (CISNET),[53] an NCI-funded consortium to "use statistical modeling to improve our understanding of cancer control interventions in prevention, screening, and treatment and their effects on population trends in incidence and mortality." These efforts integrate basic, clinical, and population science efforts. There are currently five participating sites, and Dr. Zauber is the clinical coordinator of the colon cancer site.

In a recent microsimulation modeling study in colon cancer,[54] her team assessed potential effects of changes in public health measures on survival rates in Blacks compared to Whites. In 1975, before the introduction of systematic screening efforts, mortalities were similar in the two groups. In 2006, however, there was a very clear difference in age-adjusted mortalities, with rates in Blacks exceeding those in Whites by more than 20 percent.

Dr. Zauber's team within CISNET carried out several modeling efforts that revealed that public health measures that would lead to similar screening rates and overall survivorship in Blacks compared to Whites would be able to achieve a 54 percent reduction of the observed disparity. This illustrates how hypothetical scenarios from computer models can provide epidemiologists with estimates regarding the expected benefits of different interventions, in this case screening and improved clinical care. Additional questions that can be addressed are issues of tumor subtypes and the predicted effects of interventions. This research constitutes one existing effort to translate diverse sources of basic research data into direct clinical action.

Dr. Zauber also has worked with the CMS to evaluate the cost-effectiveness of colonoscopy and has just started a new project that incorporates polygenic risks into the analysis.

Dr. Katrina Goddard reviewed the current major challenges encountered in knowledge integration in epidemiology and potential solutions.

The rapid development of new technologies and the associated increase in the amounts of available data pose tremendous challenges, because each systematic review carried out today in the field of genomics can take 1-2 person years to complete, quickly exhausting the available resources.

Dr. Goddard has worked on new methods to carry out targeted reviews that can be completed within a matter of days and are scalable to a genome-wide scope. She further noted the challenge to distinguish true signals from noise and problems encountered when trying to achieve large enough sample sizes for rare conditions or genetic subgroups.

As potential solutions to be implemented within the next 12 years, Dr. Goddard mentioned the development of electronic tools to improve automation and search algorithms. She further concurred with previous speakers about the necessity for

---

[53] http://cisnet.cancer.gov/.

[54] Lansdorp-Vogelaar, I., Kuntz, K.M., Knudsen, A.B., Van Ballegooijen, M., Zauber, A.G., and Jemal, A. (2012). Contribution of screening and survival differences to racial disparities in colorectal cancer rates. Cancer Epidemiol. Biomarkers Prev. 21, 728–736.

greater standardization in nomenclature, coding, and reporting of research data. Furthermore, there should be a clear definition regarding the required level of evidence to move findings into clinical practice. Epidemiologists should use tools such as value-of-information (VOI) analysis and decision analysis to identify critical issues. They also should make increasing use of alternative study designs, including observational studies, data from EMRs, and "natural experiments" across nations.

Dr. Martin Brown provided a health economist's perspective. Increasing health care costs have been identified as a major problem in his field, although the recently launched Patient-Centered Outcome Research Institute (PCORI), for which he carries out work, has an explicit mandate not to take cost-effectiveness into account when prioritizing research on new treatment strategies. Dr. Brown posited that epidemiology has a lot to contribute to health services research and comparative effectiveness research and that health services research has a significant impact on epidemiology.

Epidemiology can help efforts in health services research and comparative effectiveness research by providing large, representative cohorts with long-term longitudinal follow-up, instead of pure reliance on cross-sectional data aggregation. Observational studies and randomized controlled trials offer important complementary approaches with unique strengths and weaknesses. Epidemiological know-how is crucial to understand how to analyze these data and how to avoid confounders. It is further needed to harmonize, standardize, and validate findings.

Health services research can contribute to epidemiology by providing data sources such as administrative, EMR and EHR data, as well as health system interventions. Health services research can further generate data for simulation models and can validate populations for risk models, markers, subset analyses, and effect sizes.

Opportunities on the horizon, according to Dr. Brown, include the establishment of research requirements for meaningful use of EMR/EHR data. Finally, the uneven implementation of health care reform currently provides the opportunity to study the impact of health care services as a natural experiment.

General Panel Discussion

Big data and the challenge to prevent epidemiology from becoming "incidentalomics"
Dr. Khoury emphasized the need for epidemiologists to take an active role in knowledge integration. Geneticists took care of the genomic data, and epidemiologists must be careful not to lose additional ground in the application of new and emerging technologies.

The speed of discovery: large data projects must be allowed to take their time
- Dr. Ioannidis commented that the reward systems are faulty: if researchers are rewarded for producing a lot of publications quickly, then that is what they will do. The field must find a way to put much greater emphasis on reproducibility. If a hypothetical "reproducibility index" were weighed higher than the publication or citation index, then a reversal in thinking and prioritization could be achieved.
- Dr. Hiatt added that a balance must be found between rapid release of new and promising findings, which eventually might become replicated by the field and established, and cautiousness because of the fear of false positives.

- Dr. Kramer noted that he was pessimistic regarding the ability of replication to take care of the problem and demanded greater care by epidemiologists to provide a comprehensive account of all study limitations.

Heterogeneity and implications for data aggregation

- Dr. Goddard commented that attention needs to be given to report all stratifications that have been applied to a dataset.
- Dr. Brown added that the laboratory for clinical research should be the entire health care system. For this to work, an investment in a health care system database is required. A key characteristic would be to invest in methods to convert "found" data into the rigorous scientific data elements that are required for analysis.
- Dr. Khoury observed that there is a perceived conflict between evidence-based medicine and prevention that applies to the average person in the population. Individual patients are interested in their own treatment and not in effects measured in an entire population that might not apply to them.
- Dr. Zauber commented that CISNET always attempts to include information about heterogeneity into the models, but it needs solid data before they can be incorporated into realistic models. Dr. Brown mentioned that CISNET has led to a standardization of modeling efforts and better annotation of models to overcome the problem of inconsistent findings.
- Dr. Olopade contended that current data collections from diverse populations are insufficient to provide a solid enough base to understand the causes of disparities. Much more data are needed to make universally applicable health recommendations across heterogeneous populations.
- Dr. Ioannidis noted confusion in the field regarding the term heterogeneity. The field must move beyond its struggles with issues of noise and bias to "genuine" heterogeneity that refers to actual differences between individuals. Currently, the noise is still too dominant to make inferences about the latter definition of the term.
- Dr. Hiatt emphasized the need to integrate the clinical and population health enterprises into the research efforts. The complexity in social and population determinants is just as large as genetic heterogeneity, and tools must be developed to address both types of complexity.

Questions from the Online Audience

- One online participant observed that librarians have been missing from the discussion on knowledge integration.
- Another question referred to the use of natural language processing to deal with unclean data. Dr. Brown noted that the method can be used once it has been validated, as one of many other approaches.
- Another comment referred to the use of sensitivity analysis in epidemiology. Dr. Zauber stated that sensitivity analysis is an essential component of her modeling efforts that is always performed at the onset. Dr. Brown emphasized the importance of sensitivity analysis to obtain realistic estimates of actual ranges of parameters in the real world, which can be distorted in nonobservational studies.

Moderated Discussion to Develop 12 Broad Recommendations for Action in the Next 10-20 Years

Dr. Patricia Hartge explained that the purpose of this session was to begin to prioritize ideas derived from the discussions during the past 2 days. The intended audience consisted of the workshop participants including individuals following online and through social media; early career epidemiologists establishing their research careers; and funders at the NIH and elsewhere.

Dr. Hartge remarked on the substantial overlap between the issues and priorities discussed online and on-site. She then presented a summary of 12 priorities from the online community (Appendix 3).

The participants conducted a quick poll among themselves to get a snapshot of the issues that were perceived as most urgent by this group. Dr. Khoury noted that these priorities will be the basis for the development of a "12 in 12" list of ideas that should be implemented within the next 12 years. He clarified that these efforts are considered a "work in progress" and that additional exchanges with the participants are expected to occur by e-mail to further develop and refine this list. The following priorities were mentioned, which are clustered by theme and not in any particular order:

Study Design

- Evaluate risks and benefits from the design stage of each study.
- Think totally outside the box. Look at Canada. Conduct thought experiments with unlimited resources; employ much broader thinking.
- Invest in scalable infrastructure and interdisciplinary collaborations.
- Consider multiple aspects of the system from early on.
- Maximize returns from each study. Studies should be creative, flexible, and cost-efficient.
- Develop research designs in the context of health care systems.
- Develop a rapid response strategy. Create a funding mechanism to respond to ongoing national experiments as laws change and interpretations emerge. Make use of the experiment as long as it is possible. Concrete example: health disparities during the ACA rollout, but can be more general, e.g., economic status, insurance status, national disasters. Opportunity can be lost by the time it goes through the current bureaucracy.

Open Access/Sharing/Harmonization

- Require data sharing and open access.
- Establish incentives and reward systems to promote efficient sharing of data.
- Clearly articulate what is new, and insist on harmonization and open access of the "old" core features.
- Employ cross-study techniques, and apply harmonization to maximize outcomes.
- Create more synergies with the NIH and other investigators.

- Because replication is so central, there must be not only data sharing, but also replication plans in each research design. The NIH should advocate that journals make this a requirement and consider including this as a condition of funding.
- Establish an independent replication center.

Cohorts

- Build a nationalized epidemiology network. Give everybody the option to opt in, and ensure that a substantial proportion of the U.S. population participates. Learn from countries with a unified health care system.
- Establish a North American cohort. Study participants in the general public.
- Build an international epidemiology network. Some issues are actionable in other structures, but not here, at least not now.

RCTs/Observational Studies

- Encourage rational novel small studies using existing cohorts and repositories.
- Study extreme phenotypes.
- Think of studies that can be launched immediately; consider how to piggyback onto existing studies.
- Undertake comprehensive exposome profiling from birth to death with as large a sample as possible.
- Incorporate all the exciting new technologies.

Data Integration

- Apply interdisciplinary data integration, so that data from different levels truly can be integrated.
- Integrate GWAS with functional data.
- Carefully consider methods and structure.

Data Quality

- Emphasize the importance of high-quality data from any study design.

Specific Research Content

- Put more emphasis on research on stress, which plays a role in a lot of systems; ask questions about stress and social support.
- Study early life programming of adult disease and discover new biomarkers.
- Address comprehensive "–omics."
- Discover intermediate biomarkers.
- Give greater emphasis to pediatric cancers.

Translation of Research Findings

- Find a greater balance between new discoveries and translational research.
- Overhaul the health care educational system.
- Do a better job in promoting what we know works.
- Train and support a new cadre of epidemiologists in T3 (implementation and health services) science.
- Encourage new designs in implementation and dissemination science, led by epidemiologists.
- Expand the boundaries of what epidemiology does.
- Use epidemiologic methods to realize the promise of precision medicine.

- Establish community-based best practices, but leave them dynamic.
- Make an impact on public health.

Improve Communication

- Better communicate to the public and seek input about what they think is important and how they understand risk; find new ways to better foster two-way communication.

Dr. Khoury concluded with thoughts about reinvigorating epidemiology, adapting to the new era, and expanding the boundaries of epidemiology.

Appendix 1: Workshop Agenda

Wednesday, December 12, 2012	
12:00 pm to 1:00 pm	**Registration**
1:00 pm to 1:10 pm	**Welcome** Robert T. Croyle, PhD *Division of Cancer Control and Population Sciences, NCI*
1:10 pm to 1:30 pm	**Charge to Participants** Muin J. Khoury, MD, PhD *Epidemiology and Genomics Research Program, DCCPS, NCI*

Session 1: Setting the stage: the evolution of epidemiology and its applications to cancer
Moderator: Robert T. Croyle, PhD, *DCCPS, NCI*

1:30 pm to 2:00 pm	**Historical perspectives on the evolution of cancer epidemiology** Robert N. Hoover, MD, ScD *Division of Cancer Epidemiology and Genetics, NCI*
2:00 pm to 3:00 pm	**Panel and Audience Discussion** **What lessons and success stories have we learned from 20th century cancer epidemiology?** **What are the major scientific questions that cancer epidemiology should address in the next decade to impact public health?** Panelists: David Hunter, ScD, MPH, *Harvard University* Timothy Rebbeck, PhD, *University of Pennsylvania* Margaret R. Spitz, MD, *Baylor College of Medicine* *Audience and Web Participation*
3:00 pm to 3:15 pm	**Break**

Session 2: The impact of new methods and technologies on epidemiologic research
Moderator: Stephen J. Chanock, MD, *DCEG, NCI*

3:15 pm to 3:45 pm	**Technology-driven epidemiology: a paradigm shift** Geoffrey S. Ginsburg, MD, PhD *Institute for Genome Sciences and Policy, Duke University*
3:45 pm to 5:00 pm	**Panel and Audience Discussion** **Which technologies do you feel are ready for "prime time" in epidemiologic research and for what purpose?** **What criteria would you use to determine when emerging technologies should be integrated into epidemiologic research?**

Panelists:

Zdenko Herceg, PhD, *International Agency for Research on Cancer*

Thomas A. Sellers, PhD, MPH, *Moffitt Cancer Center*

Michael Snyder, PhD, *Stanford University*

Georgia D. Tourassi, PhD, *Oak Ridge National Laboratory*

Audience and Web Participation

6:00 pm to 8:00 pm	**Optional Dinner at Bethesda Restaurant**
	Note: Attendees will be responsible for meals at their own cost

Thursday, December 13, 2012

7:30 am to 8:00 am	**Registration**

Session 3: The evolution of epidemiologic cohorts in the study of natural history of cancer and other diseases
Moderator: Deborah M. Winn, PhD, *DCCPS, NCI*

8:00 am to 8:30 am	**What have we learned from epidemiology cohorts and where should we be going next?**
	Julie Buring, ScD, MS
	Harvard School of Public Health
8:30 am to 9:30 am	**Panel and Audience Discussion**
	What developments are needed to make epidemiologic cohorts a cornerstone of the discovery to practice continuum—bridging the transition from etiology to outcomes to policy and practice?
	How should NCI and NIH facilitate multidisciplinary collaboration to integrate these developments into the research portfolio?
	Panelists:
	Julie R. Palmer, ScD, MPH, *Boston University School of Public Health*
	Lyle Palmer, PhD, *Ontario Institute for Cancer Research*
	Leslie L. Robison, PhD, *St. Jude Cancer Center*
	Daniela Seminara, PhD, MPH, *DCCPS, NCI*
	Audience and Web Participation

Session 4: Use of epidemiologic research to advance clinical and public health practice: bridging the evidence gap with observational studies and randomized clinical trials
Moderator: Sheri D. Schully, PhD, *DCCPS, NCI*

9:30 am to 10:00 am	**Epidemiology and evidence-based research along the cancer care contiuum**
	David F. Ransohoff, MD
	University of North Carolina at Chapel Hill
10:00 am to 11:00 am	**Panel and Audience Discussion**
	What are new ways in which epidemiology can be used to fill evidence gaps between discoveries and population health impact in the cancer care continuum?
	How can observational epidemiology make the greatest scientific contributions in understanding cancer-related risk factors that cannot be studied through randomized clinical trials?
	Panelists:
	Barry Kramer, MD, MPH, *Division of Cancer Prevention, NCI*

Michael Lauer, MD, *Division of Cardiovascular Sciences, NHLBI*
Jeffrey A. Meyerhardt, MD, MPH, *Harvard Medical School*
Olufunmilayo I. Olopade, MD, FACP, *University of Chicago*

Audience and Web Participation

11:00 am to 11:15 am	**Break**

Session 5: Use of epidemiology in knowledge integration and meta-research
Moderator: Muin J. Khoury, MD, PhD, *DCCPS, NCI*

11:15 am to 11:45 am	**The role of epidemiology in knowledge integration and meta research**

John Ioannidis, MD, DSc
Stanford Prevention Research Center

11:45 am to 12:45 pm	**Panel and Audience Discussion**

How can epidemiology help integrate knowledge from basic, clinical and population sciences to accelerate translation from research to practice?

Panelists:
Martin Brown, PhD, *DCCPS, NCI*
Katrina Goddard, PhD, *Kaiser Permanente Northwest*
Robert A. Hiatt, MD, PhD, *University of California San Francisco*
Ann Zauber, PhD, *Memorial Sloan Kettering Cancer Center*

Audience and Web Participation

12:45 pm to 1:15 pm	**Break to grab lunch from cafeteria**

Note: Attendees will be responsible for meals at their own cost

Session 6: Where do we go from here? 12 recommendations for epidemiology in the next 12 years
Moderator: Patricia Hartge, ScD, DCEG, NCI

1:15 pm to 2:30 pm	**Working Lunch and moderated discussion to come up with 12 broad recommendations for action in the next 10-20 years**

Appendix 2: List of Workshop Participants

John Beresny
Principal
ICF International

Lewis Berman
Vice President
ICF International

Diane Bild
Associate Director for Prevention &
Population Sciences
Division of Cardiovascular Sciences
National Heart, Lung, & Blood Institute, NIH

Julie Buring
Professor of Medicine
Harvard Medical School/Brigham &
Women's Hospital

Robert Burk
Professor
Pediatrics, Microbiology & Immunology,
Epidemiology & Population Health &
Obstetrics, Gynecology & Women's Health
Albert Einstein College of Medicine

Celia Byrne
Associate Professor
Epidemiology & Biostatistics Division
Uniformed Services University

Yongmei Chen
Senior Biostatistician, Researcher
Center for Prostate Disease Research

Linda Cook
Professor
Epidemiology & Biostatistic,
Internal Medicine
University of New Mexico

Jennifer Cullen
Director, Epidemiologic Research
Surgery
DOD, Center for Prostate Disease Reseach

Richard Fabsitz
Deputy Chief, Epidemiology Branch
Epidemiology Branch
National Heart, Lung, & Blood Institute, NIH

Jo Freudenheim
Distinguished Professor & Chair
Department of Social & Preventive Medicine
University at Buffalo

Lisa Gallicchio
Epidemiologist
The Prevention & Research Center
Mercy Medical Center

Geoffrey Ginsburg
Executive Director, Center for Personalized
Medicine
Director, Genomic Medicine
Professor of Pathology & Medicine
Institute for Genome Sciences & Policy
Duke University

Katrina Goddard
Senior Investigator
Center for Health Research
Kaiser Permanente

Laurel Habel
Section Chief, Cancer Research
Division of Research
Kaiser Permanente

Jane Harman
Epidemiologist
Division of Cardiovascular Sciences
National Heart, Lung, & Blood Institute, NIH

Terryl Hartman
Professor
Nutritional Sciences
Penn State University

Zdenko Herceg
Head of the Mechanisms of Carcinogenesis
Section and the Epigenetics Group
International Agency for Research on
Cancer

Robert Hiatt
Professor, Chair
Epidemiology & Biostatistics
University of California, San Francisco

David Hunter
Dean for Academic Affairs
Harvard School of Public Health

Terry Hyslop
Director, Division of Biostatistics
Thomas Jefferson University

John Ioannidis
Professor of Medicine, Health Research &
Policy & Statistics
Stanford University

Steven Krosnick
Medical Officer
National Institute of Biomedical Imaging &
Bioengineering, NIH

Michael Lauer
Director
Division of Cardiovascular Sciences
National Heart, Lung, & Blood Institute, NIH

Gertraud Maskarinec
Professor of Epidemiology
University of Hawaii Cancer Center

Susan Mayne
Professor of Epidemiology
Yale School of Public Health & Yale Cancer
Center

Jeffrey Meyerhardt
Associate Professor of Medicine
Medical Oncology
Dana-Farber Cancer Institute

Marian Neuhouser
Full Member, Cancer Prevention
Fred Hutchinson Cancer Research Center

Olufunmilayo Olopade
Professor of Medicine & Human Genetics
Director, Cancer Risk Clinic
Medical Center
The University of Chicago

Jean Olson
Deputy Chief, Epidemiology Branch
Division of Cardiovascular Sciences
National Heart, Lung, & Blood Institute, NIH

Julie Palmer
Professor
Slone Epidemiology Center
Boston University

Lyle Palmer
Executive Scientific Director
Ontario Health Study
Ontario Institute for Cancer Research

Jong Park
Associate Member
Cancer Epidemiology
Moffitt Cancer Center

Edward Peters
Program Director, Epidemiology
Louisiana State University School of Public
Health

Susan Pinney
Professor, Environmental Health
University of Cincinnati, College of Medicine

David Ransohoff
Professor of Medicine
University of North Carolina

Timothy Rebbeck
Professor
University of Pennsylvania

Kim Robien
Associate Professor
Epidemiology & Biostatistics
George Washington University

Betsy Rolland
Project Manager, Asia Cohort Consortium
Public Health Sciences Division
Fred Hutchinson Cancer Research Center

Laura Rozek
Assistant Professor, Environmental Health
University of Michigan

Thomas Sellers
Center Director
H Lee Moffitt Cancer Center & Research
Institute

Michael Snyder
Chair of Genetics
Stanford University

Paul Sorlie
Epidemiologist
Epidemiology Branch
National Heart, Lung, & Blood Institute, NIH

Margaret Spitz
Professor
Dan L. Duncan Cancer Center
Baylor College of Medicine

Pothur Srinivas
Program Director
Cardiovascular Sciences, Epidemiology
National Heart, Lung, & Blood Institute, NIH

Patricia Thompson
Associate Professor, Cellular & Molecular
Medicine
Leader, Cancer Prevention & Control
Program
The University of Arizona Cancer Center

Georgia Tourassi
Director
Biomedical Science & Engineering Center
Oak Ridge National Laboratory

Gina Wei
Medical Officer
Division of Cardiovascular Sciences
National Heart, Lung, & Blood Institute, NIH

Mary White
Chief, Epidemiology & Applied Research
Branch
Division of Cancer Prevention & Control
Centers for Disease Control & Prevention

John Witte
Professor
University of California, San Francisco

Huichun Xu
Research Fellow
National Human Genome Research
Institute, NIH

Ann Zauber
Biostatistician
Department of Epidemiology & Biostatistics
Memorial Sloan-Kettering Cancer Center

NATIONAL CANCER INSTITUTE STAFF
DIVISION OF CANCER CONTROL & POPULATION SCIENCES (DCCPS)

Robert Croyle
Director

Deborah Winn
Deputy Director

Audrey Babkirk
Health Communications Fellow
Epidemiology and Genomics Research
Program

Rachel Ballard-Barbash
Associate Director
Applied Research Program

Dacia Beard
Health Communications Fellow
Epidemiology & Genomics Research
Program

Heather Bowles
Epidemiologist
Applied Research Program

Martin Brown
Chief, Health Services & Economics Branch
Applied Research Program

Danielle Carrick
Program Director
Epidemiology & Genomics Research
Program

Christine Chang
CRTA Fellow
Epidemiology & Genomics Research
Program

Rao (Mahesh) Divi
Program Director
Epidemiology & Genomics Research
Program

Vincent Doria-Rose
Epidemiologist
Applied Research Program

Paul Ebohon
Presidential Management Fellow
Epidemiology & Genomics Research
Program

Brenda Edwards
Senior Advisor Surveillance
Surveillance Research Program

Stephanie George
Postdoctoral Fellow
Applied Research Program

Elizabeth Gillanders
Chief, Host Susceptibility Factors Branch
Epidemiology & Genomics Research
Program

Carolyn Hutter
Program Director
Epidemiology & Genomics Research
Program

Christie Kaefer
Communications Coordinator
Epidemiology & Genomics Research
Program

Erin Kent
Program Director
Applied Research Program

Muin Khoury
Acting Associate Director
Epidemiology & Genomics Research
Program

Tram Lam
Program Director
Epidemiology & Genomics Research
Program

Leah Mechanic
Program Director
Epidemiology & Genomics Research
Program

Stefanie Nelson
Program Director
Epidemiology & Genomics Research
Program

Nancy Potischman
Nutritional Epidemiologist
Applied Research Program

Jill Reedy
Nutritionist
Applied Research Program

Sheri Schully
Knowledge Integration Team Lead
Epidemiology & Genomics Research
Program

Daniela Seminara
Scientific Consortia Coordinator
Epidemiology & Genomics Research
Program

Amy Subar
Nutritionist
Applied Research Program

Mukesh Verma
Chief, Methods & Technologies Branch
Epidemiology & Genomics Research
Program

Gordon Willis
Cognitive Psychologist
Applied Research Program

Giovanna Zappalà
Medical Affair Scientist
Basic Biobehavioral & Psychological
Sciences Branch

DIVISION OF CANCER EPIDEMIOLOGY & GENETICS

Margaret Tucker
Acting Director
Director, Human Genetics Program

Zaynah Abid
Research Analyst
Epidemiology & Biostatistics Program

Amy Berrington
Senior Investigator
Radiation Epidemiology Branch

Stephen Chanock
Chief
Laboratory of Translational Genomics

Michael Cook
Investigator
Hormonal & Reproductive Epidemiology
Branch

Jessica Faupel-Badger
Senior Biomedical Scientist
Cancer Prevention Fellowship Program

Patricia Hartge
Deputy Director
Epidemiology & Biostatistics Program

Robert Hoover
Director
Epidemiology & Biostatistics Program

Scott Kelly
Pre-doctoral Fellow
Hormonal & Reproductive Epidemiology
Branch

Barnett Kramer
Director, Division of Cancer Prevention
Division of Cancer Prevention

Jennifer Loukissas
Communications Manager
Office of Communications & Special
Initiatives

Jianbing Wang
Postdoctoral Fellow
Nutritional Epidemiology Branch

Appendix 3: Summary of Priorities from the Online Community

Excerpted from presentation by Patricia Hartge, Sc.D., Epidemiology and Biostatistics Program, Division of Cancer Epidemiology and Genetics, NCI

1. **Big Data**
   - Improved analytical tools, data infrastructure harmonization, and models

2. **Biorepositories**
   - Epidemiologically sound biorepositories that can capitalize on emerging technologies

3. **Unravel the "Exposome"**
   - Reliable methods to quantify exposures and to identify the combined effects of genetic and other factors on complex diseases

4. **Cohorts**
   - Support cohort studies that focus on multiple outcomes

5. **Natural History**
   - Initiatives to better understand natural history of disease

6. **Multilevel Analysis**
   - Multiple levels of influences into epidemiologic research

7. **Knowledge Integration**
   - Tools that integrate knowledge from disparate disciplines

8. **RCTs/Observational Studies**
   - Embed RCTs and observational studies into flexible designs and leverage existing resources when applicable

9. **Open-Access and Sharing**
   - Data and specimen sharing to promote collaboration and accelerate discovery and translation across disciplines

10. **Education and Training**
    - Train with eye toward collaboration, translation, and global health

11. **Career Advancement**
    - Career advancement that promotes collaboration and rewards translation

12. **Global Health and Health Disparities**
    - Assess issues relevant to health disparities and global health

NIH...Turning Discovery Into Health®