



National
Comprehensive
Cancer
Network®

JNCCN

Journal of the National Comprehensive Cancer Network

SUPPLEMENT

Data Needs in Oncology: “Making Sense of The Big Data Soup”

Jessica K. DeMartino, PhD, and Jonathan K. Larsen, MPP

The National Comprehensive Cancer Network® (NCCN®) appreciates that supporting companies recognize NCCN®'s need for autonomy in the development of the content of NCCN resources. All NCCN content is produced completely independently. NCCN Guidelines® are not intended to promote any specific therapeutic modality. The distribution of this report is supported by a grant from Abbott Laboratories; supported by an educational donation provided by Amgen; supported by Apotex Corp, Genentech, Inc., Janssen, Novartis Oncology, Onyx Pharmaceuticals, Inc., and Teva Oncology; sponsored by Boehringer Ingelheim Pharmaceuticals, Inc., Helsinn Therapeutics (U.S.), Inc., and Xcenda, an AmerisourceBergen Consulting Services Company; supported by a charitable donation from Bristol-Myers Squibb; supported by an unrestricted grant from Eisai Inc.; supported by a contribution from Lilly USA, LLC.; supported by a scientific grant from Millennium: The Takeda Oncology Company; and supported in part by ProStrakan Inc.



Harborside
Press®

NCCN.org | JNCCN.org

A supplement to JNCCN—Journal of the National Comprehensive Cancer Network

Masthead

Editorial

Editor-in-Chief:

Harold J. Burstein, MD, PhD

National Comprehensive Cancer Network® (NCCN®)

Senior Director, Professional and Patient Publications:

Kimberly A. Callan, MS, ELS

Assistant Managing Editor:

Kerrin M. Green, MA

Journal Production Specialist:

Genevieve Emberger Hartzman, MA

National Comprehensive Cancer Network

Chairman of the Board:

Thomas A. D'Amico, MD

Vice Chair of the Board:

Samuel M. Silver, MD, PhD

Chief Executive Officer:

Robert W. Carlson, MD

Executive VP/Chief Operating Officer:

Patricia J. Goldsmith

Senior VP, Finance/Chief Financial Officer:

Lisa G. Kimbro, CPA, MBA

NCCN Guidelines®

Senior VP, Clinical Information & Publications:

Joan S. McClure, MS

VP, Clinical Information Operations:

Kristina M. Gregory, RN, MSN, OCN

Director, Patient and Clinical Information Operations:

Dorothy A. Shead, MS

Guidelines Coordinators:

Nicole R. McMillian, MS

Mary A. Dwyer, MS

Mary Anne Bergman

Oncology Scientists/Sr. Medical Writers:

Miranda Hughes, PhD

Hema Sundar, PhD

Rashmi Kumar, PhD

Maria Ho, PhD

Deborah A. Freedman-Cass, PhD

Maoko Naganuma, MSc

Administrative Coordinators:

Jean Marie Dougherty

Donna Black

Business Development and Marketing

VP, U.S. & Global Development:

C. Lyn Fitzgerald, MJ

Director, Marketing:

Jennifer P. Tredwell, MBA

Advertising

Harborside Press®

Director of Business Development:

David Horowitz

Publishing and Design

Harborside Press®

Executive Editor:

Conor Lynch

Layout Artist:

Kristina O'Toole

Production Coordinator:

Wendy McGullam

President:

Anthony Cutrone

Publisher:

John A. Gentile, Jr.

Postal and Contact Information

JNCCN (ISSN 1540-1405, USPS Publication Number 240), the official journal of the National Comprehensive Cancer Network® (NCCN®), is published monthly, except semi-monthly in May, by Harborside Press®, 37 Main Street, Cold Spring Harbor, NY 11724. Periodicals Postage paid at Cold Spring Harbor, NY and additional mailing offices.

Change of Address: Postmaster: send address changes to JNCCN, c/o Harborside Press®, 37 Main Street, Cold Spring Harbor, NY 11724. Recipient: to change your address contact subscriptions@harborsidepress.com or fax 631-692-0805. Please state that this change of address request is for JNCCN.

© National Comprehensive Cancer Network, Inc. 2013. All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means now or hereafter known, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from NCCN.

Subscriptions: Prices for yearly subscriptions (13 issues plus supplements) are: **Individual:** Print only or online only, US \$572; Can/Mex \$644; Int'l \$716; print and online, US \$627; Can/Mex \$699; Int'l \$771. **Institutional:** Print only or online only, US \$913; Can/Mex \$985; Int'l \$1057; print and online, US \$990; Can/Mex \$1062; Int'l \$1134. **Single Copy:** US \$105; Can/Mex \$111; Int'l \$117. **Subscription Inquiries** should be directed to Wendy McGullam, Harborside Press®, at: 631-935-7651 or wendy@harborsidepress.com. **Online access** is available to subscribers through HighWire Press (JNCCN.org).

Contact Information

Editorial Office: Manuscripts, correspondence, and commentaries to be considered for publication should be sent to JNCCN, 275 Commerce Drive, Suite 300, Fort Washington, PA 19034; or e-mail JNCCN@nccn.org. Correspondence can also be faxed: 215-690-0283 (attn: JNCCN). Questions about requirements for publication or topic suitability can be directed as above or to Harold J. Burstein, MD, PhD, Editor-in-Chief, JNCCN, 275 Commerce Drive, Suite 300, Fort Washington, PA 19034; or e-mail burstein@nccn.org.

Instructions for authors are published in JNCCN as space allows and can be found online at JNCCN.org. They can also be requested by calling 215-690-0270 or e-mailing JNCCN@nccn.org.

Advertising

To purchase advertising space: Contact David Horowitz, Director of Business Development, Harborside Press®, 37 Main Street, Cold Spring Harbor, NY 11724; phone 631-935-7652; fax 631-692-0805; or e-mail david@harborsidepress.com.



To send ad materials: Ship to Harborside Press®, Attn: Wendy McGullam, (JNCCN, Vol ___ Issue ___), 37 Main Street, Cold Spring Harbor, NY 11724; phone 631-935-7651; fax 631-692-0805; or e-mail wendy@harborsidepress.com.

To send pre-printed inserts: Ship to Publishers Press, Inc., Attn: Tammy Baugh, 13487 South Preston Highway, Lebanon Junction, KY 40150.

Production

Reprints: Reprints of individual articles are available. Orders must be for a minimum of 100 copies. Please contact David Horowitz, Director of Business Development, Harborside Press®, 37 Main Street, Cold Spring Harbor, NY 11724; phone 631-935-7652; fax 631-692-0805; or e-mail david@harborsidepress.com.

Permissions

For information about photocopying, republishing, reprinting, or adapting material, please go online to NCCN.org/about/permissions.

Indexing

JNCCN is indexed by MEDLINE/PUBMED®, Chemical Abstracts, EMBASE, EmCare, and Scopus. This paper meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper) effective with Volume 1, Issue 1, 2003.

JNCCN is a member of the Medscape Publisher's Circle®, an alliance of leading medical publishers whose content is featured on Medscape (<http://www.medscape.com>). Medscape is part of the WebMD Medscape Health Network, a leading online healthcare resource for professionals and consumers.

Disclaimer

The treatment algorithms presented in JNCCN are a statement of evidence and consensus of the authors regarding their views of currently accepted approaches to treatment. **Any clinician seeking to apply or consult these guidelines is expected to use independent medical judgment in the context of individual circumstances to determine any patient's care or treatment.** The research articles, reviews, and other individually authored papers presented herein are the work of the authors listed. Furthermore, the reader is advised that, except where specifically stated, all of the ideas and opinions expressed in JNCCN are the authors' own and do not necessarily reflect those of NCCN, the member organizations, the editor, or the publisher. Publication of an advertisement or other product mention in JNCCN should not be construed as an endorsement of the product or the manufacturer's claims.

The information contained in JNCCN is presented for the purpose of educating our readership on cancer treatment and management. The information should not be relied on as complete or accurate, nor should it be relied on to suggest a course of treatment for a particular individual. It should not be used in place of a visit, call, consultation, or the advice of a licensed physician or other qualified health care provider. Patients with health care-related questions or concerns are advised to contact a physician or other qualified health care provider promptly.

Although every attempt has been made to verify that information presented within is complete and accurate, the information is provided "AS IS" without warranty, express or implied. NCCN hereby excludes all implied warranties of merchantability and fitness for a particular use or purpose with respect to the Information. Furthermore, NCCN makes no warranty as to the reliability, accuracy, timeliness, usefulness, adequacy, completeness, or suitability of the information.

NCCN Member Institutions

City of Hope Comprehensive Cancer Center
Los Angeles, California

Dana-Farber/Brigham and Women's Cancer Center
Massachusetts General Hospital Cancer Center
Boston, Massachusetts

Duke Cancer Institute
Durham, North Carolina

Fox Chase Cancer Center
Philadelphia, Pennsylvania

Huntsman Cancer Institute at the University of Utah
Salt Lake City, Utah

Fred Hutchinson Cancer Research Center/
Seattle Cancer Care Alliance
Seattle, Washington

The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins
Baltimore, Maryland

Robert H. Lurie Comprehensive Cancer Center of Northwestern University
Chicago, Illinois

Memorial Sloan-Kettering Cancer Center
New York, New York

Moffitt Cancer Center
Tampa, Florida

The Ohio State University Comprehensive Cancer Center – James Cancer Hospital and Solove Research Institute
Columbus, Ohio

Roswell Park Cancer Institute
Buffalo, New York

Siteman Cancer Center at Barnes-Jewish Hospital and Washington University School of Medicine
St. Louis, Missouri

St. Jude Children's Research Hospital/The University of Tennessee Health Science Center
Memphis, Tennessee

Stanford Cancer Institute
Stanford, California

University of Alabama at Birmingham Comprehensive Cancer Center
Birmingham, Alabama

UC San Diego
Moore's Cancer Center
La Jolla, California

UCSF Helen Diller Family Comprehensive Cancer Center
San Francisco, California

University of Colorado Cancer Center
Aurora, Colorado

University of Michigan Comprehensive Cancer Center
Ann Arbor, Michigan

UNMC Eppley Cancer Center at The Nebraska Medical Center
Omaha, Nebraska

The University of Texas MD Anderson Cancer Center
Houston, Texas

Vanderbilt-Ingram Cancer Center
Nashville, Tennessee

For more information, visit NCCN.org

Mission Statement

JNCCN is dedicated to improving the quality of cancer care locally, nationally, and internationally while enhancing the collaboration between academic medicine and the community physician. JNCCN is further committed to disseminating information across the cancer care continuum by publishing clinical practice guidelines and reporting rigorous outcomes data collected and analyzed by experts from the world's leading care centers. JNCCN also provides a forum for original research and review papers focusing on clinical and translational research and applications of the NCCN Guidelines in everyday practice, as well as correspondence and commentary.



NCCN
275 Commerce Drive
Suite 300
Fort Washington, PA 19034
215-690-0300
NCCN.org

About NCCN

The National Comprehensive Cancer Network® (NCCN®), a not-for-profit alliance of 23 of the world's leading cancer centers, is dedicated to improving the quality and effectiveness of care provided to patients with cancer. Through the leadership and expertise of clinical professionals at NCCN Member Institutions, NCCN develops resources that present valuable information to the numerous stakeholders in the health care delivery system. As the arbiter of high-quality cancer care, NCCN promotes the importance of continuous quality improvement and recognizes the significance of creating clinical practice guidelines appropriate for use by patients, clinicians, and other health care decision-makers. The primary goal of all NCCN initiatives is to improve the quality, effectiveness, and efficiency of oncology practice so patients can live better lives. For more information, visit NCCN.org.

Data Needs in Oncology Work Group Members

Amy Abernethy, MD
Duke Cancer Institute

Peter Bach, MD
Memorial Sloan-Kettering Cancer Center

Mara G. Bloom, JD, MS*
Massachusetts General Hospital Cancer Center

Suzanne Demko, PA-C*
US Food and Drug Administration

Erling Donnelly, PhD
Pfizer Inc.

Elaine Jeter, MD
Palmetto GBA

Joel Kallich, PhD*
Big Health Data
(employee of Amgen Inc. at the time of writing)

Susan Krivacic, MPAFF
PBG Consulting

Kimary Kulig, PhD, MPH*
Bristol-Myers Squibb
(employee of National Comprehensive Cancer Network
at the time of writing)

Jennifer Malin, MD, PhD*
WellPoint, Inc.

Lee N. Newcomer, MD, MHA
UnitedHealthcare

Joyce Niland, PhD
City of Hope Comprehensive Cancer Center

Lee Schwartzberg, MD
The University of Tennessee Health Science Center

Samuel M. Silver, MD, PhD
University of Michigan Comprehensive Cancer Center

Art Small, MD*
Genentech Inc.

NCCN Staff:
Jessica K. DeMartino, PhD
Jonathan K. Larsen, MPP

*Provided editorial support

Disclosure of Affiliations and Significant Relationships

Dr. Abernethy has disclosed that she receives clinical research funding from Alexion Pharmaceuticals, Inc.; Amgen Inc.; Biovex; Bristol-Myers Squibb; DARA; Helsinn; KangLaiTe USA; MiCo; and Pfizer Inc. She is a consultant for Amgen Inc.; Helsinn; Proventys; and Pfizer Inc.

Dr. Bach has disclosed that he has no financial interests, arrangements, or affiliations with the manufacturers of any products or devices discussed in this report or their competitors.

Ms. Bloom has disclosed has no financial interests, arrangements, or affiliations with the manufacturers of any products or devices discussed in this report or their competitors.

Dr. DeMartino has disclosed that she has no financial interests, arrangements, or affiliations with the manufacturers of any products or devices discussed in this report or their competitors. She is an employee of the National Comprehensive Cancer Network.

Ms. Demko has disclosed that she has no financial interests, arrangements, or affiliations with the manufacturers of any products or devices discussed in this report or their competitors.

Dr. Donnelly has disclosed that he and his spouse are employees of Pfizer, Inc.

Dr. Jeter has disclosed that she has no financial interests, arrangements, or affiliations with the manufacturers of any products or devices discussed in this report or their competitors.

Dr. Kallich has disclosed that he is an employee of Amgen Inc.

Ms. Krivacic has disclosed that she has no financial interests, arrangements, or affiliations with the manufacturers of any products or devices discussed in this report or their competitors.

Dr. Kulig has disclosed that she is a Pfizer, Inc. stockholder.

Mr. Larsen has disclosed that he has no financial interests, arrangements, or affiliations with the manufacturers of any products or devices discussed in this report or their competitors. He is an employee of the National Comprehensive Cancer Network.

Dr. Malin has disclosed that she has no financial interests, arrangements, or affiliations with the manufacturers of any products or devices discussed in this report or their competitors.

Dr. Newcomer has disclosed that he is an employee of United Healthcare.

Dr. Niland has disclosed that she has no financial interests, arrangements, or affiliations with the manufacturers of any products or devices discussed in this report or their competitors.

Dr. Schwartzberg has disclosed that he has no financial interests, arrangements, or affiliations with the manufacturers of any products or devices discussed in this report or their competitors.

Dr. Silver has disclosed that he is an employee of BlueCross Blue Shield of Michigan and is a consultant for 3M.

Dr. Small has disclosed that he is an employee of Genentech.

JNCCN—Journal of the National Comprehensive Cancer Network Editorial Board

Editor-in-Chief

Harold J. Burstein, MD, PhD
Dana-Farber Cancer Institute

Associate Editors

Robert W. Carlson, MD
Medical Oncology
National Comprehensive Cancer Network

Wui-Jin Koh, MD
Radiation Oncology
Seattle Cancer Care Alliance

Joan S. McClure, MS
NCCN Guidelines Program
National Comprehensive Cancer Network

Deborah Schrag, MD, MPH
Healthcare Research
Dana-Farber Cancer Institute

Marshall M. Urist, MD
Surgical Oncology
University of Alabama at Birmingham
Comprehensive Cancer Center

Sharon M. Weinstein, MD
Supportive Care
Huntsman Cancer Institute at the
University of Utah

Editorial Board Members

Jaffer A. Ajani, MD
The University of Texas
MD Anderson Cancer Center

Joseph C. Alvarnas, MD
City of Hope Comprehensive
Cancer Center

Kenneth C. Anderson, MD
Dana-Farber/Brigham and Women's
Cancer Center

Lindsey Robert Baden, MD
Dana-Farber/Brigham and Women's
Cancer Center

Tanios Bekaii-Saab, MD
The Ohio State University - James Cancer
Hospital and Solove Research Institute

William Bensinger, MD
Fred Hutchinson Cancer Research Center/
Seattle Cancer Care Alliance

Al B. Benson III, MD
Robert H. Lurie Comprehensive Cancer
Center of Northwestern University

Ann M. Berger, PhD, APRN, AOCNS
UNMC Eppley Cancer Center at
The Nebraska Medical Center

Therese B. Bevers, MD
The University of Texas
MD Anderson Cancer Center

Christopher K. Bichakjian, MD
University of Michigan
Comprehensive Cancer Center

J. Sybil Biermann, MD
University of Michigan
Comprehensive Cancer Center

William Blum, MD
The Ohio State University - James Cancer
Hospital and Solove Research Institute

Patrick A. Brown, MD
The Sidney Kimmel Comprehensive Cancer
Center at Johns Hopkins

Randall W. Burt, MD
Huntsman Cancer Institute at the
University of Utah

Peter R. Carroll, MD, MPH
UCSF Helen Diller Family
Comprehensive Cancer Center

Peter E. Clark, MD
Vanderbilt-Ingram Cancer Center

Peter F. Coccia, MD
UNMC Eppley Cancer Center at
The Nebraska Medical Center

Daniel G. Coit, MD
Memorial Sloan-Kettering Cancer Center

Jeffrey Crawford, MD
Duke Cancer Institute

Michael I. D'Angelica, MD
Memorial Sloan-Kettering Cancer Center

Mary B. Daly, MD, PhD
Fox Chase Cancer Center

Crystal S. Denlinger, MD
Fox Chase Cancer Center

David S. Ettinger, MD
The Sidney Kimmel Comprehensive Cancer
Center at Johns Hopkins

William J. Gradishar, MD
Robert H. Lurie Comprehensive Cancer
Center of Northwestern University

Peter L. Greenberg, MD
Stanford Cancer Institute

Benjamin E. Greer, MD
Seattle Cancer Care Alliance

Charles R. Handorf, MD, PhD
St. Jude Children's Research Hospital/
The University of Tennessee Health Science
Center

Jimmie C. Holland, MD
Memorial Sloan-Kettering Cancer Center

Richard T. Hoppe, MD
Stanford Cancer Institute

Arti Hurria, MD
City of Hope Comprehensive Cancer Center

Gregory P. Kalemkerian, MD
University of Michigan
Comprehensive Cancer Center

Philip W. Kantoff, MD
Dana-Farber Cancer Institute

Ella Kazerooni, MD
University of Michigan
Comprehensive Cancer Center

Wui-Jin Koh, MD
Fred Hutchinson Cancer Research Center/
Seattle Cancer Care Alliance

Matthew H. Kulke, MD
Dana-Farber/Brigham & Women's
Cancer Center

Michael H. Levy, MD, PhD
Fox Chase Cancer Center

James L. Mohler, MD
Roswell Park Cancer Institute

Robert J. Morgan Jr, MD
City of Hope Comprehensive Cancer Center

Robert J. Motzer, MD
Memorial Sloan-Kettering Cancer Center

Louis Burt Nabors, MD
University of Alabama at Birmingham
Comprehensive Cancer Center

Susan O'Brien, MD
The University of Texas
MD Anderson Cancer Center

Margaret R. O'Donnell, MD
City of Hope Comprehensive Cancer Center

Edward E. Partridge, MD
University of Alabama at Birmingham
Comprehensive Cancer Center

David G. Pfister, MD
Memorial Sloan-Kettering Cancer Center

Jerald P. Radich, MD
Fred Hutchinson Cancer Research Center

George M. Rodgers III, MD, PhD
Huntsman Cancer Institute at the
University of Utah

Michael B. Streiff, MD
The Sidney Kimmel Comprehensive
Cancer Center at Johns Hopkins

Robert A. Swarm, MD
Siteman Cancer Center at Barnes-Jewish
Hospital and Washington University School
of Medicine

Margaret A. Tempero, MD
UCSF Helen Diller Family
Comprehensive Cancer Center

Richard Theriault, DO, MBA
The University of Texas
MD Anderson Cancer Center

R. Michael Tuttle, MD
Memorial Sloan-Kettering Cancer Center

Alan P. Venook, MD
UCSF Helen Diller Family
Comprehensive Cancer Center

Margaret von Mehren, MD
Fox Chase Cancer Center

Douglas E. Wood, MD
University of Washington/
Seattle Cancer Care Alliance

Andrew D. Zelenetz, MD, PhD
Memorial Sloan-Kettering Cancer Center

Data Needs in Oncology: “Making Sense of The Big Data Soup”

Jessica K. DeMartino, PhD, and Jonathan K. Larsen, MPP

Abstract

Rising health care costs and continued concerns about safety, efficacy, and quality have resulted in the demand for more data and evidence by payors, regulators, providers, and patients alike. Stakeholders with different objectives for the use of data are driving the need for more and “better” data. It is important for organizations to not only understand how to handle and collect data but also translate it into actionable information that can help transform health care delivery. Appropriate use of data can lead to reduced health care costs and increased value to all stakeholders. In June 2012, NCCN assembled a work group composed of thought leaders from NCCN Member Institutions and other organizations to identify and examine the challenges of data generation, collection, and application for clinical, regulatory, and coverage decision-making. The NCCN Data Needs Work Group identified 4 main areas for consideration: data sources, patient-derived data, payor-collected data, and regulatory policy toward data generation and use. (*JNCCN* 2013;11(Suppl 2):S1–S12)

Executive Summary

Rising health care costs and continued concerns about safety, efficacy, and quality have resulted in the demand for more data and evidence by payors, regulators, providers, and patients alike. Stakeholders with different objectives for the use of data are driving the need for more and “better” data. A common foundation of high-quality data that are available in real-time is a high priority for multiple stakeholders, along with the ability to analyze and use the data for decision-making. Data should be available that can simultaneously be used to improve the delivery of health care, yield measures for quality improvement, provide a foundation for clinical research, and deliver information for health care re-

imbursement and policy decision-making. In an ideal world, data would be acquired and aggregated and, in turn, synthesized into information and knowledge, and finally, wisdom to benefit patients and society.

Possible sources of data include both interventional studies (ie, clinical trials) and noninterventional studies (ie, registries, observational studies, patient-reported outcomes, postmarketing surveillance studies). These noninterventional data sources have varying levels of richness, different time frames for collection, different data collection rules, and varying data abstraction, and are not usually open access. Ideally, an integrated coordinated database would exist that could solve multiple problems simultaneously and provide information that could assist in health care optimization and redesign along with improving value and quality of care.

To identify and examine the challenges of data generation, collection, and application for clinical, regulatory, and coverage decision-making, NCCN assembled a work group composed of thought leaders from NCCN Member Institutions and those representing patients, manufacturers, payors, and government agencies. The NCCN Data Needs Work Group identified key areas of concern and, when possible, identified recommendations for improvement. The NCCN Data Needs Work Group, along with NCCN, recognizes that data have numerous uses in health care, and this report is not meant to be an exhaustive review. For the purposes of discussion, the NCCN Data Needs Work Group agreed to focus on observational and noninterventional sources of data and will touch on other topics as they relate to these sources of data.

The contents of this document represent the work of NCCN and may not necessarily reflect the opinions of the external work group members or the organizations with which they are affiliated. This paper will focus primarily on topics that were identified as important by NCCN and the NCCN Data Needs Work Group and further discussed at the NCCN Data Needs Policy Summit.

Overview and Background

The appropriate collection and use of data is central to oncology research, clinical treatment, and coverage and reimbursement decisions. Different stakeholders and disparate forces are behind the incomprehensible amounts of data currently being collected on a daily basis. The oncology community needs a better understanding of the data currently being collected to develop much-needed analytical tools to advance personalized medicine. It is important for organizations to not only understand how to handle, collect, store, and disseminate these data but also translate it into actionable information that can help transform health care delivery. Appropriate use of data can lead to improved outcomes for patients, reduction of health care costs, and increased value to all stakeholders.

The NCCN Data Needs Work Group was tasked with recognizing and identifying challenges in data generation, collection, and use in oncology. The 4 main areas identified for consideration are data sources, patient-derived data, payor-collected data, and regulatory policy toward data generation and use. The decision was made to focus primarily on noninterventional sources of data and touch on additional topics as they relate to noninterventional sources. Noninterventional data can both complement and sometimes replace other types of data collection, including clinical trials that are not only expensive but also often require years to mature and produce data necessary for making health care decisions.

Noninterventional data used for the purposes of clinical, regulatory, and coverage decision-making come from a variety of sources, including retrospective medical chart reviews, cancer registries, electronic health records (EHRs), administrative systems, patient surveys, and uniform data collection. These data sources have varying levels of richness, different time frames for collection, different

data collection rules, and varying data abstraction, and are not usually open access. EHRs are quickly becoming widespread and adopted as standard care within many doctors' offices and hospitals; however, the data formats and collection of structured versus unstructured data vary widely. They offer data not typically found in disease registries, claims records, or prescription databases, and in some cases may be easier to analyze. Noninterventional data are also frequently used for quality control or performance improvement.

Patient-derived data and, more specifically, patient-reported outcomes (PROs) have garnered more attention in the past few years. PROs provide a way for collecting structured data within the context of clinical practice. PROs can provide understanding and detail regarding the impact of new treatments in both active treatment and supportive care settings. The analysis of PROs in clinical trials has led to labeling claims for new therapeutic agents.¹ Researchers and regulators have cited several methodological issues that must be addressed to improve the measurement and interpretation of PROs.² PROs have the capability to provide data that could be useful for clinical, regulatory, and coverage decision-making.

The challenges of data collection, aggregation, and reporting for regulatory decision-making are diverse and complicated. The FDA must handle large amounts of data in varying formats to make regulatory decisions quickly and effectively. Both drug and device companies are making the move to electronic data capture and standardized data formats, but the transition will take several years unless the FDA issues new requirements. The FDA and industry must continue to work together on data standards and formats to speed the regulatory approval process and ensure that safe interventions reach the public.

Increases in health care costs and continued concerns about safety and quality have resulted in payors demanding additional data, beyond that required by the FDA, to justify reimbursement of interventions. Our country's largest payor, Medicare, uses available data to determine whether an intervention is "reasonable and necessary" for a beneficiary, whereas the data was generated to meet a different standard. Payors often seek information about questions such as comparative effectiveness, suitability in types of patients not studied as part of the FDA trials, and data regarding the safety and effectiveness of

off-label uses. Payers are increasingly using data to determine concordance with national guidelines and specialized oncology pathways. Claims data may be one of the richest sources of data from an administrative perspective. It provides a view of the full episode of care, and payors may use their own claims data to derive information, but most claims data do not include enough clinical information, such as staging or molecular characteristics, to fully inform decision-making regarding coverage decisions or to support research without being augmented with a source of clinical data (eg, cancer registries).

Work Group Description

To identify and examine challenges of data generation, collection, and application, NCCN convened a work group comprising thought leaders from NCCN Member Institutions, payors, manufacturers, government regulatory agencies, and patient groups. The NCCN Data Needs Work Group met on June 18, 2012 in Philadelphia, PA. In addition, NCCN held the NCCN Oncology Policy Summit: Data Needs in Oncology - Clinical, Regulatory, Coverage, and Policy Issues on October 5, 2012 in Washington, DC. This summit included additional thought leaders representing the aforementioned groups and other relevant stakeholders.

There was interest on the part of NCCN and the NCCN Data Needs Work Group in exploring a wide range of topics within the data realm. This document encapsulates the discussion during the work group meeting and at the policy summit, including background on data sources and discussion about patient-derived data, including PROs, payor-collected data, and regulatory policy toward data generation and use.

Data Sources

Data collection is driven by stakeholders' intended uses. Physicians are not collecting data with the same intentions as clinical researchers or payors, resulting in differing qualities and attributes in the data collected by each group of stakeholders. Possible sources of data include both interventional studies (eg, clinical trials) and noninterventional studies (eg, registries, observational studies, PROs, postmarketing surveillance studies, comparative effectiveness studies, claims data, EHRs). A recent

article in the *Journal of Clinical Oncology* detailed administrative databases that provide observational data.³ Common research concepts and methods for overcoming bias in the analysis of observational data were also explained. These data sources have varying levels of richness, different time frames for collection, different data collection rules, and varying data abstraction, and are not usually open access. Ideally, an integrated coordinated data platform would exist that could solve multiple problems simultaneously and provide information that could assist in health care optimization and redesign and improve value and quality of care.

Innovative Data Sources

Several members of the NCCN Data Needs Work Group presented examples of their organizations' innovative data repositories at the policy summit. Descriptions of the data repositories are provided in the following sections.

The Duke Cancer Institute Experience: The mission at the Duke Cancer Institute is to raise the bar and collect research-quality clinical data that improves the understanding of both the clinical and research interface. Collecting PROs is one part of this process and allows for the collection of structured data within the context of clinical practice that reflects who people are and their personal experience. Four key components to collection of PROs include technology, choice of instrument, process integration, and analytics and visualization. Duke integrates PRO data into their data warehouse along with tumor registry, NCCN Outcomes Database, and administrative data. Through a NCI Grand Opportunities (GO) grant, Duke has created a registry that essentially integrates all of their different data sets across their data warehouse and health system. This continuously aggregating data set is able to create annotations for biospecimens. Although this type of registry can integrate all types of data together, the challenge that remains is to develop analytical tools that can make sense of all the data and allow stakeholders to transform data into knowledge.

Avastin Registry Investigation of Effectiveness and Safety: The Avastin Registry Investigation of Effectiveness and Safety (ARIES) study, sponsored by Genentech, is a prospective observational cohort study of patients with metastatic colorectal cancer who are treated with Avastin and chemotherapy. The study, which ended in March 2012, had 2000

patients enrolled who were followed until death or loss to follow-up, with data collected quarterly. AR-IES also includes archival tissue and blood DNA. One of the objectives of the registry was to compare survival outcomes of patients who used Avastin in both first- and second-line treatment with those of patients who used Avastin in first-line treatment and a different chemotherapy for second-line treatment. This registry adds to the understanding of Avastin's real-world use and allows for examination of safety data postlaunch, particularly in patients that may not have been well represented in clinical trials. This type of registry allows for the application of innovative methodology to observational data that may improve precision in measuring exposure outcomes and confounders.

Oncology Services Comprehensive Electronic Records: Amgen, in collaboration with IMS Health, has created a robust EMR data platform for business analytics, safety, clinical development, effectiveness, outcomes, and epidemiologic research. The core of Oncology Services Comprehensive Electronic Records (OSCER) is a data warehouse of roughly 590,000 outpatient oncology clinic patient EMRs with more than 650 variables from approximately 380 medical clinics. The core is linked with health care claims, death records, tumor registry records, and EMR records from other sites of care. OSCER uses a proprietary software package that allows linkages between disparate data sets without violating patient confidentiality and HIPAA regulations. Biospecimens and biomarker test results are being added, because the platform is constantly evolving to capture new medical tests, procedures, and treatments as they are implemented in the practice of medicine. The goal is to sustain a platform that enables the application of real-world data for effectiveness and safety research, and to serve commercial market research needs.

Molecular Diagnostic Services Program: Palmetto GBA, in coordination with the Centers for Medicare & Medicaid Services (CMS), has developed the Molecular Diagnostic Services Program (MolDx) to identify and establish coverage and reimbursement for molecular diagnostic tests in the J1 Medicare Administrative Contractor jurisdiction. The J1 jurisdiction comprises California, Nevada, Hawaii, Guam, American Samoa, and the Northern Marianas. The genesis of this program was the lack of sufficient,

adequate codes; stacked coding; and inappropriate billing for molecular tests. Palmetto GBA was unsure what tests they were reimbursing providers for and were unable to mine the data. MolDx uses the McKesson Diagnostics Exchange module to collect evidentiary and utilization information about the molecular tests that laboratories submit for coverage. Based on the information provided for the technical assessment, Palmetto GBA will apply multiple methodologies appropriate to the specific test to determine an equitable value for each submitted test. Once registered, the laboratory receives either a McKesson Z-code or Palmetto Test Identifier (PTI) that must be entered in the claim line narrative/comment field. If a claim comes in without an identifier on the claim line, it is rejected and no option for appeal is available. The future of this program is currently unknown at the time of publication because a different contractor has received the Medicare administrative contract for the J1 jurisdiction.

National Cancer Databases

There are currently 3 national cancer-related databases: the Centers for Disease Control and Prevention's (CDC) National Program of Cancer Registries (NPCR); the NCI's Surveillance, Epidemiology and End Results (SEER) Program; and the National Cancer Database (NCDB), a joint collaboration between the American College of Surgeons' Commission on Cancer (ACS CoC) and the American Cancer Society. NPCR and SEER are cancer surveillance systems with a primary mission of providing population-based estimates with which to understand the occurrence and distribution of cancer, whereas the NCDB is dedicated to looking at and measuring quality as a baseline for improvement. The data submitted and found in these cancer registries are essentially the same as what comes from the same overlapping set of sources. The registries differ in their foci, use, and subsequent reporting.

The NPCR collects data on the occurrence of cancer; the type, extent, and location of the cancer; and the type of initial treatment.⁴ In each state, medical facilities (including hospitals, physicians' offices, therapeutic radiation facilities, freestanding surgical centers, and pathology laboratories) report these data to a central cancer registry. State cancer registries are designed to monitor cancer trends over time, determine cancer patterns, guide planning and evaluation of cancer control programs, help set funding priorities, advance health services research, and provide information for a

national database of cancer incidence. These registries can be resource-intensive and difficult to maintain and augment. They often have a significant time lag, and registries may have different rules surrounding the data and have different types of data abstractors, including clinical research assistants and certified tumor registrars. Chart abstraction is labor-intensive, inefficient, and prone to error relative to prospective electronic capture of information.

SEER is a source of information on cancer incidence and survival in the United States.⁵ SEER currently collects and publishes cancer incidence and survival data from population-based cancer registries covering approximately 28% of the US population. The SEER Program registries routinely collect data on patient demographics, primary tumor site, tumor morphology and stage at diagnosis, first course of treatment, and follow-up for vital status. The SEER Program is the only comprehensive source of population-based information in the United States that includes stage of cancer at diagnosis and patient survival data. SEER data are also linked to Medicare data for patients aged 65 years and older. Linking the 2 data sets provides information about both initial diagnosis and subsequent treatment, but lacks information on oral medications taken by patients.

The NCDB is a nationwide oncology outcomes database for more than 1500 ACS CoC-accredited cancer programs in the United States and Puerto Rico.⁶ Approximately 70% of all newly diagnosed cases of cancer in the United States are captured at the institutional level and reported to the NCDB. The NCDB contains approximately 26 million records from hospital cancer registries across the United States. These data are used to explore trends in cancer care, create regional and state benchmarks for participating hospitals, and serve as the basis for quality improvement. Data elements are collected and submitted to the NCDB from CoC-accredited cancer program registries using nationally standardized data item and coding definitions, as specified in the CoC's *Facility Oncology Registry Data Standards*, and nationally standardized data transmission format specifications coordinated through the North American Association of Central Cancer Registries. These elements include patient characteristics, cancer staging and tumor histologic characteristics, type of first course treatment administered, and outcomes information.

Alternative Data Sources

Other organizations, including industry, are making efforts to create repositories of data sets. The CEO Roundtable on Cancer has started an initiative for sharing cancer clinical trial data, entitled DataSphere.⁷ This effort aims to create a repository of data sets from cancer clinical trials conducted by drug companies, academic laboratories, and other organizations. The program started with 2 data sets contributed by Sanofi, and organizers hope to make the repository available to outside researchers by April of 2013. Initially, data from comparator arms of clinical trials will be contributed, and proponents estimate that pooling data could cut costs of developing a drug by 10%. This can be seen as a step forward, because pharmaceutical companies have historically been worried about competitive advantage, and academic researchers have been concerned about the ability to publish. In a similar initiative, a group of 10 pharmaceutical companies announced that they would work together through a nonprofit named TransCelerate BioPharma to standardize clinical trial formats and execution in many ways, from patient recruitment to how to record data.⁸

An additional source of high-quality data is the NCCN Oncology Outcomes Database, which is a network-based data collection, reporting, and analytic system that describes the patterns and outcomes of care delivered in the management of patients with cancer. The NCCN Oncology Outcomes Database is composed of 5 database components: breast, colon/rectal, non-small cell lung, and ovarian cancers and non-Hodgkin's lymphomas. More than 300 HIPAA-compliant data elements tracking the continuum of care longitudinally are contained in this database, including complete patient demographics, histories, and characteristics; sites of metastases; sequencing of therapies; reasons for discontinuation of chemotherapy; and progression-free and overall survival. The NCCN Oncology Outcomes Database, which officially stopped collecting new data in February 2013, represents an early gold-standard model that could potentially benefit from modernization, such as the use of aggregating information technology solutions to ensure its relevance to the oncology community.

Electronic Health Records

EHRs are a means of collecting, documenting, and displaying information electronically. The Healthcare Information and Management Systems Society

has defined the essential attributes and requirements of EHRs.⁹ Information collected often includes basic demographics, billing and coding requirements, medications administered, laboratory values, patient symptoms and history, and diagnostic information. EHRs offer data not typically found in disease registries, claims records, or prescription databases. Most EHRs have been designed to support clinical workflow needs, while supporting administrative and financial systems. With the greater amount of data found in EHRs, they are being used increasingly for observational research, for postmarketing safety evaluation, and to inform decision-making.¹⁰ The availability of data varies across EHRs and depends on the design and completeness of data entry. Because EHRs are designed primarily for patient care and billing, details that are important to health research may not be collected as rigorously as is required for all types of research.

There is currently a lack of consensus and standardization in EHRs. The Clinical Oncology Requirements for the EHR (CORE) initiative, a collaboration among ASCO, the NCI Center for Biomedical Informatics and Information Technology, and the NCI Community Cancer Center Program, identified essential elements and features of EHRs for cancer care.¹¹ The Commission for the Certification of Health Information Technology Workgroup transitioned many of these elements into criteria for certification of EHRs for support of cancer care.¹²

The use of EHRs has increased since 2009 because of the Health Information Technology for Economic and Clinical Health Act (HITECH Act). The HITECH Act authorized incentive payments to Medicare- and Medicaid-participating physicians who adopted and used EHR systems. Physicians must show that they are meeting “meaningful use” requirements to receive incentive payments. As of the end of November 2012, \$9.2 billion had been distributed to 177,100 eligible health care providers and hospitals through the meaningful use incentive program. According to a recent CDC Survey, 72% of office-based physicians used EHR systems in 2012.¹³ Many hospitals and large health systems are also using EHR systems either purchased from a vendor and customized for the particular hospital, or have designed and implemented their own system.

The Kaiser Permanente health system provides an example of the uptake of EHRs. Their EHR system, purchased from Epic Systems, cost \$4 billion to

install and took 5 years to fully implement.¹⁴ The system links 37 Kaiser hospitals, 15,857 physicians, and 9 million members. Kaiser completed the gradual switch to an electronic system, but not without some initial objections from doctors, system outages, and periods of low productivity. Kaiser’s experience with EHRs shows the positives of EHRs and the limitations that still exist. The insurer says it has used its data to improve care, but still has the problem of not being able to communicate electronically with other health systems with which their doctors interact. The inability of EHR systems to interoperate can be explained partly by the vast number of vendors and the lack of guidance from the federal government. There are 551 certified medical information software companies in the United States, selling 1137 software programs.¹⁴ Many of these vendors are large, such as General Electric and Epic, but there are also hundreds of niche players. An additional significant obstacle to interoperability stems from clinical documentation practices.¹⁵ Many providers document their clinical notes in prose-like narratives rather than prespecified formats. To most easily use the information entered into an EHR, every piece of information should be codified. This approach would not appeal to most physicians, and a compromise must be found that incorporates both discretized documentation and narrative documentation.

An additional use of EHRs is in accountable care organizations (ACOs). Introduced in the Affordable Care Act by President Obama, ACOs in their most basic form consist of networks of health care facilities and physicians who share responsibility for delivering quality medical care in a cost-effective manner. ACOs have to coordinate patient care across their networks, and EHRs may enable providers within the ACO to more easily share and exchange information, coordinate care, and increase value. In addition to allowing for more streamlined care coordination and individual patient management, the use of EHRs and electronic information exchange within an ACO will allow interventions to be evaluated on a population basis to identify those that are successful and those that are not.

In a survey conducted by Harris Interactive, 85% of Americans have some kind of anxiety related to EHRs. They worry that their digitized health data could be lost, damaged, or corrupted, and that their physician may not be able to access their records dur-

ing a power outage or if computers crash. Worries about data breaches are valid. The Department of Health & Human Services (HHS) has received reports of data breaches that have affected 21 million people's medical records since September 2009.¹⁶ Providers, hospitals, and the government must work to assure patients that their data are safe and are being used appropriately.

Patient-Reported Data

There is growing recognition that PRO measures can complement traditional biomedical outcome measures (eg, overall survival, disease-free survival) in conveying important information about the overall burden of cancer and the effectiveness of interventions.¹⁷ The collection of patient-reported data is a way to accumulate structured data within the context of clinical practice. PROs include information such as symptoms, quality of life, value, satisfaction, medication use, and compliance. PROs have recently garnered more attention in the clinical community because of the increased focus on personalized medicine and the growing interest in bringing the patient's perspective to cancer care decision-making. PROs also offer real-time information to aid provider decision-making that affects clinical care and psychosocial factors. Some examples of outcomes that PROs target include preventing readmissions, reducing depression or anxiety, and ensuring compliance with oral chemotherapy protocols.

PRO measures have multiple potential uses in oncology. One of the primary areas of application is randomized clinical trials. PROs have the ability to add value to randomized clinical trials, in some types of trials more than others. Ganz and Goodwin¹⁸ concluded that PROs can play an important, even pivotal, role in studies in which biomedical outcomes are essentially equivalent. Conversely, PROs may be of secondary importance in comparing interventions that are curative in intent and differ substantially in biomedical outcomes. PROs can also play a pivotal role when studies have not provided evidence of a leading therapeutic option, such as prostate cancer.¹⁹

A second area of application of PRO measures is the regulatory approval process. The FDA PRO Guidance document entitled "Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims" provides advice to drug manufacturers regarding the FDA's preferred approach to conducting clinical trials when an FDA-

approved PRO-based labeling claim is desired.²⁰ The Center for Medical Technology and Policy has released an effectiveness guidance document that provides recommendations for incorporating PROs into the design of clinical trials for adults with cancer.²¹ The authors recognized that, "Patients' subjective experiences constitute information that is essential to any study examining the real-world outcomes of existing treatments or process interventions. PROs offer value added to standard clinical research."²¹

Another potential use of PROs is in clinical practice. There is growing recognition that routine measurement of PROs in oncology practice may improve cancer care planning, monitoring, and management for patients and survivors.¹⁷ The use of PROs in everyday clinical practice can promote better communication and shared decision-making by patients and providers, and help distinguish physical, emotional, and social issues that can be addressed.

The advent of PROs has been largely due to technology that allows for more efficient, reliable, and valid reporting. PROs can be collected by electronic means, including Web-based surveys and hand-held devices. The collection of PROs has improved because of better reporting instruments integrated directly into the care process. Survey questions have been refined and validated. Patients may be asked to complete a questionnaire while waiting to see their physician, or complete a survey at home using the Web. To gather the most relevant data, the data capture process and technology must be matched to where the patient is in terms of their education, illness, and preferred site of care and site of interaction.

Furthermore, the increase in the use of PROs for clinical decision-making is partly because of improved analytics and visualization techniques. Data must be presented to physicians in a format that is effective for how they conduct their practice and interact with patients. This may require trying various visualization formats. Although visualization is critically important, predictive modeling or decision-support related to PROs is also important to trigger recommended interventions. In addition to the use of PROs by providers, patient-reported data can be used by a multidisciplinary care team, banked for a series of aggregated tasks, or integrated into a larger data warehouse.

Challenges to the use of patient-reported data and outcomes in clinical trials do exist.² The validity and reliability of patient-reported data can be chal-

lenged by researchers. Measures must be validated and the psychometrics of PRO data must be constantly checked by a team that includes statisticians, clinicians, clinical trialists, and patients. An additional challenge is the capture and cost of capture for research participants and personnel.

To ensure that patients fully participate in reporting their outcomes, they must understand the value in reporting how they feel, whether they can work and perform self-care tasks, and whether they have been adhering to their treatment regimen. A feedback loop should exist between patients and their physicians so that patients know their self-reported data are being used to improve communication and shared decision-making. It is also important for patients to know whether their data are being used to help make treatment decisions for other patients with cancer. It was stated at the policy summit that we have historically been scientifically and regulatorily focused on objective measures, whereas the patient experience may in fact be the most important measure of quality care.

Regulatory Challenges for Oncology Data

Regulatory agencies across the globe, tasked with ensuring the safety and effectiveness of drugs and medical products, receive data in a variety of formats, often conforming to proprietary industry data standards that may differ from company to company. In practice, this variance in formats adds complexity to the review process for regulatory agencies attempting to maintain efficiency, because each new drug or product submission in a different data format must be considered separately to make sense of the data. Hard copies of data, whether on compact disc or in some other medium, must be analyzed and processed, taking valuable time from the regulatory decision-making process, ultimately slowing the time it takes to get new treatments to the public.

In the United States, the FDA, which does not currently require a standard data format for submission, is experiencing an increase in submissions using Clinical Data Interchange Standards Consortium (CDISC) data, a standardized data format developed by CDISC, a “global, open, multidisciplinary, non-profit organization that has established standards to support the acquisition, exchange, submission and

archive of clinical research data and metadata.”²² This is representative of an increasing trend to standardize medical data formats to begin to address data format disparities that muddle regulatory review.

Since 2008, the FDA’s Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER) have recommended submitting using the electronic Common Technical Document (eCTD), an International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH)–developed standard. The eCTD uses CDISC as its preferred data format for submission. The FDA’s recent draft guidance for industry requiring electronic regulatory submissions using the eCTD format for certain submissions further supports the FDA’s goal of increased regulatory review efficiency.²³

In the context of regulatory review, standardized data formats allow for easier comparison of different studies, and therefore may help increase the efficiency of existing review processes and accuracy of reviews, potentially increasing safety. Although this trend toward standardized data, like CDISC data, may be increasing, industry has not yet fully adopted this format and still needs to change multiple types of reporting forms before it can submit all applications for new drugs or devices to FDA using the CDISC standards.

Further complicating this standardization effort is the regimenting of data within the FDA depending on whether it was acquired from industry for drugs, devices, or diagnostics applications. Data sharing across FDA offices, for example between CDER and the Center for Devices and Radiological Health, requires special permissions that slow data-sharing efforts. During the policy summit, Dr. Andrew von Eschenbach advised that the FDA should consider infrastructure improvements to allow for greater interoperability between divisions.

Data standardization itself is far from simple. For data to be useful across multiple platforms, numerous elements must be standardized, including data dictionaries, vocabulary, and rules for collecting and reporting data. To aid in standardization efforts, the FDA is organizing its data in a data warehouse called Janus that uses CDISC data as its standard to help address some of the issues noted earlier.²⁴ This data warehouse is currently in its testing phase. This data

warehouse will primarily support regulatory decision-making, providing not only data standardization but also analytic tools to speed review.

Although the Janus data warehouse has the potential to increase efficiency of the FDA's review process, companies continue to own the data being analyzed, and the FDA would need permission to publish data or any studies based on the data. Ultimately, it has been suggested that Janus data could be the basis for the control group of a clinical trial, and especially useful in single-arm studies. It may also be possible to de-risk the research and development process for industry through sharing toxicity data using Janus. However, Janus relies on CDISC data, so older study data in need of conversion may remain outside the data warehouse, because it is extremely labor-intensive and expensive to convert non-CDISC data into the standardized format.

During the policy summit, Dr. Amy Abernethy of the Duke Cancer Institute encouraged emulating the data standardization that has made international commerce and banking seamless as data is moved and communicated across the globe. Ideally, this could be a valid model for cancer data, but more work must be performed to realize this goal, because cancer data elements are much more complex than most banking transactions. Projects such as Janus may be the start of a more nimble, responsive data system in medicine that could increase the speed and efficiency of knowledge sharing.

The FDA has also recently released draft guidance for clinical trial data collection that “promotes capturing source data in electronic form, and it is intended to assist in ensuring the reliability, quality, integrity, and traceability of electronic source data.”²⁵ This is another example of the many facets of data standardization that must be considered to increase regulatory review efficiency. It has been suggested that drug development could be expedited if data from clinical trials could be combined and analyzed through a trusted third party.²⁶ However, as noted earlier, these data are proprietary and permission would need to be obtained to use it in this way.

Another issue that arises is that the regulatory evidence requirements from the FDA may conflict with payors' evidence objectives and could potentially burden innovation in drug development.²⁷ Drug developers must manage both regulatory and payor data requirements to ensure that new treatments

make it to market and are reimbursed. Data must show that a new treatment will improve outcomes, and payors must agree with the data to reimburse for the new treatment. Doctors must be comfortable with data before prescribing a new intervention, and patients, in turn, must feel comfortable taking it. Overcoming the regulatory hurdle alone does not ensure that a new treatment will be used in practice, partly because of these competing data needs.

Oncology Data Challenges for Payors

All stakeholders have different objectives for the data that they collect. For payors, the data elements needed to determine whether a treatment should be reimbursed are the most critical. Payor data have traditionally been of 3 kinds: claims data, pharmacy data, and administration data (from prior authorization). Other data elements may be collected by payors, but are not likely to be as complete as the critical reimbursement data, and are likely not as comprehensive as a medical chart. Often staging, biomarker status, outcomes, and patient preference data will be absent from payor-collected data. However, claims data have unique advantages. For example, claims data will follow an insured patient wherever treatment is sought. An individual hospital or provider's patient data may only provide a snapshot of one segment of a patient's care, because care may be sought in multiple settings. Payor claims data will follow an insured patient across their health care experience. However, payor data aims to capture financial transactions, which will likely not provide a complete picture of the health care experience.

Payors vary in what data they collect beyond claims data. For example, Wellpoint plans to collect stage data when approving imaging requests and plans for more robust preauthorization for cancer therapy at a regimen level, and will collect clinical status and biomarker status data as part of this program. United Healthcare (UHC), the largest private payor in the United States, has collected stage, biomarker, and current status as a voluntary process for breast, colon, lung, and prostate cancers. In the past 3 years, this has accounted for data from more than 34,000 patients. Like Wellpoint, UHC will start to collect staging and biomarker status data during the preauthorization process for cancer therapy. UHC has developed algorithms to help decipher lines

of therapy, progression-free survival, and survival. Challenges remain when multiple cancer diagnoses are involved, when processing codes for diagnostic tests, and in analyzing pharmacy data, which often have nonuniform facility-based revenue codes and nonstandard reporting procedures for pharmacy benefit managers.

As the setting of patient care has changed, the sources of billing data have also changed, causing variations in the reporting and quality of claims data. Although the ability to combine data collected by a physician with data collected by a pharmacist may be improving, claims data collected across different settings will often have missing data, preventing it from being a complete accurate account of the health care experience. As mentioned earlier, UHC is addressing this gap by collecting additional information, such as stage, status, adjuvant therapy, metastatic state, and relevant genetic markers, through voluntary processes and new preauthorization requirements. With these additional data, UHC hopes to address issues regarding quality gaps, utilization management, and health services research.

Tumor registries may function as an important data source to augment claims data, especially for diagnosis, treatment, and staging information. Discussion during the policy summit suggests that more payors are looking to collaborate with tumor registries to address areas of missing data, particularly survival data. Improving the quality of data allows payors to make better-informed choices, which impact the cost of care and factor into efforts to improve the safety and quality of care. Mortality data sources, such as credit agencies or the National Death Index Status Files, may also serve an important role in completing data.

Payors often collect data to gain a better understanding of the variation in health care. They wish to identify outliers and the source of variation, may it be perverse incentives, knowledge gaps, or something else. Claims and other data help increase understanding of how procedures are being used, and that knowledge can be used to impact health care decisions. However, the quality and validity of payor-collected data, even when augmented by other data sources, may not be useable in outcomes research if too many pieces of data are missing. Criticisms of the use of payor data for research arose at the policy summit. The rigorousness and robustness of

the data was questioned along with the use of proxy sources and algorithms to account for missing data. Despite this, it is critical to use even imperfect payor-collected data to manage trends in the rapidly evolving oncology space and health care more broadly. Lee Newcomer of UHC supported the use of imperfect data because of pending financial collapse discussed in a recent *Annals of Family Medicine* article.²⁸ The research suggests that we are on track for insurance premiums to surpass average household income by 2037, and that private health insurance will become increasingly unaffordable to low-to-middle-income Americans unless major changes are made in the US health care system. Payors may argue that needed changes to prevent this scenario can only occur using their imperfect data.

Another payor effort, pioneered by the CMS, is to allow for coverage of promising technologies that may not pass Medicare's "reasonable and necessary" standard because of a lack of evidence. This "coverage with evidence development" (CED) process provides a mechanism to increase data and evidence about the safety and effectiveness of new technologies that allow patients to access potentially beneficial care that without coverage would not be an option.²⁹ This may also allow for comparative effectiveness research that payors have not typically engaged in, as more data are gathered and processed on competing procedures. CED has been used in at least 2 key clinical areas in oncology: PET scans and drugs for the treatment of colorectal cancer.^{30,31} An additional effort by a CMS contractor was detailed earlier. Palmetto GBA established MolDx to better-identify molecular and genomic tests and determine coverage and reimbursement. Medicare itself is limited in how it can use data it collects, although its health care data has proven to be some of the most useful available for others to examine health care trends (ie, SEER/Medicare data).

Ultimately, it is important for payors to make effective use of data they are collecting to improve the quality of the health care delivery system. In many cases, using third-party data to augment existing data may help improve efficiency and may increase the safety and quality of care being delivered.

Closing Statement

The NCCN Data Needs Work Group, along with others, have identified numerous challenges re-

lated to collecting and using data and synthesizing knowledge that were not detailed within this report. A recent Institute of Medicine forum focused on the informatics needs and challenges in cancer research.³² It was noted that clinical data systems and research data systems do not routinely interoperate. Ideally, these 2 sources of data would be linked or at least have the ability to be linked to complete studies. Technical obstacles identified include the sheer quantity and nonuniformity of the data being collected.³³ Currently, too many standards exist to choose from and none are universally adopted. Additionally, the absence of incentives to build interconnected systems is a limiting factor in encouraging collaboration.

The NCCN Data Needs Work Group discussed a variety of noninterventional data sources, including registries, EHRs, patient-reported data, and payors' claims data. Each data source has inherent strengths and weaknesses, including varying levels of richness, different time frames for collection, different data collection rules, varying data abstraction, and varying rules on public access. Noninterventional data can complement other types of data collection, including clinical trials in an expedited, less-expensive manner. The oncology community must understand the wealth of data that are currently being collected, work to limit the amount of redundancy found in the current data systems, and strive to form collaborative relationships that use current data sources.

Patient-derived data and, more specifically, PROs have garnered more attention over the past few years, because they can provide understanding and detail regarding the impact of new treatments in both active treatment and supportive care settings. PRO measures have multiple potential uses in oncology, including in randomized clinical trials, regulatory approvals, and clinical care. To gather the most relevant data, the data capture process and technology must be matched to where the patient is in terms of their education, illness, and preferred site of care and site of interaction.

The challenges of data collection, aggregation, and reporting for regulatory decision-making are numerous. The FDA must balance the processing of large amounts of data in varying formats with the need to make regulatory decisions quickly and effectively. Although industry is beginning to standardize its data formats, such as using CDISC, the tran-

sition may take several years, depending on future FDA data requirements. The FDA and industry must continue to work together on data standards and formats and on projects, such as the Janus clinical trial repository, to speed the regulatory approval process and ensure that safe interventions reach the public.

Payors are collecting additional information, such as stage, status, adjuvant therapy, metastatic state, and relevant genetic markers through voluntary processes and new preauthorization requirements. With this additional data, payors hope to address issues regarding quality gaps, use management, and health services research. It is important for payors to make effective use of data they are collecting to improve the quality of the health care delivery system. Using third-party data, such as state cancer registries, to augment existing data may help improve efficiency and increase the safety and quality of care being delivered.

Data must go beyond simply its acquisition, and turn information into actionable knowledge that will positively impact clinical care. Data should be moved from silos of information to an integrated system. Innovation must occur to create more capacity and capability to analyze and apply data in ways that improve patient care and increase the value of care. The oncology community must embrace the idea of collaboration to combine available sources of data and improve the current health care system.

References

1. Arnould B, Caron M, Emery MP, et al. Do patient-reported outcomes contribute to regulatory decisions in the USA and Europe? A systematic review of guidance documents and authorizations of medicinal products from 2006 to 2010. Available at: <https://smdm.confex.com/smdm/2011ch/webprogram/Paper6550.html>. Accessed March 21, 2013.
2. Garcia SF, Cella D, Clauser SB, et al. Standardizing patient-reported outcomes assessment in cancer clinical trials: a patient-reported outcomes measurement information system initiative. *J Clin Oncol* 2007;25:5106–5112.
3. Hershman DL, Wright JD. Comparative effectiveness in oncology methodology: observation data. *J Clin Oncol* 2012;30:4215–4222.
4. National Program of Cancer Registries Web site. About the program. Available at: <http://www.cdc.gov/cancer/npcr/about.htm>. Accessed March 21, 2013.
5. Surveillance, Epidemiology, and End Results Web site. Overview of the SEER program. Available at: <http://seer.cancer.gov/about/overview.html>. Accessed March 21, 2013.
6. American College of Surgeons Web site. National Cancer Data Base. Available at: <http://www.facs.org/cancer/ncdb/index.html>. Accessed March 21, 2013.

Data Needs in Oncology

7. Bhattacharjee Y. Pharma firms push for sharing of cancer trial data. *Science* 2012;338:29.
8. FierceMarkets PharmaTech Web site. 10 pharma giants form nonprofit to gang up on common R&D foes. Available at: <http://www.fiercebiotech.com/story/10-pharma-giants-form-nonprofit-gang-common-rd-foes/2012-09-19>. Accessed March 21, 2013.
9. Handler J, Holtmeier R, Metzger J, et al. HIMSS electronic health record definition model Version 1.0. Available at: http://www.pranagy.com/uploads/HIMSS_EMR_Definition_Model_v1-0.pdf. Accessed March 21, 2013.
10. Dean BB, Lam J, Natoli JL, et al. Use of electronic medical records for health outcomes research: a literature review. *Med Care Res Rev* 2010;66:611–638.
11. Clinical Oncology Requirements for the EHR. Available at: <http://www.asco.org/ASCOv2/Department%20Content/Cancer%20Policy%20and%20Clinical%20Affairs/CORE%20white%20paper%20100609.pdf>. Accessed March 21, 2013.
12. Yu PP. The evolution of oncology electronic health records. *Cancer J* 2011;17:197–202.
13. Hsiao CJ, Hing E. Use and characteristics of electronic health record systems among office-based physician practices: United States 2001–2012. *NCHS Data Brief* 2012;(111):1–8.
14. Leonard D, Tozzi J. Why don't more hospitals use electronic health records? Available at: <http://www.businessweek.com/articles/2012-06-21/why-dont-more-hospitals-use-electronic-health-records>. Accessed March 26, 2013.
15. Miriovsky BJ, Shulman LN, Abernethy AP. Importance of health information technology, electronic health records, and continuously aggregating data to comparative effectiveness research and learning health care. *J Clin Oncol* 2012;30:4243–4248.
16. Lowes R. Fear of data theft blunts public acceptance of EHRs. Available at: <http://www.medscape.com/viewarticle/769778>. Accessed August 27, 2012.
17. Lipscomb J, Gotay CC, Snyder CF. Patient-reported outcomes in cancer: a review of recent research and policy initiatives. *CA Cancer J Clin* 2007;57:278–300.
18. Ganz PA, Goodwin PJ. Quality of life in breast cancer: what have we learned and where do we go from here. In: Lipscomb J, Gotay CC, Snyder C, eds. *Outcomes Assessment in Cancer: Measures, Methods, and Applications*. Cambridge, UK: Cambridge University Press; 2005:93–126.
19. Litwin MS, Talcott JA. Measuring quality of life in prostate cancer: progress and challenges. In: Lipscomb J, Gotay CC, Snyder C, eds. *Outcomes Assessment in Cancer: Measures, Methods, and Applications*. Cambridge, UK: Cambridge University Press; 2005:160–178.
20. U.S. Food and Drug Administration Web site. Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Available at: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf>. Accessed March 26, 2013.
21. Center for Medical Technology Policy Web site. Recommendations for Incorporating Patient-Reported Outcomes into the Design of Clinical Trials in Adult Oncology. Available at: <http://www.cmpnet.org/wp-content/uploads/downloads/2012/05/PRO-EGD.pdf>. Accessed March 26, 2013.
22. Clinical Data Interchange Standards Consortium Web site. Mission & Principles. Available at: <http://www.cdisc.org/mission-and-principles>. Accessed March 26, 2013.
23. Draft revision of guidance for industry on providing regulatory submissions in electronic format—certain human pharmaceutical product applications and related submissions using the electronic common technical document specifications; availability. *Fed Regist* 2013;78:310–311
24. U.S. Food and Drug Administration Web site. Janus Clinical Trials Repository Project. Available at: <http://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/ucm155327.htm>. Accessed March 26, 2013.
25. U.S. Food and Drug Administration Web site. Guidance for Industry. Electronic Source Data in Clinical Investigations. Available at: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM328691.pdf>. Accessed March 26, 2013.
26. Sandburg B. Clinical Trial Regs: FDA Seeks Advice On How to Make Them Better, Faster, Stronger. *The Pink Sheet*, Vol. 74, No. 13, March 26, 2012.
27. Kelly C. “Blossoming” Evidence Demands For Drugs A Burden on Innovation, Pfizer Exec Says. *The Pink Sheet*, Vol. 74, No. 14, April 2, 2012.
28. Young RA, DeVoe JE. Who will have health insurance in the future? An updated projection. *Ann Fam Med* 2012;10:156–162.
29. Tunis SR, Pearson SD. Coverage options for promising technologies: Medicare's ‘coverage with evidence development’. *Health Aff (Millwood)* 2006;25:1218–1230.
30. Hillner BE, Siegel BA, Liu D, et al. Impact of positron emission tomography/computed tomography and positron emission tomography (PET) alone on expected management of patients with cancer: initial results from the national oncologic PET registry. *J Clin Oncol* 2008;26:2155–2161.
31. Carino T, Williams RD II, Colbert AM, Bridger P. Medicare's coverage of colorectal cancer drugs: a case study in evidence development and policy. *Health Aff (Millwood)* 2006;25:1231–1239.
32. Nass SJ, Wizemann T. National Cancer Policy Forum, Board on Health Care Services, Institute of Medicine. *Informatics Needs and Challenges in Cancer Research: Workshop Summary*. Washington, DC: The National Academies Press; 2012.
33. Kean MA, Abernethy AP, Clark AM, et al. Achieving data liquidity in the cancer community: proposal for a coalition of all stakeholders. Washington, DC: Institute of Medicine of the National Academies; 2012. Available at: <http://www.iom.edu/~media/Files/Perspectives-Files/2012/Discussion-Papers/NCPF-Achieving-Data-Liquidity.pdf>. Accessed March 26, 2013.