Report calls for creation of a biomedical research and patient data network for more accurate classification of diseases

November 2 2011

A new data network that integrates emerging research on the molecular makeup of diseases with clinical data on individual patients could drive the development of a more accurate classification of disease and ultimately enhance diagnosis and treatment, says a new report from the National Research Council. The "new taxonomy" that emerges would define diseases by their underlying molecular causes and other factors in addition to their traditional physical signs and symptoms. The report adds that the new data network could also improve biomedical research by enabling scientists to access patients' information during treatment while still protecting their rights. This would allow the marriage of molecular research and clinical data at the point of care, as opposed to research information continuing to reside primarily in academia.

"Currently, a disconnect exists between the wealth of scientific advances in research and the incorporation of this information into the clinic," said Susan Desmond-Hellmann, co-chair of the committee that authored the report and chancellor of the University of California, San Francisco. "Often it can take years for biomedical research information to trickle to doctors and patients, and in the meantime wasteful health care expenditures are carried out for treatments that are only effective in specific subgroups. In addition, researchers don't have access to comprehensive and timely information from the clinic. Overall, opportunities are being missed to understand, diagnose, and treat diseases more precisely, and to better inform health care decisions."

"Developing this new network and the associated classification system will require a long-term perspective and parallels the challenges of building Europe's great cathedrals -- one generation will start building them, but they will ultimately be completed by another, with plans changing over time," said committee co-chair Charles Sawyers, a Howard Hughes Medical Institute investigator and the inaugural director of the Human Oncology and Pathogenesis Program at Memorial Sloan-Kettering Cancer Center. "Dramatic advances in biology and technology have enabled rapid, comprehensive, and cost-efficient analysis of patients' health information, which has resulted in an explosion of data that could dramatically alter disease classification. Health care costs have also steadily increased without translating into significantly improved clinical outcomes. These circumstances make it a perfect time to modernize disease classification."

Typically, disease taxonomy refers to the International Classification of Diseases (ICD), a system established more than 100 years ago that is used to track and diagnose disease and determine reimbursement for care. Under ICD, which is in its 10th edition, disease classifications are primarily based on signs and symptoms and seldom incorporate rapidly emerging molecular data, incidental patient characteristics, or socio-environmental influences on disease.

This approach may have been adequate in an era when treatments were largely directed toward symptoms rather than underlying causes, but diagnosis based on traditional signs and symptoms alone carries the risk of missing or misclassifying diseases, the committee said. For instance, symptoms in patients are often nonspecific and rarely identify a disease unambiguously, and numerous diseases, such as cancer and HIV infection, are asymptomatic in the early stages. Moreover, many subgroups of certain diseases have diverse molecular causes and are classified as one disease and, conversely, multiple diseases share a common molecular cause and are not categorized in the same disease

classification.

The committee noted several areas where classification of diseases based on genetic makeup is already happening with new drug approvals. For example, in a set of trials on patients with non-small-cell lung cancer, a drug was shown to produce dramatic anti-tumor effects in approximately 10 percent of the patients while other patients did not respond at all. Aided by the dramatic tumor shrinkage, the drug was approved and used on a broad range of lung cancer patients but did nothing for most other than increase costs and side effects. Subsequently, it was discovered that patients who responded to the drug carried specific genetic mutations. This allowed doctors to predict which patients would respond and led to the design of more effective clinical trials, reduced treatment costs, and increased treatment effectiveness. Since then, other studies have further divided lung cancers into subsets that are defined by driver mutations.

Framework to Achieve a New Disease Taxonomy

The committee recommended a modernization and reorientation of the information systems used by researchers and health care providers to attain the new taxonomy and move toward precision medicine. It suggested a framework for creating a "knowledge network of disease" that integrates the rapidly expanding range of information on what causes diseases and allows researchers, health care providers, and the public to share and update this information. The first stage in developing the network would involve creating an "information commons" that links layers of molecular data, medical histories, including information on social and physical environments, and health outcomes to individual patients. The second stage would construct the network and require data mining of the information commons to highlight the data's interconnectedness and integrate it with evolving research. Fundamentally, data would be continuously deposited by the research community and extracted directly from the medical records of

participating patients.

To acquire information for the knowledge network, the committee recommended designing strategies to collect and integrate disease-relevant information; implementing pilot studies to assess the feasibility of integrating molecular parameters with medical histories in the ordinary course of care; and gradually eliminating institutional, cultural, and regulatory barriers to widespread sharing of individuals' molecular profiles and health histories while still protecting patients' rights. Much of the initial work necessary to develop the information commons should take the form of observational studies, which would collect molecular and other patient data during treatment. Having this access at point of care could reduce the cost of research, make scientific advances relevant to real-life medicine, and facilitate the use of electronic health records.

The committee noted that moving toward individualized medicine requires that researchers and health care providers have access to very large sets of health and disease-related data linked to individual patients. These data are also critical for developing the information commons, the knowledge network of disease, and ultimately the new taxonomy.

Provided by National Academy of Sciences