



# **Examining the Impact of Real-World Evidence on Medical Product Development: A Workshop Series**

## **Workshop 3: Application**

***Workshop Briefing Materials***



# **Table of Contents**

1. Workshop Information
2. Real-World Evidence Workshop Series Background Document
3. Forum on Drug Discovery, Development, and Translation Background Document  
(*Parent Activity*)
4. Workshop Agenda
5. Workshop Series Planning Committee Member Biographies
6. Workshop Speaker and Discussant Biographies
7. Literature Informing the Discussion of Draft Decisions Aids at the Workshop:
  - a. Establishing a Framework to Evaluate Real-World Endpoints (Friends of Cancer Research, July 10, 2018)
  - b. The Salford Lung Study Protocol: A Pragmatic, Randomised Phase III Real-World Effectiveness Trial in Chronic Obstructive Pulmonary Disease (Bakerly et al., 2015)
  - c. Effectiveness of Fluticasone Furoate–Vilanterol for COPD in Clinical Practice (Vestbo et al. for the Salford Lung Study Investigators, 2016)
  - d. High-Dose Influenza Vaccine to Reduce Clinical Outcomes in High-Risk Cardiovascular Patients: Rationale and Design of the INVESTED Trial (Vardeny et al., 2018)



- e. Association of Initiation of Basal Insulin Analogs vs. Neutral Protamine Hagedorn Insulin with Hypoglycemia-Related Emergency Department Visits or Hospital Admissions and with Glycemic Control in Patients with Type 2 Diabetes (Lipska et al., 2018)
  - f. Clozapine Treatment for Suicidality in Schizophrenia: International Suicide Prevention Trial (InterSePT) (Meltzer et al., 2003)
  - g. Design and Rationale of the Paliperidone Palmitate Research in Demonstrating Effectiveness (PRIDE) Study: A Novel Comparative Trial of Once-Monthly Paliperidone Palmitate Versus Daily Oral Antipsychotic Treatment for Delaying Time to Treatment Failure in Persons with Schizophrenia (Alphs et al., 2014)
  - h. Real-World Outcomes of Paliperidone Palmitate Compared to Daily Oral Antipsychotic Therapy in Schizophrenia: A Randomized, Open-Label, Review Board–Blinded 15-Month Study (Alphs et al., 2015)
  - i. Effectiveness and Duration of Protection Provided by the Live-Attenuated Herpes Zoster Vaccine in the Medicare Population Ages 65 Years and Older (Izurieta et al., 2017)
8. Discussion Draft Decision Aids (*Specific to each workshop session*):
- a. Session II: When is a Real-World Data Element Fit for Assessment of Eligibility, Treatment Exposure, or Outcome? (*2 pages*)
  - b. Session III: Obscuring Intervention Allocation in Trials to Generate Real-World Evidence: Why, Who, and How?
  - c. Session IV: How Tightly Should Investigators Attempt to Control or Restrict Treatment Quality in a Pragmatic or “Real World” Trial?
  - d. Session V: How Can Bias in Observational Comparisons be Assessed and Minimized?
9. Proceedings from Previous Workshops in the Real-World Evidence Series
- a. Series Workshop 1: Incentives
  - b. Series Workshop 2: Practical Approaches



## **Workshop Information**

### **Dates:**

July 17-18, 2018

### **Time:**

Day 1: 8:15am – 5:00pm

Day 2: 8:00am – 12:00pm

### **Location:**

National Academy of Sciences Building  
Lecture Room  
2101 Constitution Ave NW  
Washington, DC 20418

### **Website:**

<http://nationalacademies.org/hmd/Activities/Research/DrugForum/2018-July-17.aspx>



## Examining the Impact of Real-World Evidence on Medical Product Development: A Three-Part Workshop Series

Washington, D.C.

### Background and Objectives

Randomized, controlled clinical trials (RCTs) have traditionally served as the gold standard for evidence generation in support of medical product development and approval. However, it is increasingly recognized that RCTs have inherent limitations, particularly with regard to generalizability, and time and monetary investment. Data from sources supplemental to RCTs, such as safety surveillance, observational studies, registries, claims, or patient-centered outcomes research, would be valuable to support biomedical research, including medical product development and evaluation.

This three-part workshop series will provide a format for examining the practicalities of collection of data from such real-world sources and deriving real-world evidence for the evaluation of medical products, including drugs, biologics, and devices. Each 1.5 day workshop will include presentations and perspectives from thought and knowledge leaders representing a range of disciplines, including but not limited to federal regulatory and funding agencies, clinical and academic medicine and research, medical professional organizations, the regulated biopharmaceutical industry, patients and patient-focused and disease-advocacy organizations, payers, consumer organizations, health systems, and other interested stakeholders that represent the myriad views of those involved in drug, biologic, and device discovery, development, translation, and regulation. The workshop audiences are expected to be similarly diverse, and they will have opportunities to engage in discussion during the workshops. The series will employ case studies to illustrate the current state and to illuminate potential ways forward; staff or invited experts will prepare background papers describing the characteristics of, and gaps in, current data generation efforts. Thought leaders will be invited to react to and build on the papers.

### Workshop Topics and Flow

- **Workshop One** (*September 19-20, 2017*) focused on how to align incentives to support collection and use of real-world evidence in health product review, payment, and delivery. Incentives need to address barriers impeding the uptake of real-world evidence, including barriers to transparency.
- **Workshop Two** (*March 6-7, 2018*) was a “town-hall” style meeting with active audience participation to illuminate what types of data are appropriate for what specific purposes and to suggest approaches for data collection and evidence use by developing and working through example use cases.
- **Workshop Three** (*July 17-18, 2018*) will examine and suggest approaches for operationalizing the collection and use of real-world evidence.

### Planning Committee

**Mark McClellan (Co-Chair)**, Duke-Margolis Center for Health Policy  
**Gregory Simon (Co-Chair)**, Kaiser Permanente Washington Health Research Institute  
**Jeff Allen**, Friends of Cancer Research  
**Andrew Bindman**, UCSF  
**Adam Haim**, NIMH, NIH  
**Michael Horberg**, Kaiser Permanente Mid-Atlantic Medical Group  
**John Graham**, GlaxoSmithKline  
**Richard Kuntz**, Medtronic, Inc.  
**Elliott Levy**, Amgen Inc.  
**David Madigan**, Columbia University  
**Deven McGraw**, Ciitizen  
**Richard Platt**, Harvard Medical School  
**Joanne Waldstreicher**, Johnson&Johnson  
**Marcus Wilson**, HealthCore, Inc

# ABOUT THE FORUM



Forum on  
**DRUG DISCOVERY, DEVELOPMENT,  
and TRANSLATION**

The Forum on Drug Discovery, Development, and Translation of the National Academies of Sciences, Engineering, and Medicine was created in 2005 by the Board on Health Sciences Policy to provide a unique platform for dialogue and collaboration among thought leaders and stakeholders in government, academia, industry, foundations, and patient advocacy with an interest in improving the system of drug discovery, development, and translation. The Forum brings together leaders from private sector sponsors of biomedical and clinical research, federal agencies sponsoring and regulating biomedical and clinical research, the academic community, and patients, and in doing so serves to educate the policy community about issues where science and policy intersect. The Forum convenes several times each year to identify, discuss, and act on key problems and strategies in the discovery, development, and translation of drugs. To supplement the perspectives and expertise of its members, the Forum also holds public workshops to engage a wide range of experts, members of the public, and the policy community. The Forum also fosters collaborations among its members and constituencies. The activities of the Forum are determined by its members, focusing on the major themes outlined below.

## **INNOVATION AND THE DRUG DEVELOPMENT ENTERPRISE**

Despite exciting scientific advances, the pathway from basic science to new therapeutics faces challenges on many fronts. New paradigms for discovering and developing drugs are being sought to bridge the ever-widening gap between scientific discoveries and translation of those discoveries into life-changing medications. There is also increasing recognition of the need for new models and methods for drug development and translational science, and “precompetitive collaborations” and other partnerships, including public-private partnerships, are proliferating. The Forum offers a venue to discuss effective collaboration in the drug discovery and development enterprise and also hosts discussions that could help chart a course through the turbulent forces of disruptive innovation in the drug discovery and development “ecosystem.”

Key gaps remain in our knowledge about science, technology, and methods needed to support drug discovery and development. Recent rapid advances in innovative drug development science present opportunity for revolutionary developments of new scientific techniques, therapeutic products, and applications. The Forum provides a venue

to focus ongoing attention and visibility to these important drug development needs and facilitates exploration of new approaches across the drug development lifecycle. The Forum has held workshops that have contributed to the defining and establishment of regulatory science and have helped inform aspects of drug regulatory evaluation.

## **CLINICAL TRIALS AND CLINICAL PRODUCT DEVELOPMENT**

Clinical research is the critical link between bench and bedside in developing new therapeutics. Significant infrastructural, cultural, and regulatory impediments challenge efforts to integrate clinical trials into the health care delivery system. Collaborative, cross-sector approaches can help articulate and address these key challenges and foster systemic responses. The Forum has convened a multiyear initiative to examine the state of clinical trials in the United States, identify areas of strength and weakness in our current clinical trial enterprise, and consider transformative strategies for enhancing the ways in which clinical trials are organized and conducted. In addition to sponsoring multiple symposia and workshops, under this initiative, the Forum is fostering innovative, collaborative efforts to facilitate needed change in areas such as improvement of clinical trial site performance.

## **INFRASTRUCTURE AND WORKFORCE FOR DRUG DISCOVERY, DEVELOPMENT, AND TRANSLATION**

Considerable opportunities remain for enhancement and improvement of the infrastructure that supports the drug development enterprise. That infrastructure, which includes the organizational structure, framework, systems, and resources that facilitate the conduct of biomedical science for drug development, faces significant challenges. The science of drug discovery and development, and its translation into clinical practice, is cross-cutting and multidisciplinary. Career paths can be opaque or lack incentives such as recognition, career advancement, or financial security. The Forum has considered workforce needs as foundational to the advancement of drug discovery, development, and translation. It has convened workshops examining these issues, including consideration of strategies for developing a discipline of innovative regulatory science through the development of a robust workforce. The Forum will also host an initiative that will address needs for a workforce across the translational science lifecycle.

## Forum on Drug Discovery, Development, and Translation

---

**Russ Altman (Co-Chair)**  
Stanford University

**Robert Califf (Co-Chair)**  
Duke University and  
Verily Life Sciences

**Christopher Austin**  
National Center for Advancing  
Translational Sciences, NIH

**Linda Brady**  
National Institute of Mental Health,  
NIH

**Tanisha Carino**  
FasterCures, Milken Institute

**Richard Davey**  
National Institute of Allergy and  
Infectious Diseases, NIH

**James Doroshow**  
National Cancer Institute, NIH

**Jeffrey Drazen**  
*New England Journal of Medicine*

**Steven Galson**  
Amgen Inc.

**Carlos Garner**  
Eli Lilly and Company

**Julie Gerberding**  
Merck & Co., Inc.

**Lynn Hudson**  
Critical Path Institute

**S. Claiborne (Clay) Johnston**  
University of Texas, Austin

**Gregory Keenan**  
AstraZeneca

**Katharine Knobil**  
GlaxoSmithKline

**Freda Lewis-Hall**  
Pfizer Inc.

**Allison McElvaine**  
American Diabetes Association

**Ross McKinney**  
Association of American Medical  
Colleges

**Joseph Menetski**  
Foundation for the NIH

**Bernard Munos**  
InnoThink Center for Research in  
Biomedical Innovation

**Michael Severino**  
AbbVie, Inc.

**Rachel Sherman**  
Office of the Commissioner,  
U.S. FDA

**Ellen Sigal**  
Friends of Cancer Research

**Lana Skirboll**  
Sanofi

**Brian Strom**  
Rutgers, The State University of New  
Jersey

**Amir Tamiz**  
National Institute of Neurological  
Disorders and Stroke, NIH

**Pamela Tenaerts**  
Clinical Trials Transformation  
Initiative

**John Wagner**  
Takeda Pharmaceuticals

**Joanne Waldstreicher**  
Johnson & Johnson

**Carrie Wolinetz**  
National Institutes of Health, Office  
of Science Policy

**Janet Woodcock**  
Center for Drug Evaluation and  
Research, U.S. FDA

## Project Staff

---

**Carolyn Shore, Ph.D.**  
Forum Director  
cshore@nas.edu  
202-334-2669

**Rebecca English, M.P.H.**  
Program Officer

**Amanda Wagner Gee, M.S.**  
Program Officer

**Melvin Joppy**  
Senior Program Assistant

**For more information, please visit:**  
[NATIONALACADEMIES.ORG/DRUGFORUM](https://www.nationalacademies.org/drugforum)

Health and Medicine Division  
Board on Health Sciences Policy

*The National Academies of*  
SCIENCES • ENGINEERING • MEDICINE





## Examining the Impact of Real-World Evidence on Medical Product Development: A Three-Part Workshop Series

### Workshop Three: Application

July 17 – 18, 2018

National Academy of Sciences Building, Lecture Room  
2101 Constitution Ave. NW, Washington, DC 20418

The National Academies of Sciences, Engineering, and Medicine (National Academies) is convening a three-part workshop series, sponsored by FDA, examining how real-world evidence development and uptake can enhance medical product development and evaluation. The workshops will advance discussions and common knowledge about complex issues relating to the generation and utilization of real-world evidence, including fostering development and implementation of the science and technology of real-world evidence generation and utilization.

- [Workshop One](#) (*September 19-20, 2017*) focused on how to align incentives to support collection and use of real-world evidence in health product review, payment, and delivery. Incentives need to address barriers impeding the uptake of real-world evidence, including barriers to transparency.
- [Workshop Two](#) (*March 6-7, 2018*) illuminated what types of data are appropriate for what specific purposes and suggested practical approaches for data collection and evidence use by developing and working through example use cases.
- [Workshop Three](#) (*July 17-18, 2018*) will examine and suggest approaches for operationalizing the collection and use of real-world evidence through discussing and revising “decision aids” about specific topics in study design. The decision aids will be question lists to help inform stakeholders about study design choices, including potential risks, costs, and reporting/transparency expectations.

## DAY 1: July 17, 2018

8:00 a.m. Breakfast available outside the Lecture Room

8:15 a.m. **Welcome and opening remarks**  
MARK MCCLELLAN, *Workshop Series Co-Chair*  
Director  
Duke-Margolis Center for Health Policy

GREGORY SIMON, *Workshop Series Co-Chair*  
Investigator  
Kaiser Permanente Washington Health Research Institute



---

## SESSION I KEY CONSIDERATIONS FOR REAL-WORLD EVIDENCE APPLICATION

---

### Session Objectives:

- Examine how some organizations are currently considering traditional and real-world evidence.
- Discuss factors that may be influencing overall cost and time investment required by traditional evidence generation.
- Consider when nontraditional datasources may be beneficial to assess outcomes.

8:45 a.m.      **Update on IMI's GetReal and view from NICE**  
PALL JONSSON  
Associate Director, Research and Development  
National Institute for Health and Care Excellence

9:05 a.m.      **Drivers of expense and delay**  
ELLIOTT LEVY  
Senior Vice President, Global Development  
Amgen, Inc.

9:25 a.m.      **Patient-collected and owned data**  
KOMATHI STEM  
Chief Executive Officer and Founder  
monARC Bionetworks

9:45 a.m.      **BREAK**

---

## SESSION II WHEN IS A REAL-WORLD DATA ELEMENT FIT FOR ASSESSMENT OF ELIGIBILITY, TREATMENT EXPOSURE, OR OUTCOMES?

---

### Session Objectives:

- Discuss potential bias-introducing steps in evidence generation from real-world data.
- Suggest key considerations in the data collection and evidence generation processes that influence reliability of RWD.
- Discuss how a decision aid laying out key questions and considerations might help inform current and future studies.

10:05 a.m.      **Introduction: A proposed framework for a decision aid**  
PALL JONSSON, *Session Moderator*  
Associate Director, Research and Development  
National Institute for Health and Care Excellence

10:15 a.m.      **Looking back: How might a decision aid inform a real-world example?**  
JEFF ALLEN  
President and Chief Executive Officer  
Friends of Cancer Research

10:35 a.m. **Looking forward: How decision aid might apply to future studies?**

***Panel discussion and audience Q&A***

AYLIN ALTAN

Senior Vice President of Research  
OptumLabs

ROBERT BALL

Deputy Director, Office of Surveillance and Epidemiology  
Center for Drug Evaluation and Research  
U.S. Food and Drug Administration

LUCA FOSCHINI

Co-founder and Chief Data Scientist  
Evidation Health

BRANDE YAIST

Sr. Director, Global Patient outcomes and Real-World Evidence  
Eli Lilly and Company

12:00 p.m. **BREAK** (Lunch available Outside the Lecture Room)

---

**SESSION III    OBSCURING INTERVENTION ALLOCATION IN TRIALS TO GENERATE REAL-WORLD EVIDENCE: WHY, WHO, AND WHEN?**

---

Session Objectives:

- Discuss how variability in knowledge of treatment assignment group affects
  - Provider and patient adherence and outcomes
  - Study cost and reliability.
- Suggest key factors that could affect decisions to obscure intervention allocation.
- Discuss how a decision aid laying out key questions and considerations might help inform current and future studies.

1:00 p.m. **Introduction: A proposed framework for a decision aid**

JONATHAN WATANABE, *Session Moderator*

Associate Professor of Clinical Pharmacy

National Academy of Medicine Anniversary Fellow in Pharmacy

University of California San Diego

1:10 p.m. **Looking back: How might a decision aid inform a real-world example?**

JOHN GRAHAM (*invited*)

Head, Value Evidence and Outcomes

GlaxoSmithKline

ORLY VARDENY

Minneapolis VA Center for Chronic Disease outcomes Research

Associate Professor of Medicine

University of Minnesota

1:30 p.m. **Looking forward: How decision aid might apply to future studies?**

***Panel discussion and audience Q&A***

CATHY CRITCHLOW

Vice President, Center for Observational Research  
Amgen, Inc.

NANCY DREYER

Chief Scientific Officer  
IQVIA

ALEX JOHN LONDON

Clara L. West Professor of Ethics and Philosophy  
Carnegie Mellon University

JAMES P. SMITH

Deputy Director, Division of Metabolism and Endocrinology Products  
Center for Drug Evaluation and Research  
U.S. Food and Drug Administration

2:50 p.m. **BREAK**

---

**SESSION IV HOW TIGHTLY SHOULD INVESTIGATORS ATTEMPT TO CONTROL OR RESTRICT TREATMENT QUALITY IN A PRAGMATIC OR REAL-WORLD TRIAL?**

---

Session Objectives:

- Discuss how variability in treatment delivery and adherence can affect results, including
  - Potential influence of variation in standard treatment practice, and
  - Considerations for balancing participant autonomy and safety.
- Suggest key factors that could help determine the base comparison and level of control suited to a particular trial.
- Discuss how a decision aid laying out key questions and considerations might help inform current and future studies.

3:10 p.m. **Introduction: A proposed framework for a decision aid**

JENNIFER GRAFF, *Session Moderator*

Vice President of Comparative Effectiveness Research  
National Pharmaceutical Council

3:20 p.m. **Looking back: How might a decision aid inform a real-world example?**

LARRY ALPHS

Deputy Chief Medical Officer  
Newron Pharmaceuticals

3:40 p.m. **Looking forward: How decision aid might apply to future studies?**

***Panel discussion and audience Q&A***

JUDITH CARRITHERS

Director of Regulatory Services

Advarra

W. BENJAMIN NOWELL

Director, Patient-Centered Research

Global Healthy Living Foundation

Co-PI: ArthritisPower Patient Powered Research Network

PETER STEIN

Deputy Director, Office of New Drugs

Center for Drug Evaluation and Research

U.S. Food and Drug Administration

4:50 p.m. **Day 1 wrap up and concluding thoughts/discussion with audience**

5:00 p.m. **ADJOURN WORKSHOP DAY 1**

---

## DAY 2: July 18, 2018

7:30 a.m. Breakfast Available Outside the Lecture Room

8:00 a.m. **Welcome**

MARK MCCLELLAN, *Workshop Series Co-Chair*  
Director  
Duke-Margolis Center for Health Policy

GREGORY SIMON, *Workshop Series Co-Chair*  
Investigator  
Kaiser Permanente Washington Health Research Institute

---

### SESSION V HOW CAN BIAS IN OBSERVATIONAL COMPARISONS BE ASSESSED AND MINIMIZED?

---

#### Session Objectives:

- Discuss methods to assess presence of and optimally reduce bias from unmeasured confounding.
- Suggest key considerations for assessing—and communicating—uncertainty in observational studies.
- Discuss how a decision aid laying out key questions and considerations might help inform current and future studies.

8:10 a.m. **Introduction: A proposed framework for a decision aid**

DAVID MARTIN  
Associate Director for Real-World Evidence Analytics  
U.S Food and Drug Administration

8:20 a.m. **Looking back: How might a decision aid inform a real-world example?**

HECTOR IZURIETA  
Epidemiologist, Office of Biostatistics and Epidemiology  
Center for Biologics Evaluation and Research  
U.S. Food and Drug Administration

8:40 a.m. **Looking forward: How decision aid might apply to future studies?**

***Panel discussion and audience Q&A***

GREGORY DANIEL, *Session Moderator*  
Deputy Director  
Duke-Margolis Center for Health Policy

JESSICA FRANKLIN  
Assistant Professor of Medicine  
Harvard Medical School

NICOLE GORMLEY  
Team Lead, Division of Hematologic Products  
Office of Hematology and Oncology Products  
Center for Drug Evaluation and Research  
U.S. Food and Drug Administration

JAVIER JIMENEZ  
Vice President and Global Head for Real-World Evidence and Clinical Outcomes  
Sanofi

HENG LI  
Mathematical Statistician  
Center for Devices and Radiological Health  
U.S. Food and Drug Administration

MARK VAN DER LAAN  
Professor, Biostatistics and Statistics  
University of California Berkeley

10:00 a.m.     **BREAK**

---

## **SESSION VI    FDA PANEL**

---

### Session Objectives:

- Hear updates and perspective of current thinking about real-world evidence in Europe.
- Discuss challenges, opportunities, and remaining gaps for moving forward with real-world evidence application.

10:15 a.m.     **A European perspective**  
ALASDAIR BRECKENRIDGE, *Session Moderator*  
Emeritus Professor of Clinical Pharmacology  
University of Liverpool

10:30 a.m.     **Reflections from FDA**  
STEVEN ANDERSON  
Director, Office of Biostat. and Epidem., Center for Biologics Evaluation and Research  
U.S. Food and Drug Administration

JACQUELINE CORRIGAN-CURAY  
Director, Office of Medical Policy, Center for Drug Evaluation and Research  
U.S. Food and Drug Administration

JEFF SHUREN  
Director, Center for Devices and Radiological Health  
U.S. Food and Drug Administration

11:15 a.m. **Panel discussion with audience**

11:50 p.m. **Synthesis of workshop discussions**

MARK MCCLELLAN , *Workshop Series Co-Chair*

Director

Duke-Margolis Center for Health Policy

GREGORY SIMON, *Workshop Series Co-Chair*

Investigator

Kaiser Permanente Washington Health Research Institute

12:00 p.m. **ADJOURN WORKSHOP DAY 2**



## **Examining the Impact of Real-World Evidence on Medical Product Development: A Workshop Series**

### **Workshop 3: Application**

#### **PLANNING COMMITTEE BIOGRAPHIES**

##### **CO-CHAIRS:**

**MARK MCCLELLAN, M.D., Ph.D.**, is the Robert J. Margolis Professor of Business, Medicine, and Policy, and Director of the Duke-Margolis Center for Health Policy at Duke University with offices at Duke and in Washington DC. Dr. McClellan is a doctor and an economist, and his work has addressed a wide range of strategies and policy reforms to improve health care, including payment reforms to promote better outcomes and lower costs, methods for development and use of real-world evidence, and approaches for more effective drug and device innovation. Dr. McClellan is a former administrator of the Centers for Medicare & Medicaid Services (CMS) and former commissioner of the U.S. Food and Drug Administration (FDA), where he developed and implemented major reforms in health policy. He was also a Senior Fellow at the Brookings Institution and a professor of economics and medicine at Stanford University.

**GREGORY SIMON, M.D., M.P.H.**, is an investigator at Kaiser Permanente Washington Health Research Institute and a psychiatrist in Kaiser Permanente's Behavioral Health Service. He is also a Research Professor in the Department of Psychiatry and Behavioral Sciences at the University of Washington and co-chair of the national scientific advisory board of the Depression and Bipolar Support Alliance. Dr. Simon completed residency training in internal medicine at the University of Washington, residency training in psychiatry at the Massachusetts General Hospital, and fellowship training in the Robert Wood Johnson Clinical Scholars program at the University of Washington. Dr. Simon's research focuses on improving access to and quality of mental health care, especially for mood disorders. Specific areas of research include improving adherence to medication, increasing the availability of effective psychotherapy, evaluating peer support by and for people with mood disorders, identifying and reducing risk of suicidal behavior, cost-effectiveness of treatment, and comorbidity of mood disorders with chronic medical conditions. Dr. Simon currently leads the Mental Health Research Network, a National Institute of Mental Health-funded cooperative agreement supporting population-based mental health research across 13 large health systems.

##### **PLANNING COMMITTEE:**

**JEFF ALLEN, Ph.D.**, serves as the President and CEO of Friends of Cancer Research (Friends). During the past 20 years, Friends has been instrumental in the creation and implementation of policies ensuring patients receive the best treatments in the fastest and safest way possible. As a thought leader on many issues related to the Food and Drug Administration, regulatory strategy and healthcare policy, he is regularly published in prestigious medical journals and policy publications, and has contributed his expertise to the legislative process on multiple occasions. Recent Friends initiatives include the establishment of the new Breakthrough Therapies designation and the development of the Lung Cancer Master Protocol, a unique partnership that will accelerate and optimize clinical trial conduct for new drugs. Dr. Allen received his Ph.D. in cell and molecular biology from Georgetown University, and holds a Bachelors of Science in Biology from Bowling Green State University.

**ANDREW BINDMAN, M.D.**, is Professor of Medicine, Health Policy, Epidemiology and Biostatistics, at the University of California San Francisco (UCSF). He is also a senior advisor in the Office of the Assistant Secretary for Planning and Evaluation within the US Department of Health and Human Services where he works on issues related to the health care workforce, graduate medical education, and Medicaid; from May

2016 until January 2017, Dr. Bindman served as the Director of the Agency for Healthcare Research and Quality. He has practiced, taught and performed health services research at UCSF's affiliated San Francisco General Hospital for over 25 years. Dr. Bindman has published extensively on evaluations of Medicaid health policies with a focus on access to care and health outcomes. During 2009-2010, he served as a Robert Wood Johnson Health Policy Fellow on the staff of the Energy and Commerce Committee within the US House of Representatives where he was intimately involved in the drafting of legislative language for the federal health reform law, the Patient Protection and Affordable Care Act (ACA).

**ADAM HAIM, Ph.D.**, is the Chief of the Treatment and Preventive Intervention Research Branch within the Division of Services and Intervention Research at the National Institute of Mental Health (NIMH). Dr. Haim manages a broad portfolio of research focused on evaluating the efficacy and effectiveness of pharmacologic, psychosocial and combination interventions on mental and behavior disorders. He is also a thought leader in the development, evaluation and implementation of technology enhanced mental health interventions. Dr. Haim is a licensed clinical psychologist and earned his doctoral degree in clinical psychology from State University of New York at Albany and completed his research fellowship at the NIMH Intramural Program in the Division of Clinical Neuroendocrinology.

**MICHAEL HORBERG, M.D., M.A.S, FACP**, is Executive Director Research and Community Benefit of Mid-Atlantic Permanente Medical Group (MAPMG) and the director of the Mid-Atlantic Permanente Research Institute (MAPRI). He is also director of HIV/AIDS for Kaiser Permanente. Dr. Horberg has been appointed to serve on the Presidential Advisory Council on HIV/AIDS (PACHA), and co-chairs the Access to Care and Improved Outcomes Committee of PACHA. Dr. Horberg is a Fellow of the American College of Physicians, and he presently serves as Vice-Chair of the Board of Directors of the HIV Medicine Association of the Infectious Disease Society of America. He has co-chaired the NCQA/AMA/HRSA/IDSA sponsored Expert Panel on HIV-related provider performance measures. He is Assistant Clinical Professor at Stanford University Medical School. Dr. Horberg is past-president of the national Gay and Lesbian Medical Association. His HIV research interests are health service outcomes for HIV-infected patients (including HIV quality measures and care improvement, and determinants of optimized multidisciplinary care for maximized HIV outcomes), medication adherence issues in these patients, and epidemiology of the disease. He graduated from Boston University's College of Liberal Arts and School of Medicine (with honors of Summa cum Laude and Phi Beta Kappa) and completed his internal medicine residency at Michael Reese Hospital in Chicago (University of Chicago affiliate). He received his Master of Advanced Studies (Clinical Research) from University of California San Francisco.

**JOHN GRAHAM, PHARM.D.**, is a Senior Vice President at GlaxoSmithKline and leads the Medical Engagement and Value Evidence and Outcomes (VEO) organization within GSK's R&D division. The VEO organization is accountable for providing strategic input into the progression of assets as well as ensuring the value demonstration through evidence generation of the assets that do progress. This evidence includes accountability for real world evidence across GSK products and across all regions globally. In addition, his organization is accountable for GSK's Patient Centered Outcomes and alignment with Patients in Partnership. His interests include real world evidence (RWE), particularly the integration of traditional clinical evidence with non-traditional RWE, to answer questions from patients, providers, and payers. Dr. Graham sits on various Advisory Boards relating to Real World Evidence including the National Academies of Science working group on RWE, OPERAND, as well as the RWE Forum. He has published scientific communications across cardiovascular and metabolic diseases and across both clinical as well as RWE areas. Dr. Graham joined GSK in 2014 as the Vice President, CV/Met, NS, Rare Disease VEO. He is located in the GSK R&D HUB in Collegeville, PA. Prior to GSK, Dr. Graham worked for over 16 years at Bristol-Myers Squibb where he was most recently the Head of the U.S. HEOR group and previously had roles across R&D and commercial with both Global and local accountabilities. Prior to industry, Dr. Graham worked in academia most recently as Assistant Professor of Clinical Pharmacy at St. Louis College of Pharmacy. Dr. Graham holds a Doctorate of Pharmacy degree from Idaho State University and completed a Primary Care Residency at the University of Nebraska Medical Center.

**RICHARD KUNTZ, M.D.**, is Senior Vice President and Chief Scientific, Clinical and Regulatory Officer of Medtronic and serves as a member of the company's Executive Committee. In this role, which he assumed in

August 2009, Dr. Kuntz oversees the company's global regulatory affairs, health policy and reimbursement, clinical research activities, and corporate technology. Dr. Kuntz joined Medtronic in October 2005, as Senior Vice President and President of Medtronic Neuromodulation, which encompasses the company's products and therapies used in the treatment of chronic pain, movement disorders, spasticity, overactive bladder and urinary retention, benign prostatic hyperplasia, and gastroparesis. In this role he was responsible for the research, development, operations and product sales and marketing for each of these therapeutic areas worldwide. Dr. Kuntz brings to Medtronic a broad background and expertise in many different areas of healthcare. Prior to Medtronic he was the Founder and Chief Scientific Officer of the Harvard Clinical Research Institute (HCRI), a university-based contract research organization which coordinates National Institutes of Health (NIH) and industry clinical trials with the Food and Drug Administration (FDA). Dr. Kuntz has directed over 100 multicenter clinical trials and has authored more than 250 original publications. His major interests are traditional and alternative clinical trial design and biostatistics. Dr. Kuntz also served as Associate Professor of Medicine at Harvard Medical School, Chief of the Division of Clinical Biometrics, and an interventional cardiologist in the division of cardiovascular diseases at the Brigham and Women's Hospital in Boston, MA. Dr. Kuntz has served as a member of the Board of Governors of PCORI (Patient Centered Outcomes Research Institute) since it was established in 2010 as part of the Affordable Care Act. Dr. Kuntz graduated from Miami University, and received his medical degree from Case Western Reserve University School of Medicine. He completed his residency and chief residency in internal medicine at the University of Texas Southwestern Medical School, and then completed fellowships in cardiovascular diseases and interventional cardiology at the Beth Israel Hospital and Harvard Medical School, Boston. Dr. Kuntz received his master's of science in biostatistics from the Harvard School of Public Health.

**ELLIOTT LEVY, M.D.**, is senior vice president, Global Development, at Amgen. He is responsible for the clinical development of Amgen's investigative and marketed products. Before joining Amgen, Dr. Levy spent 17 years at Bristol-Myers Squibb (BMS) in clinical development and pharmacovigilance. He has contributed to the development and approval of numerous new therapies for cardiovascular, metabolic, inflammatory, and malignant diseases, and led large organizations through periods of transformative change. Dr. Levy is a graduate of the Yale School of Medicine, where he also trained in internal medicine and nephrology. Dr. Levy was also a member of the Renal Division at Brigham and Women's Hospital in Boston, Massachusetts, where he was an investigator in federally sponsored outcomes research as well as industry-sponsored clinical trials.

**DAVID MADIGAN, Ph.D.**, is Professor of Statistics and Executive Vice President and Dean of the Faculty of Arts & Sciences at Columbia University in New York City. He received a bachelor's degree in Mathematical Sciences and a Ph.D. in Statistics, both from Trinity College Dublin. He has previously worked for AT&T Inc., Soliloquy Inc., the University of Washington, Rutgers University, and SkillSoft, Inc. He has over 180 publications in such areas as Bayesian statistics, text mining, Monte Carlo methods, pharmacovigilance and probabilistic graphical models. He is an elected Fellow of the American Statistical Association, the Institute of Mathematical Statistics, and the American Association for the Advancement of Science. He has served terms as Editor-in-Chief of Statistical Science and of Statistical Analysis and Data Mining – the ASA Data Science Journal.

**DEVEN MCGRAW, J.D., M.P.H.**, is the General Counsel and Chief Regulatory Officer for Ciitizen, a consumer health technology start-up. Prior to joining Ciitizen, she directed U.S. health privacy and security through her roles as Deputy Director, Health Information Privacy at the HHS Office for Civil Rights (the office that oversees HIPAA policy development and enforcement) and Chief Privacy Officer (Acting) of the Office of the National Coordinator for Health IT. Widely recognized for her expertise in health privacy and security, she directed the Health Privacy Project at the Center for Democracy & Technology (a nonprofit civil liberties organization) for six years and led the privacy and security policy work for the HITECH Health IT Policy Committee. She also served as the Chief Operating Officer of the National Partnership for Women and Families. She has also advised health industry clients on HIPAA compliance and data governance while a partner at Manatt, Phelps & Phillips, LLP. Deven graduated magna cum laude from Georgetown University Law Center and has a Masters of Public Health from Johns Hopkins University.

**RICHARD PLATT, M.D., M.Sc.**, is Professor and Chair of the Harvard Medical School Department of Population Medicine and Executive Director of the Harvard Pilgrim Health Care Institute. He is Principal Investigator of

the FDA Sentinel System. He led the development, with the Massachusetts Department of Public Health, of ESPnet, a system for doing real time EHR-based surveillance for both syndromes of interest and individually notifiable conditions. He is also co-Principal Investigator of the National Patient Centered Clinical Research Network (PCORnet) Coordinating Center, which is developing standard methods for extracting and using EHR data for multiple uses. Dr. Platt also co-leads the coordinating center of the NIH Health Care System Research Collaboratory and leads a CDC Prevention Epicenter. He co-chairs the CER Innovation Collaborative of the National Academy of Medicine's Leadership Consortium for a Value & Science-Driven Health System, and is a member of the American Medical Colleges Advisory Panel on Research.

**JOANNE WALDSTREICHER, M.D.**, is Chief Medical Officer, Johnson & Johnson. In this role, she has oversight across pharmaceuticals, devices and consumer products for safety, epidemiology, clinical and regulatory operations transformation, internal and external partnerships and collaborations supporting development of the ethical science, technology and R&D policies, including those related to clinical trial transparency and compassionate access. Joanne also chairs the Pharmaceuticals (Janssen) R&D Development Committee and supports the Device and Consumer Development Committees, which review late stage development programs in the pipeline. She also holds an appointment as a Faculty Affiliate of the Division of Medical Ethics, Department of Population Health, New York University School of Medicine. Among her prior roles in Janssen, the pharmaceutical sector of Johnson & Johnson, Dr. Waldstreicher was responsible for late-stage development spanning the areas of neuroscience, cardiovascular and metabolism including INVOKANA®, XARELTO®, INVEGA SUSTENNA®, and INVEGA TRINZA®. Before joining Johnson & Johnson in 2002, Dr. Waldstreicher was head of the Endocrinology and Metabolism clinical research group at Merck Research Laboratories, and responsible for overseeing clinical development of MEVACOR®, ZOCOR®, PROSCAR® and PROPECIA®, and for clinical development programs in atherosclerosis, obesity, diabetes, urology and dermatology. During that time, she received numerous awards and distinctions, including the Merck Research Laboratory Key Innovator Award. Dr. Waldstreicher received both the Jonas Salk and Belle Zeller scholarships from the City University of New York and graduated Summa Cum Laude from Brooklyn College. Joanne graduated Cum Laude from Harvard Medical School in 1987, and completed her internship and residency at Beth Israel Hospital, and her endocrinology fellowship at MGH. She has won numerous awards and scholarships, and has authored numerous papers and abstracts. In December, 2016, Dr. Waldstreicher was named Healthcare Champion of the Year for Women by the National Association of Female Executives.

**MARCUS D. WILSON, Pharm.D.**, is President of HealthCore, Anthem's wholly-owned outcomes research subsidiary. He has been extensively involved in efforts to utilize real-world data environments to accelerate healthcare evidence development and to facilitate clinical decision support for more than 20 years. Prior to co-founding HealthCore in 1996, Dr. Wilson spent seven years within an integrated delivery system owned by BCBS of Delaware where he oversaw the physician and patient clinical decision support, pharmacy policy and clinical trials programs. Dr. Wilson is active on a number of boards and national committees including serving as chair of the Joint Research Committee for the Academy of Managed Care Pharmacy (AMCP) & the AMCP Foundation; member of the FDA Sentinel Initiative Planning Board; Board of Directors for the Center for Medical Technology Policy (CMTP); a member of the Dean's Roundtable, College of Science, Virginia Tech; and member of the Planning Committee for the National Academy of Sciences, Engineering, and Medicine's Real-World Evidence Workshop Series that is being conducted for FDA in response to the 21st Century Cures legislation. Dr. Wilson is a past member of numerous boards and committees including the Board of Directors for the International Society of Pharmacoeconomics and Outcomes Research (ISPOR), the eHealth Initiative and is former chair of the Innovations in Medical Evidence Development (IMEDS) Steering Committee, Reagan-Udall Foundation for the FDA. Dr. Wilson received his Bachelor of Science in Biochemistry from Virginia Tech and his Doctor of Pharmacy degree from the Medical College of Virginia. He completed a residency in Family Medicine at the Medical University of South Carolina prior to joining the faculty at the Philadelphia College of Pharmacy (now the University of Sciences in Philadelphia).



## **Examining the Impact of Real-World Evidence on Medical Product Development: A Workshop Series**

### **Workshop 3: Application**

#### **SPEAKER AND DISCUSSANT BIOGRAPHIES**

**LARRY ALPHS, M.D., Ph.D.**, is Deputy Chief Medical Officer at Newron Pharmaceuticals US, LLC. He obtained his M.D. and Ph.D. from the University of Chicago Pritzker School of Medicine. Dr. Alphas worked for 10 years as a researcher/clinician specializing in clinical research in persons with serious psychiatric disorders. For most of his career Dr. Alphas has worked in various aspects of clinical development, with positions at Novartis, Knoll Pharmaceuticals, Pfizer, Janssen and Newron. He has done Phase I-IV work in a variety of CNS disorders, including schizophrenia, bipolar disorder, suicidality, anxiety, depression, epilepsy, neuropathic pain, and traumatic brain injury. His interests have focused on the nosology and treatment of and suicidal ideation and behavior, especially that observed in schizophrenia, and depression. He is a founder of the International Society for CNS Clinical Trials and Methodology, and now serves on its board. Dr. Alphas has authored numerous peer-reviewed journal articles and book chapters, and has developed proprietary scales for use in psychiatric evaluations. Throughout his career Dr. Alphas has been interested in real world clinical trials. He has designed and led two major real world clinical trials: the InterSePT Study and the PRIDE study. The InterSePT study was a randomized, open-label, suicide monitoring board-blinded study comparing clozapine versus olanzapine for the treatment of suicide behavior in schizophrenic/schizoaffective patients at high risk for suicide. The PRIDE study was a randomized, open-label, relapse monitoring board-blinded study comparing once-monthly injectable paliperidone palmitate with oral antipsychotics for relapse prevention in schizophrenic patients who had recently been incarcerated. Both studies showed the treatment of interest to be superior to comparators. Dr. Alphas has also developed numerous scale for CNS clinical trials. The ASPECT-R has been developed to characterize the relative "pragmaticness" of clinical trials.

**AYLIN ALTAN, Ph.D.**, is Senior Vice President of Research at OptumLabs. Dr. Altan specializes in the use of real world data and real world evidence generation. Her role within OptumLabs includes working with partners as a consultant and co-investigator, overseeing the OptumLabs team of scientists and research analysts, and leading workshops, issues panels, and training sessions both for OptumLabs Partners and at clinical, policy, and industry conferences and meetings. Dr. Altan's research areas of focus include opioid use & pain management, connected care in diabetes management, and patients with complex chronic illness. Prior to joining Optum Labs in 2015, Dr. Altan spent 15 years in Optum's Health Economics and Outcomes Research practice. As head of the North American HEOR organization, Dr. Altan was responsible for working with Life Sciences companies to develop evidence generation strategies in support of their key assets, with the end goal of publication in the peer-reviewed, clinical and policy literature. Dr. Altan holds degrees from the University of Minnesota and Stanford University. She was a contributing author to *Managed Care and the Treatment of Chronic Illness*, published by Sage Publications, and has more than 50 publications in peer-reviewed journals and more than 90 posters, workshops, and podium presentations.

**STEVEN ANDERSON, Ph.D., M.P.P.**, is currently the Director of the Office of Biostatistics and Epidemiology (OBE) at the FDA Center for Biologics Evaluation and Research (CBER). He provides leadership for all CBER statistical, epidemiological and benefit-risk assessment programs. Previously, Dr. Anderson had been the Deputy Director for OBE since 2005. In 2001, he was hired by CBER as the Associate Director for Risk Assessment to establish a program in quantitative risk assessment for biologic products including vaccines, blood products and others. Since his arrival at FDA he has led numerous important risk assessment projects and epidemiological studies. He led the first studies at FDA using Centers for Medicare & Medicaid Services (CMS) data to estimate blood

utilization and address important blood product safety questions of regulatory concern. He has conducted collaborative studies using the FDA Sentinel system and CMS data to evaluate the safety of blood products and vaccines and worked to integrate use of these data systems into CBER's regulatory processes to improve biologic product safety evaluations and surveillance. Dr. Anderson earned a Master's Degree in Public Policy (MPP) at Georgetown University and while there developed the first quantitative risk assessment for antimicrobial resistant pathogens. Dr. Anderson received his PhD in Biology from the University of Cincinnati where he worked on biochemistry, drug resistance and ion pumps, pathogenicity and genomics of unique tropical disease pathogens. He has published a number of articles in biologic product safety, risk assessment, epidemiology, pharmacoepidemiology, infectious diseases, biologics safety, and genomics and protein structure/targeting.

**ROBERT BALL, M.D., M.P.H., Sc.M.**, is Deputy Director, Office of Surveillance and Epidemiology (OSE), Center for Drug Evaluation and Research (CDER), FDA. Dr. Ball shares in the responsibilities for leading OSE staff evaluating drug risks and promoting the safe use of drugs by the American people, including managing FDA's Sentinel System. From 2008 to 2013, Dr. Ball served as the Director, Office of Biostatistics and Epidemiology (OBE), Center for Biologics Evaluation and Research (CBER), FDA, where he led statistical and epidemiological evaluation of biologic products, including post-marketing safety programs for vaccines and blood.

**SIR ALASDAIR BRECKENRIDGE, CBE, FRCP, FRCPE, FRSE, FMedSci**, is the Emeritus Professor of Clinical Pharmacology at the University of Liverpool. He is a clinical pharmacologist and medicines regulator who has worked with the UK Academy of Medical Sciences and the European Medicines Agency. He is former Chair of the Medicines and Healthcare Products Regulatory Agency (MHRA) in the UK, the Department of Health agency responsible for regulating all medicines and medical devices. During his tenure at MHRA, Sir Dr. Breckenridge served as the first Chair of the Department of Health's Emerging Science and Bioethics Advisory Committee (ESBAC), an organization responsible for providing advice to UK health departments on emerging healthcare scientific developments and their ethical, legal, social and economic implications. He was a member of the Medical Research Council, and he worked closely on several programs of the European Union, the World Health Organization and the International Union of Pharmacology, and as Chair of the Committee on Safety of Medicines. Prior to MHRA, he was Professor of Clinical Pharmacology at the University of Liverpool and a member and then Chairman of local and regional health authorities in the north-west of England, including Chairman of the North West Regional Office. Sir Dr. Breckenridge is a clinical pharmacologist and after training at the Royal Postgraduate Medical School, he was appointed to the Chair of Clinical Pharmacology at the University of Liverpool. Sir Dr. Breckenridge was knighted in 2004.

**JUDITH CARRITHERS, J.D.**, serves as the Director of Oncology Services in the Central Oncology Review Division at Advarra (formerly Schulman IRB). Prior to her current position, she worked in the Johns Hopkins Office of Human Subjects Research and served as an Assistant dean. At Hopkins, she was responsible for human subjects research oversight and the Institutional Review Board system at the Johns Hopkins University School of Medicine (JHUSOM IRB). The JHUSOM IRB serves six hospitals and several schools within Hopkins, and a clinical research network comprised of five regional hospitals. The JHUSOM IRB has seven IRB committees and approximately 5500 active studies. In addition, the JHUSOM IRB coordinates ancillary review of research protocols (for example, conflict of interest committee review and clinical radiation review). Ms. Carrithers received her J.D. degree from Stanford University Law School (California) in 1987 and her Master's degree in public administration from Seattle University (Washington) in 1987.

**JACQUELINE CORRIGAN-CURAY, J.D., M.D.**, serves as Director of the Office of Medical Policy (OMP) in the Center for Drug Evaluation and Research (CDER) at FDA. She leads the development, coordination, and implementation of medical policy programs and strategic initiatives. She works collaboratively with other CDER program areas, FDA centers, and stakeholders on enhancing policies to improve drug development and regulatory review processes. OMP is comprised of the Office of Prescription Drug Promotion (OPDP) and the Office of Medical Policy Initiatives (OMPI). OPDP oversees the regulation of prescription drug promotion and advertising. OMPI provides oversight and direction for new and ongoing policy initiatives in broad-based medical and clinical policy areas. Dr. Corrigan-Curay brings to the position a unique legal, scientific policy, and clinical background with expertise in risk and scientific assessment, and clinical trial design and oversight. Before joining FDA, she served as supervisory medical officer with the Immediate Office of the Director,

National Heart, Lung and Blood Institute (NHLBI), at the National Institute of Health's (NIH) and served in director and acting director roles with the Office of Biotechnology Activities (OBA), Office of Science Policy at NIH, where she was executive secretary of the NIH Recombinant DNA Advisory Committee. She has held positions as an attending physician with the VA Medical Center, a policy analyst with the Congressional Office of Technology Assessment, and a practicing attorney in Washington, D.C. Dr. Corrigan-Curay earned her law degree from Harvard Law School, her medical degree from University of Maryland School of Medicine, and a bachelor's degree in history of science from Harvard/Radcliffe College in Cambridge, MA. She completed her training in internal medicine at Georgetown University Medical Center, where she also served as a clinical assistant professor of medicine. She continues to practice internal medicine part-time at the Veterans Affairs Medical Center in Washington, D.C.

**CATHY CRITCHLOW, Ph.D., M.S.**, serves as the director of the of the Center for Observational Research (CfOR) at Amgen. In this position, Dr. Critchlow provides operational and strategic leadership for the design and conduct of observational research within Amgen. The CfOR Real World Data (RWD) Platform provides widespread access to patient health data and visualization and analytic tools based on innovative technologies to aid teams in the generation of real world evidence in support of drug development and commercialization of Amgen products. Dr. Critchlow joined Amgen in 2004 where she led a number of Therapeutic Areas within Global Epidemiology prior to her being named Head of CfOR in 2012. Prior to joining Amgen, Dr. Critchlow was a faculty member in Epidemiology at the University of Washington. Dr. Critchlow was a member of the Endocrinologic and Metabolic Advisory Committee of the Food and Drug Administration and has served on a number of research review committees for the National Institutes of Health. Dr. Critchlow earned her bachelor's degree from Stanford University, and both her master's degree in biomathematics and her doctorate degree in epidemiology from the University of Washington. Dr. Critchlow is an Affiliate Professor of Epidemiology at the University of Washington and a Fellow of the American College of Epidemiology.

**GREGORY DANIEL, Ph.D., M.P.H., M.S.**, is the Deputy Director of the Duke-Robert J. Margolis, MD Center for Health Policy and a Clinical Professor in Duke's Fuqua School of Business. Dr. Daniel directs the DC-based office of the Center and leads the Center's pharmaceutical and medical device policy portfolio, which includes developing policy and data strategies for improving development and access to innovative pharmaceutical and medical device technologies. This includes post-market evidence development to support increased value, improving regulatory science and drug development tools, optimizing biomedical innovation, and supporting drug and device value-based payment reform. Dr. Daniel is also Adjunct Associate Professor in the Division of Pharmaceutical Outcomes and Policy at the UNC Eshelman School of Pharmacy. Previously, he was Managing Director for Evidence Development & Biomedical Innovation in the Center for Health Policy and Fellow in Economic Studies at the Brookings Institution and Vice President, Government and Academic Research at HealthCore (an Anthem, Inc. company). In addition to health and pharmaceutical policy, Dr. Daniel's research expertise includes real world evidence (RWE) development utilizing electronic health data in the areas of health outcomes and pharmacoepidemiology, comparative effectiveness, and drug safety and pharmacoepidemiology. Dr. Daniel received a Ph.D. in pharmaceutical economics, policy and outcomes from the University of Arizona, as well as an M.P.H., M.S., and BS in Pharmacy all from The Ohio State University.

**NANCY DREYER, Ph.D., M.P.H.**, is the Chief Scientific Officer and Global Chief of Scientific Affairs for IQVIA Real-World & Analytic Solutions. She leads the Center for Advanced Evidence Generation, focusing on the use of real-world evidence for regulators, payers, clinicians, and patients using minimally interventional and non-interventional study design that rely on primary and/or secondary data collection. She has worked with the FDA, most recently helping to plan a medical device evaluation network, and also has worked with the European Medicines Agency testing new methods for pharmacovigilance. She is a Fellow of both the International Society of Pharmacoepidemiology and the Drug Information Association, and is Adjunct Professor of Epidemiology at the UNC Gillings School of Global Public Health in North Carolina.

**LUCA FOSCHINI, Ph.D., M.S., M.E.**, is the Co-founder and Chief Data Scientist at Evidation Health, responsible for data analytics and research and development. At Evidation he has driven research collaborations resulting in numerous publications in the fields of machine learning, behavioral economics, and medical informatics. Previously, Dr. Foschini held research positions in industry and academic institutions, including Ask.com, Google, ETH Zurich, and UC Santa Barbara. He has co-authored several papers and patents on efficient



algorithms for partitioning and detecting anomalies in massive networks. Dr. Foschini holds M.S. and Ph.D. degrees in Computer Science from UC Santa Barbara, and M.E. and B.E. degrees from the Sant'Anna School of Pisa, Italy.

**JESSICA FRANKLIN, Ph.D.**, is an Assistant Professor of Medicine at Harvard Medical School and biostatistician in the Division of Pharmacoepidemiology and Pharmacoeconomics at Brigham and Women's Hospital. Her research focuses on developing and applying statistical methods for the study of the comparative effectiveness, safety, and utilization of medicines based on large electronic healthcare databases, including insurance claims and electronic health records. Dr. Franklin has developed several novel approaches to evaluating the performance of causal inference methods in these data, including plasmode simulation. Currently, she is leading the RCT DUPLICATE project, which aims to build an empirical basis for causal inference methods applied to real world data analyses of medications through the large-scale replication of randomized trials in real world data. She received her Bachelor's degree in mathematics at the University of Georgia and her doctorate in biostatistics at the Johns Hopkins Bloomberg School of Public Health.

**NICOLE GORMLEY, M.D.**, completed fellowship training in both hematology and critical care at the National Institutes of Health. While at NIH, her research involved investigation of the non-infectious pulmonary complications that occur after hematopoietic stem cell transplant and served as the principal investigator of clinical trials of a novel treatment for bronchiolitis obliterans in the post-transplant setting. After her training, Dr. Gormley served as the Deputy Clinical Director at the National Heart, Lung and Blood Institute before joining the Food and Drug Administration. Dr. Gormley currently works in the Division of Hematology Products at the FDA as the multiple myeloma Clinical Team Leader and serves as a scientific liaison in the Office of Hematology and Oncology Products. In these roles, Dr. Gormley has actively engaged the multiple myeloma community on the development of novel endpoints, including minimal residual disease.

**JENNIFER GRAFF, Pharm.D.**, is the National Pharmaceutical Council's (NPC) vice president of comparative effectiveness research. In this role, she leads research and policy initiatives to advance the use of evidence to inform health care decision-making. Her areas of focus include research and education to support increased access to and use of high-quality data, development and adoption of good research methods, and policies to enable the exchange of truthful and non-misleading information to support stakeholder decision-making. Prior to joining NPC in 2009, Dr. Graff led strategic health economic and outcomes research activities at MedImmune and Pfizer Pharmaceuticals. She has authored over 20 peer-reviewed articles and presents frequently on policy issues affecting the biopharmaceutical industry. She currently serves as an associate editor of the AcademyHealth journal eGEMS and as a member of the Academy of Managed Care Pharmacy Format Executive Committee. Dr. Graff holds a Doctorate of Pharmacy from the University of Nebraska Medical Center, and completed a Health Outcomes and Pharmacoeconomics fellowship at the University of Michigan.

**HECTOR IZURIETA, M.D., M.P.H.**, serves as a senior epidemiologist in the Office of Biostatistics and Epidemiology, Center for Biologics Evaluation and Research, FDA. Prior to FDA, Dr. Izurieta worked in Africa for five years with the French Doctors without Borders and with the Swiss cooperation. He subsequently joined CDC, working on vaccine preventable diseases for 11 years, including 4 years seconded at the Pan American Health Organization (PAHO/WHO), where he played a key role in the interruption of measles circulation in the Americas. He joined CBER/FDA in 2003. He has over 80 publications in international scientific journals, mainly on the use of real world data for the investigation of vaccine and other biologics effectiveness and safety, including leading papers in all major medical journals. His research has been used by CDC, WHO, FDA, and other organizations for policy and public health decision-making. He holds a foreign medical degree, a Master's in Public Health degree from Harvard University, and was trained in epidemiology and preventive medicine at CDC's Epidemic Intelligence Service (EIS) and the Preventive Medicine programs in Atlanta, GA, respectively.

**JAVIER JIMENEZ, M.D., M.P.H.**, is a specialist in Preventive Medicine with more than 20 years of experience in epidemiology and 15 years in the pharmaceutical industry at the country, regional and global level. Throughout his career, Dr. Jimenez has applied his In-depth knowledge of epidemiology and Real World Evidence to generate the required evidence to demonstrate the value of medicines for patients, health care providers and society. Through staffing, training, and the development of high performing teams, collaboration with internal functions, and partnering with scientific societies and providers, Dr. Jimenez led the development of a world

class RWE function responsible for design and implementation of high quality innovative RWE programs. Dr. Jiminez held the role of VP, Medical Evidence and Observational Research at AstraZeneca, where he was responsible for worldwide RWE strategy, Epidemiology support and delivery of Medical Evidence (RWE, Phase III/IV and External Sponsored Research) for the AstraZeneca portfolio. He was also a key representative on Global Medical Affairs Leadership team and Medical Affairs studies Governance committees. Most recently, Dr. Jiminez joined Sanofi in January 2017 to lead the RWE and Clinical Outcomes team reporting to Bernard Hamelin in the Evidence Generation team (CMO). The mission of the RWE and CO team is to: 1) Build and maintain an integrated RWE platform, source key data sets and use of analytical capabilities and partners required to generate high quality medical evidence; 2) Champion the use of innovative medical evidence generation capabilities by GBUs, function and countries by educating them on the use of new capabilities, enhancement and collaboration to ensure the integration of RWE strategy in all projects; 3) Generate meaningful insights to support internal and external decision making required to ensure the best outcome for patients by appropriate use of our products in clinical practice; 4) Deliver innovative and impactful medical evidence required by healthcare decision makers to understand medical value of Sanofi therapies by knowing our products and data best, identify gaps and potential challenges early on; and 5) Enhance external collaboration with healthcare decision makers, patients, and academic institutions to shape medical evidence generation methodology and policies.

**PÁLL JÓNSSON, Ph.D.**, is Associate Director for Research and Development at UK's National Institute for Health and Care Excellence (NICE). He leads NICE's contribution to a portfolio of international research projects which support NICE in adapting to policy developments in health and social care delivery. Dr. Jónsson has a doctorate degree in biochemistry and bioinformatics from University College London and has expertise in health technology assessments of drugs and diagnostics at NICE. Before joining NICE he gained research experience in academia, the not-for-profit sector and the pharmaceutical industry. Dr. Jónsson sits on advisory boards of several international projects in the area of health technology assessment and evidence generation.

**HENG LI, Ph.D.**, is mathematical statistician team leader at the Center for Devices and Radiological Health, Food and Drug Administration. He received a PhD in statistics from Harvard University in 1996. He is interested in both randomized and non-randomized trial design issues for medical device studies.

**ALEX JOHN LONDON, Ph.D.**, is the Clara L. West Professor of Ethics and Philosophy and Director of the Center for Ethics and Policy at Carnegie Mellon University. An elected Fellow of the Hastings Center, Professor London's work focuses on ethical and policy issues surrounding the development and deployment of novel technologies in medicine, biotechnology and artificial intelligence. His papers have appeared in *Mind*, *Science*, *JAMA*, *The Lancet*, *The MBJ*, *PLoS Medicine*, *Statistics In Medicine*, and numerous other journals and collections. He is also co-editor of *Ethical Issues in Modern Medicine*, one of the most widely used textbooks in medical ethics. Professor London has helped to shape key ethical guidelines for the oversight of research with human participants for over a decade. From 2012-2016 he was a member of the Working Group on the Revision of CIOMS 2002 International Ethical Guidelines for Biomedical Research Involving Human Subjects. Prior to that he was an expert commentator at three World Medical Association meetings for the revision of the 2013 Declaration of Helsinki. From 2007-2018 he was a member of the ethics working group of the U.S. HIV Prevention Trials Network where he was part of the group that drafted the HIV Prevention Trials Network Ethics Guidance for Research. From 2016-2017 he was part of the U.S. National Academies of Sciences, Engineering, and Medicine Committee on Clinical Trials During the 2014-15 Ebola Outbreak.

**DAVID MARTIN, M.D., M.P.H.**, is the Associate Director for Real-World Evidence Analytics with the Office of Medical Policy, Center for Drug Evaluation and Research, Food and Drug Administration. Previously, he was assigned to FDA/CDER as the FDA Liaison to the Reagan Udall Foundation Innovation in Medical Evidence Development and Surveillance program. IMEDS enables routine private-sector queries related to medical products using a distributed database approach modeled on the FDA's Sentinel system. Dr. Martin is also the principal investigator for a Patient Centered Outcomes Trust Fund project that is incorporating patient data collected through a mobile device application into Sentinel and PCORnet. As a former Branch Chief and Division Director in the FDA Center for Biologics Evaluation and Research, he led analyses of spontaneous reports, formalized risk management planning, and played a key role in the development of the Sentinel system. Before joining the FDA, Dr. Martin served in the U.S. Air Force as a flight and occupational medicine physician. He

received his bachelor's degree at the Citadel, his medical degree at the Johns Hopkins University School of Medicine, and his master of public health degree at the Johns Hopkins University Bloomberg School of Public Health.

**W. BENJAMIN NOWELL, Ph.D.**, is the Director of Patient-Centered Research at the Global Healthy Living Foundation (GHLF), CreakyJoints and co-Principal Investigator of ArthritisPower Patient-Powered Research Network with Jeffrey Curtis, MD, MS, MPH (University of Alabama at Birmingham) and GHLF patient co-founder Seth Ginsberg. Dr. Nowell leads all research activities conducted by the organization, including facilitating studies conducted with academic and industry partners. Recent projects include rheumatoid arthritis (RA) shared decision making projects with Dr. Liana Fraenkel, M.D., M.P.H., at Yale and a survey of the family planning concerns of patients with inflammatory arthritis with Dr. Megan Clowse, M.D., M.P.H., at Duke. Dr. Nowell also leads a hip/knee joint replacement patient research engagement project called BeTTER SAID with co-PI Thomas Concannon, Ph.D., at RAND. His research interests include examination of the factors that facilitate patient engagement as research partners, and patient-reported outcomes and shared decision making in rheumatologic conditions. Prior to joining GHLF, Dr. Nowell worked as a medical social worker and Community and Long-Term Care Coordinator for the Ottawa Regional Stroke Centre and as Research Coordinator for an evaluation of participant outcomes in Arthritis Foundation chronic disease self-management programs for arthritis.

**JEFFREY E. SHUREN, M.D., J.D.**, became the director of the Center for Devices and Radiological Health at the Food and Drug Administration (FDA) in January 2010. He previously served as Acting Center Director, beginning in September 2009. The center is responsible for assuring the safety, effectiveness, and quality of medical devices; assuring the safety of radiation-emitting products (such as cell phones and microwave ovens); and fostering device innovation. Dr. Shuren received his B.S. and M.D. degrees from Northwestern University under its Honors Program in Medical Education. He completed his medical internship at Beth Israel Hospital in Boston, his neurology residency at Tufts New England Medical Center, and a fellowship in behavioral neurology and neuropsychology at the University of Florida. He received his J.D. from the University of Michigan. Dr. Shuren has held various policy and planning positions within FDA from 1998 to 2009, including acting deputy commissioner for policy, planning, and budget; associate commissioner for policy and planning; special counsel to the principal deputy commissioner; assistant commissioner for policy; and medical officer in the Office of Policy. Dr. Shuren has served in a leadership role at FDA or on behalf of the agency on numerous initiatives, including reauthorization of the Medical Device User Fee Act, which dramatically shortens review times for device applications; creation of the Sentinel Initiative, which works toward a national electronic system for monitoring medical product safety; development of FDA's Pandemic Influenza Preparedness Strategic Plan; development of FDA's Counterfeit Drug Task Force Report; development of the Interagency Food Safety Working Report to the President; implementation of FDA provisions of the Medicare Prescription Drug Improvement and Modernization Act; development and implementation of the Interagency Import Safety Working Group's Report to the President: Action Plan for Import Safety. From 1999 to 2000, Dr. Shuren served as a detailee on Senator Edward Kennedy's staff on the Senate Health, Education, Labor, and Pensions Committee. From 1998 to 2003, he also was a staff volunteer in the National Institutes of Health's Cognitive Neuroscience Section where he supervised and designed clinical studies on human reasoning. As director of the Division of Items and Devices, Coverage and Analysis Group at the Centers for Medicare and Medicaid Services, Dr. Shuren oversaw the development of Medicare national coverage determinations for drugs, biologics, and non-implantable devices.

**JAMES P. SMITH, M.D., M.S.**, is the Deputy Director of the Division of Metabolism and Endocrinology Products (DMEP) within the Office of New Drugs, Center for Drug Evaluation and Research (CDER), FDA. In this capacity, he primarily oversees development programs targeting lipid disorders and obesity. Prior to joining FDA in February 2011, he was a faculty member in the Division of Nephrology of the University of Michigan Health System. Dr. Smith is a graduate of the University of Michigan Medical School, and he completed his residency in Internal Medicine at the same institution. Subsequently, he completed fellowships in both nephrology and clinical pharmacology at Vanderbilt University Medical Center, as well as a master's degree in Clinical Research Design and Statistical Analysis at the University of Michigan School of Public Health.

**PETER P. STEIN, M.D.**, serves as Deputy Director, Office of New Drugs (OND), Center for Drug Evaluation and Research (CDER), Food and Drug Administration (FDA). Dr. Stein earned his medical degree from University of Pennsylvania and trained at Yale-New Haven Hospital in internal medicine, and in endocrinology and metabolism. He was on faculty at Yale in the Section of Endocrinology and served as the associate program director for the Primary Care Residency Program. Subsequently, Dr. Stein was the section chief for endocrinology and the program director for internal medicine residency program at the Medical College of Georgia. Dr. Stein joined Bristol-Myers Squibb in 1999, subsequently working at Merck, Janssen, and finishing his career in industry at Merck as Vice-president for late-stage development in Diabetes and Endocrinology. During his industry career, Dr. Stein led development programs for several currently approved diabetes medications, and has worked on a wide range of programs from discovery through early and late clinical development. He is a clinical associate professor at the Robert Wood Johnson Medical School, where he maintained a practice in endocrinology for many years. Dr. Stein joined FDA in late 2016 as the Deputy Director, Office of New Drugs, CDER.

**KOMATHI STEM, M.A.**, is passionate about transforming drug development by rethinking the way clinical trials are designed and executed. As an entrepreneurial systems thinker, she is focused on sparking an ecosystem change that accelerates research through greater collaboration across the healthcare ecosystem. She is currently the Founder and CEO of monARC Bionetworks, a healthcare data collection and analytics company focused on helping patients share their health and digital data to accelerate research and participate in clinical trials anywhere, anytime. She has more than 25 years of pharmaceutical and biotech industry experience as a senior global leader at Genentech, AstraZeneca and Eli Lilly, where she built and led multi-functional global departments. As a Global Innovation Leader for Genentech, she has tested and transferred into the business, patient centric solutions for accelerating clinical research and reducing costs. Several of these solutions have sparked new startups on the use of digital tools in clinical trials. Ms. Stem is a biomedical engineer with diverse pharma/biotech leadership experience ranging from global clinical operations, strategic innovation, portfolio management, medical affairs and sales/marketing. She has extensive experience in multiple therapeutic areas including: Oncology, Neuroscience, Endocrinology, Immunology, Cardiology, Respiratory, Infectious Diseases, Gastroenterology and Rare Diseases.

**MARK J. VAN DER LAAN, Ph.D., M.S.**, is the Hsu/Peace Professor of Biostatistics at the University of California, Berkeley School of Public Health. He is the recipient of the 2005 COPSS Presidents' and Snedecor Awards, as well as the 2004 Spiegelman Award, and is a Founding Editor for the International Journal of Biostatistics and the Journal of Causal Inference. His methodological research interests include censored data, causal inference, machine learning, multiple testing, semiparametric estimation theory, and targeted learning. He has authored various books, including Springer books Targeted Learning: Causal Inference for Observational and Experimental Data (2011), and the upcoming book Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies (2017), van der Laan, Rose.

**ORLY VARDENY, Pharm.D., M.S.**, is an Investigator at the Center for Chronic Disease Outcomes Research at the Minneapolis VA Health Care System and an Associate Professor of Medicine at the University of Minnesota Medical School. Dr. Vardeny's research interests include maximizing the benefit of pharmacologic therapy in patients with heart failure, and optimizing vaccination strategies in patients with cardiac disorders. She is a co-principal investigator on a 9300-patient NIH-funded clinical trial investigating vaccination strategies in patients with high-risk cardiovascular disease.

**JONATHAN H. WATANABE, Pharm.D., Ph.D.**, is an associate professor of clinical pharmacy at the University of California San Diego Skaggs School of Pharmacy and Pharmaceutical Sciences and is the National Academy of Medicine Anniversary Pharmacy Fellow. He was a contributor to the National Academy of Sciences, Engineering, and Medicine report on Ensuring Patient Access to Affordable Medications. Dr. Watanabe examines large, real-world data with the goal of developing policy solutions to improve patient care, augment population health, and reduce medical costs. Dr. Watanabe focuses on improving access to evidence-driven medication use and pharmacist-directed patient care. He serves as an advisor to the California Health Benefits Review Program for the California State Legislature. He is an investigator, faculty, and fellowship director for the federal Health Resources and Services Administration funded San Diego Geriatrics Workforce Enhancement Program and is also supported by the NIH National Institute on Aging to examine high-risk

medication use and costs in older adults. Dr. Watanabe was the inaugural recipient of the University of Washington/Allergan Global Health Economics and Outcomes Research Fellowship. He is a Board Certified Geriatric Pharmacist (BCGP). Dr. Watanabe is a clinical consultant at the St. Paul's Program of All-inclusive Care for the Elderly (PACE) Clinic in San Diego, CA and the Villa Pomerado Skilled Nursing Facility in Poway, CA. He is an advisor to the Joint Commission on pain management and assessment standards in long-term care. He received his B.S. from the University of Washington, his Pharm.D. from the University of Southern California, and his M.S. and Ph.D. from the University of Washington Comparative Health Outcomes, Policy, and Economics (CHOICE) Institute.

**BRANDE ELLIS YAIST, M.S.,** is the Senior Director of the Center of Expertise in Global Patient Outcomes and Real World Evidence (GPORWE) at Eli Lilly and Company. She leads and develops the research talent and capabilities needed to provide scientific services/expertise and support across an array of core Health Economics & Outcomes Research (HEOR) areas of focus, which include observational studies and real world evidence (RWE), Clinical Outcomes Assessment (COA), early phase health outcomes research strategy, health outcomes and health outcomes research policy. Mrs. Yaist received her B.S. in Nuclear Medicine Technology from St. Louis University and a Master of Health Science from Washington University in St. Louis. She has been with Eli Lilly and Company for 16 years. Her experience at Lilly includes oversight and monitoring of early and late phase clinical trials, supervision, medical planning for compounds marketed in the US, and strategy and implementation for Risk Evaluation and Mitigation Strategies (REMS) and Risk Minimization plans globally. Mrs. Yaist led several projects as a Lean Six Sigma Black Belt, affecting areas of clinical development, regulatory affairs, and global patient safety. Prior to her career at Lilly, she was a Clinical Research Coordinator at Siteman Cancer Center in St. Louis, MO.



## Establishing a Framework to Evaluate Real-World Endpoints

July 10, 2018  
Washington, DC

### Introduction

Advances in data analytics and data capture through electronic health records (EHRs) and medical/pharmacy claims have brought the opportunities and challenges associated with using real-world evidence (RWE) to the forefront of the US healthcare industry. Increasingly, the promise of RWE to contribute to a more complete picture of the benefits and risks associated with therapies, when paired with results from randomized, controlled clinical trials, is being realized. RWE provides an opportunity to collect data rapidly on a broader patient population outside of a strict clinical trial protocol to help identify new indications or rare safety events, provide more generalizability of clinical trial results, and confirm clinical benefit in the post-market setting. Further, integration of the various sources of real-world data (RWD), including EHRs, clinical decision and support and hospital-based systems, administrative billing and claims databases, patient registries, longitudinal cohort studies, and patient reported outcomes tools, will yield a more robust dataset of RWE. However, the methods to aggregate data and the implications of integrating these multiple data sets as they evolve (especially in often dynamic post-approval settings) needs to be validated.

Applications for RWE extend the spectrum of therapeutics development from regulatory decision-making, to clinical use, to coverage and payment decisions. In the regulatory space, RWE has been utilized most frequently to evaluate drug safety through pharmacovigilance and adverse event monitoring in pre- and post-approval settings. However, RWE has increasingly been used to support effectiveness studies, in the form of historical data, as a surrogate for control arms in clinical trials (in the rare disease setting, for instance). Beyond regulatory decisions, RWE is frequently used to support clinical trial design, development of clinical practice guidelines, confirmation of population/subgroup size, and payment decisions including formulary placement.

These current applications of RWE in healthcare are quite limited with respect to the potential uses once appropriate standards and guardrails are implemented. Indeed, the pharmaceutical industry, FDA, and Congress recognize the importance of further developing this resource as evidenced by numerous recent publications by the FDA, passage of the 21<sup>st</sup> Century Cures Act (Cures Act), and the Prescription Drug User Fee Act (PDUFA) VI reauthorization. The Cures Act, passed in December of 2016, requires the FDA to develop a framework and issue guidance regarding the use of RWE to support a new indication for an already approved drug or post-market studies as a requirement for regulatory approval. Interestingly, FDA has already issued similar guidance regarding use of RWE for medical devices which includes supporting new indications and in post-approval studies. PDUFA VI builds upon the requirements of the Cures Act by instructing the FDA to consider stakeholder input through hosting of

public workshops as it develops its guidance for use of RWE. Other uses of RWE that could be imagined for future pharmaceutical approvals include expanded labels, pragmatic clinical trial design, and confirming benefit in the case of converting an Accelerated Approval to full approval status. In addition to potential regulatory uses, RWE could provide helpful information about the long-term value of a product and could inform future value assessments. For example, long-term efficacy endpoints that may not have been incorporated in pre-market clinical trial might be able to be captured using RWE, which requires increased understanding of how time-on-treatment or treatment discontinuation rates correlate to overall survival.

Significant progress has been made in data collection efforts to support use of RWE in regulatory settings, however challenges remain, chiefly with combining, organizing, and analyzing data from various information sources. Friends of Cancer Research proposes a pilot project, comprised of six leading healthcare data organizations, to develop a dataset curation process and validation framework to operationalize RWD collection and explore potential real-world endpoints that may be fit for regulatory purposes as well as assessing long-term benefits of a product.

## Pilot Project Overview

Immunotherapies are being used to treat patients with cancers that have historically had few treatment options, which has generated high level of interest in their use and development. While immunotherapies have resulted in significant improvements in some patients, many other patients do not respond or only respond for a limited time. This has raised questions about the value of these new drugs. Applying current value frameworks to immune checkpoint inhibitors has proved difficult as they tend to underestimate the benefits of long-term survival and treatment-free survival.<sup>1</sup> This is likely due to the reliance on pivotal trial data, and in the setting of expedited approvals, assessments of the full clinical endpoints have not been completed. Thus, conclusions are often based on surrogate efficacy endpoints. At the initiation of this pilot project, three immune checkpoint inhibitors were approved for use in non-small cell lung cancer (NSCLC), which presented an opportunity to collect a robust amount of data for analysis from the post-market setting.

This pilot project was initiated to help determine whether RWD can be used to develop an early perspective on real-world outcomes, as defined by real-world endpoints from EHR and claims data, and whether these data correlate to overall survival (OS) in the context of randomized control trials (RCTs) for patients treated with novel therapies. The pilot project evaluates the performance of real-world endpoints across multiple data sets by focusing on a common question: ***What outcomes can be evaluated for aNSCLC patients treated with immune checkpoint inhibitors?***

To answer this question, a framework of necessary data elements, characteristics, and internal validation processes were proposed along with a set of definitions for real-world endpoints in the

---

<sup>1</sup> Ben-Aharon O, Magnezi R, Leshno M, Goldstein DA. Association of Immunotherapy with Durable Survival as Defined by Value Frameworks for Cancer Care. JAMA Oncol. 2018, 4(3):326–332.



context of their use in RCTs, FDA’s regulatory framework, and data availability in EHR and claims systems. The pilot project will help evaluate whether the various data sets included in this study can achieve a similar level of correlation and statistical significance using a common framework.

## Pilot Project Study Design and Objectives

This is a retrospective observational analysis of data derived from EHR and claims data. The data sets generated for the study include all relevant, retrospective patient-level data available for eligible individuals up to the data cutoff date, pending approval by a third-party de-identification.

**Objective 1:** Describe the demographic and clinical characteristics of aNSCLC patients treated with immune checkpoint inhibitors (Table 1)

**Objective 2:** Assess ability to generate real-world endpoints (OS, rwPFS, rwTTP, TTNT, TTD) in aNSCLC patients treated with immune checkpoint inhibitors, and segmented by clinical and demographic characteristics (Tables 2, 3, and 4)

**Objective 3:** Assess performance of real-world endpoints (rwPFS, rwTTP, TTNT, TTD) as surrogate endpoints for OS (Table 5)

## Methods

| Project Details                                  |   |
|--|---|
| <b>Cohort and inclusion / exclusion criteria</b> | <p>aNSCLC patients treated with an immune checkpoint inhibitor (i.e., nivolumab, pembrolizumab, atezolizumab)</p> <p><u>Inclusion:</u></p> <ul style="list-style-type: none"> <li>• At least two documented clinical visits on or after January 1, 2011 until data cutoff date</li> <li>• Pathology consistent with NSCLC<sup>2</sup></li> <li>• Has evidence of IIIB or IV NSCLC or has early stage NSCLC with a recurrence or progression described/documentated in the EHR or claims</li> <li>• Treatment with immune checkpoint inhibitor, as documented by a structured medication order or claim as evidence of having received the treatment</li> </ul> <p><u>Exclusion:</u></p> <ul style="list-style-type: none"> <li>• Incomplete historical treatment data available within the database (i.e., patients whose advanced diagnosis date is more than 90 days before first activity date)</li> </ul> |

<sup>2</sup> For claims data, to minimize misclassification of aNSCLC, treatment with an IO agent following diagnosis of lung cancer was required. During the timeframe of this project, coverage for IO agents required evidence of advanced disease defined as either stage IIIB or IV NSCLC at initial diagnosis or early stage (stages I, II, and IIIA) NSCLC with a recurrence or progression.

**EHR and Claims-derived endpoints definition and analytical guidance**

**Overall survival (OS)**

- *Data definition / computation:* length of time from the date the patient initiates the study treatment to the date of death or proxied by time to disenrollment. Patients without a date of death will be censored at their last known activity or date of disenrollment from the health plan identified and categorized as “due to death” if the date of death captured by SSA DMF was within 30 days prior or 60 days following.

**Time to Next Treatment (TTNT)**

- *Data definition / computation:* length of time from the date the patient initiates the study treatment to the date the patient initiates their next systemic treatment. When subsequent treatment is not received (e.g., continuing current treatment or disenrollment not due to confirmed death), patients will be censored at their last known activity.
- Start date of regimen immediately after PD-(L)1 line (i.e., the subsequent systemic therapy after the initial PD-(L)1-containing regimen)

**Time to Treatment Discontinuation (TTD)**

- *Data definition / computation:* length of time from the date the patient initiates the PD-(L)1 regimen to the date the patient discontinues treatment. Patients still on treatment will be censored at their last known activity.
- Event Date: Date of PD-(L)1 regimen discontinuation defined as last administration or non-cancelled order of a drug contained within the PD-(L)1 line regimen (between the line’s start and end date) among patients that discontinued their immune checkpoint inhibitor therapy. Permanent discontinuation is defined as meeting one of the following conditions:
  - Having a subsequent systemic therapy after the initial PD-(L)1-containing regimen
  - Having a date of death while on the PD-(L)1-containing regimenHaving a gap of more than 120 days between the last administration or non-cancelled order of the PD-(L)1 line and the patient’s last visit or medication administration if there is no other systemic therapy after the PD-(L)1-containing regimen
- Censor date: Patients without a discontinuation will be censored at their last known PD-(L)1 usage defined as the last administration or non-cancelled order of a drug contained within the PD-(L)1 regimen

**Progression Event**

- *Data definition / computation:* distinct episode in which the treating clinician concludes that there has been growth or worsening in the aNSCLC. The progression event (and date) is based on review of the patient chart.

**Real-world Progression Free Survival (rwPFS)**

- *Data definition / computation:* length of time from the date the patient initiates the PD-(L)1 regimen to the date that a progression event as evident in the EHR is documented in the patient’s chart or the patient passes away.

|                        |  |
|------------------------|--|
|                        | <p>Patients without a progression event or date of death will be censored at the end of the patient’s chart.</p> <p><b>Real-world Time to Progression (rwTTP)</b></p> <ul style="list-style-type: none"> <li>• <i>Data definition / computation:</i> length of time from the date the patient initiated the PD-(L)1 regimen to the date that a progression event is documented in the patient’s EHR (excludes death as an event). Patients without a progression event will be censored at the end of the patient’s chart.</li> <li>• Event date: Patient’s first progression date more than 14 days after PD-(L)1 initiation as described in the index date definition. Death will not be considered a progression event in TTP</li> <li>• Censor date: Patients without a progression date more than 14 days after the index date or date of death (for PFS) will be censored at the last date the patient could have been assessed for progression (e.g., last clinic note date)</li> </ul> <p><b>Index Date</b></p> <ul style="list-style-type: none"> <li>• <i>Data definition / computation:</i> the earliest PD-(L)1 inhibitor initiation in the advanced setting anchored to start (e.g., first administration or non-cancelled order) of the immune-checkpoint inhibitor-containing regimen (nivolumab, pembrolizumab, atezolizumab).</li> </ul>                          |
| <p><b>Analyses</b></p> | <p><u>Table 1:</u></p> <ul style="list-style-type: none"> <li>• Assess ability to identify aNSCLC patients treated with immune checkpoint inhibitors</li> <li>• Description of demographic and clinical characteristics of aNSCLC patients treated with immune checkpoint inhibitors, example characteristics include: <ul style="list-style-type: none"> <li>○ Demographic: gender, age, SES, region</li> <li>○ Clinical: histology, smoking status, group stage at time of initial diagnosis, follow up, biomarker status (e.g., ALK, EGFR, PD-L1), hepatic and renal function</li> </ul> </li> <li>• Description of population characteristics for overall population and by treatment setting / line of therapy (e.g., 1st line metastatic, 2nd line, 3rd line plus)</li> </ul> <p><u>Table 2, Table 3, and Table 4:</u></p> <ul style="list-style-type: none"> <li>• Assess ability to generate real-world endpoints (OS, TTNT, TTD) for aNSCLC patients treated with immune checkpoint inhibitors within the advanced treatment setting (range and median figures)</li> <li>• Evaluate these endpoints when patient cohort is segmented by treatment setting and demographic /clinical characteristics</li> </ul> <p><u>Table 5:</u></p> <ul style="list-style-type: none"> <li>• Assess correlation of real-world endpoints (TTNT, TTD) to overall survival (OS)</li> </ul> |

## **Summary of Data Sources for Pilot Project Study**

### **Cancer Research Network**

The Cancer Research Network originated as an NCI-funded consortium of research groups affiliated with a dozen integrated health care systems across the US, and among whom the Health Care Systems Research Network was formed. In the early 2000's, the CRN created the Virtual Data Warehouse (VDW), a common data model to facilitate collaborative research. Data in the VDW are extracted from multiple source databases and maintained by each research group with the possibility of pooling data under specific IRB-approved research protocols. For most participating institutions, the VDW has essentially complete information on care dating back to 1996 or earlier for most data domains. Domains include health plan enrollment periods, cancer registries, encounters including diagnoses and procedures, prescription and infusion medications, laboratory results, and other areas. The data provided are results from one of the participating CRN organizations.

### **Cota**

The Cota Real-World Evidence (RWE) database is a HIPAA-compliant, de-identified data source drawn from the electronic health records (EHR) of contributing academic, for-profit, and community oncologist provider sites and hospital systems. The database includes detailed demographic, diagnostic, molecular and genomic testing, treatment, and outcome data. As of 2018, Cota's RWE is comprised of rich longitudinal patient records collected from over 40 unique locations across North America. For the purposes of this pilot study, patient data was sourced from a predominantly community setting (98%).

### **Flatiron Health**

Flatiron Health is a longitudinal, demographically and geographically diverse database derived from electronic health record (EHR) data from over 265 cancer clinics (~800 sites of care) including more than 2 million active US cancer patients available for analysis. The patient-level data in the EHRs includes structured data (e.g., laboratory values, and prescribed drugs) in addition to unstructured data collected via technology-enabled chart abstraction from physician's notes and other unstructured documents (e.g., biomarker reports).

### **IQVIA™**

IQVIA™ is a leading global provider of information, innovative technology solutions and contract research services focused on using data and science to help healthcare clients find better solutions for their patients. For this engagement, IQVIA provided data sourced through Oncology Electronic Medical Records (EMR) from multiple partners, including TransMed. The data are comprised predominately of community practices (90%+). The integrated EMR platform includes activity from all payer types and all practice sizes across the United States. Results for this analysis were calculated primarily based on structured EMR fields.

### **Mayo Clinic Analysis using OptumLabs® Data Warehouse**

OptumLabs® is an open, collaborative research and innovation center founded in 2013 as a partnership between Optum and Mayo Clinic. Its core linked data assets include de-identified claims data for privately insured and Medicare Advantage enrollees and de-identified electronic health record (EHR)

data from a nationwide network of provider groups. This pilot project was a retrospective analysis of claims data from the OptumLabs® Data Warehouse (OLDW), which includes de-identified claims data for privately insured and Medicare Advantage enrollees in a large, private, U.S. health plan. The database contains longitudinal health information on enrollees, representing a diverse mixture of ages, ethnicities and geographical regions across the United States. The health plan provides comprehensive full insurance coverage for physician, hospital, and prescription drug services.

### **PCORnet Sites**

This pilot project included 11 PCORnet partner sites who had previously participated in a PCORnet Rapid Cycle Project. The 11 sites are based in healthcare systems within three PCORnet networks across 10 US states and include 10 academic medical centers. These sites were selected from 80 PCORnet partner sites because they could rapidly provide tumor registry data and linked electronic health records in PCORnet Common Data Model (CDM) format. The pooled database contributed to the RWE Endpoints Pilot Project consisted of tumor registry data from each site and linked CDM diagnosis, procedures, prescribing, dispensing, medication administration, and death data tables. Data sources for the CDM include institutional billing and electronic health record data. The study cohort includes patients with a single primary advanced stage non-small cell lung cancer (NSCLC) diagnosis who were either diagnosed at stage 3b or 4 or who had an ICD9/10 diagnosis code for secondary metastasis.

**Table 1. Description of demographic and clinical characteristics of aNSCLC patients treated with PD-(L)1 checkpoint inhibitors**

| Demographics  | Data Set A<br>PD-(L)1-treated<br>N=2595 | Data Set B<br>PD-(L)1-treated N=557 | Data Set C<br>PD-(L)1-treated N=435 | Data Set D<br>PD-(L)1-treated<br>N=6924 | Data Set E<br>PD-(L)1-treated<br>N=2860 | Data Set F<br>PD-(L)1-treated N=269 |
|---|---|-------------------------------------|-------------------------------------|---|---|-------------------------------------|
| Age at advanced diagnosis (years), median [IQR]               | 68 [15]                                 | 64 [14]                             | 66 [14]                             | 69 [14]                                 | 68 [14]                                 | 70 [14]                             |
| Age at PD-(L)1 inhibitor initiation (years), median [IQR]     | 69 [14]                                 | 65 [14]                             | 68 [14]                             | 69 [14]                                 | 69 [14]                                 | 71 [14]                             |
| Age categories at PD-(L)1 inhibitor initiation (categorical): |   |                                     |                                     |   |   |                                     |
| ≤49 years   | 120 (5%)                                | 24 (4%)                             | 21 (5%)                             | 219 (3%)                                | 80 (3%)                                 | 8 (3%)                              |
| 50-64 years   | 888 (34%)                               | 251 (45%)                           | 129 (30%)                           | 2048 (30%)                              | 863 (30%)                               | 65 (24%)                            |
| 65-74 years   | 866 (33%)                               | 198 (36%)                           | 169 (39%)                           | 2504 (36%)                              | 1047 (37%)                              | 94 (35%)                            |
| 75+ years   | 721 (28%)                               | 84 (15%)                            | 116 (27%)                           | 2153 (31%)                              | 870 (30%)                               | 102 (38%)                           |
| Age categories at PD-(L)1 inhibitor initiation (binary):      |   |                                     |                                     |   |   |                                     |
| <75 years   | 1874 (72%)                              | 473 (85%)                           | 319 (73%)                           | 4771 (69%)                              | 1990 (70%)                              | 167 (62%)                           |
| 75+ years   | 721 (28%)                               | 84 (15%)                            | 116 (27%)                           | 2153 (31%)                              | 870 (30%)                               | 102 (38)                            |
| Gender:   |   |                                     |                                     |   |   |                                     |
| Female  | 1147 (44%)                              | 276 (50%)                           | 212 (49%)                           | 3172 (46%)                              | 1351 (47%)                              | 125 (46%)                           |
| Male  | 1448 (56%)                              | 281 (50%)                           | 222 (51%)                           | 3752 (54%)                              | 1509 (53%)                              | 143 (53%)                           |
| Unknown/Missing   | 0                                       | 0                                   | ≤5                                  | 0                                       | 0                                       | 1                                   |
| Race/ethnicity:   |   |                                     |                                     |   |   |                                     |
| White   | 1704 (78%)                              | 478 (86%)                           | 284 (65%)                           | 4969 (79%)                              | 676 (87%)                               | 160 (87%)                           |
| Black or African American                                     | 282 (13%)                               | 67 (12%)                            | 37 (9%)                             | 594 (9%)                                | 44 (6%)                                 | 14 (8%)                             |
| Asian   | 52 (2%)                                 | 6 (1%)                              | 83 (19%)                            | 155 (3%)                                | 13 (2%)                                 | 9 (5%)                              |
| Other Race  | 142 (7%)                                | 6 (1%)                              | 31 (7%)                             | 580 (9%)                                | 42 (5%)                                 | 1 (1%)                              |
| Unknown/Missing   | 415                                     | 0                                   | 0                                   | 626                                     | 2085                                    | 85                                  |
| Median household income (zip-level):                          |   |                                     |                                     |   |   |                                     |
| 1 (lowest median household income)                            |   |                                     | 103 (24%)                           | 1003 (15%)                              |   |                                     |
| 2   |   |                                     | 105 (24%)                           | 1539 (22%)                              |   |                                     |
| 3   |   |                                     | 114 (26%)                           | 1833 (27%)                              |   |                                     |
| 4 (highest median household income)                           |   |                                     | 112 (26%)                           | 2525 (37%)                              |   |                                     |
| Unknown   |   |                                     | ≤5                                  | 24                                      |   |                                     |
| <b>CLINICAL CHARACTERISTICS</b>                               |   |                                     |                                     |   |   |                                     |
| Group stage at initial diagnosis:                             |   |                                     |                                     |   |   |                                     |
| Stage 0 / Occult  |   | 0                                   |                                     | 2 (0%)                                  |   |                                     |
| Stage I   |   | 23 (6%)                             |                                     | 496 (7%)                                |   | 18 (7%)                             |
| Stage II  |   | 22 (6%)                             |                                     | 426 (6%)                                |   | 17 (7%)                             |
| Stage III   |   | 88 (23%)                            | 39 (9%)                             | 1494 (22%)                              |   | 17 (7%)                             |
| Stage IV  |   | 248 (65%)                           | 396 (91%)                           | 4335 (64%)                              |   | 161 (62%)                           |
| Group stage is not reported                                   |   | 176                                 |                                     | 171                                     |   | 10                                  |
| Histology:  |   |                                     |                                     |   |   |                                     |
| Non-squamous cell carcinoma                                   |   | 370 (66%)                           | 320 (74%)                           | 4679 (70%)                              | 1981 (69%)                              | 194 (73%)                           |
| Squamous cell carcinoma                                       |   | 147 (26%)                           | 73 (17%)                            | 1983 (30%)                              | 659 (23%)                               | 61 (23%)                            |

|   |            |           |            |            |               |
|---|------------|-----------|------------|------------|---------------|
| NSCLC histology not otherwise specified (NOS)   | 40 (7%)    | 42 (10%)  | 262 (3%)   | 220 (8%)   | 10 (4%)       |
| Missing   |            |           |            |            | 4             |
| <b>Smoking status:</b>  |            |           |            |            |               |
| History of smoking  |            | 340 (78%) | 6185 (90%) | 448 (92%)  | 182 (87%)     |
| No history of smoking   |            | 94 (22%)  | 717 (10%)  | 38 (8%)    | 28 (13%)      |
| Unknown/Not documented  |            | ≤5        | 22         | 2374       | 210           |
| PD-L1 tested on or prior to PD-(L)1 inhibitor start                                       | 326 (13%)  |           | 2384 (34%) | 96         | 80/96 (83%)   |
| <b>PD-L1 expression status (among those tested):</b>                                      |            |           |            |            |               |
| PD-L1 positive  |            |           | 512 (22%)  | 45 (50%)   | 65 (68%)      |
| PD-L1 negative/not detected   |            |           | 691 (29%)  | 45 (50%)   | 29 (30%)      |
| Unsuccessful/indeterminate test   |            |           | 1012 (42%) | 0          | 2 (2%)        |
| Results pending/unknown   |            |           | 169 (7%)   | 6          | 173           |
| ALK tested on or prior to PD-(L)1 inhibitor start   | 258 (10%)  |           | 4513 (65%) | 582        | 143/173 (83%) |
| <b>ALK status (among those tested):</b>   |            |           |            |            |               |
| Rearrangement present   |            |           | 57 (1%)    | 8 (1%)     | 1 (1%)        |
| Rearrangement not present   |            |           | 4145 (92%) | 570 (99%)  | 170 (98%)     |
| Results pending/unknown   |            |           | 68 (2%)    | 0          | 2 (1%)        |
| Unsuccessful/indeterminate test   |            |           | 243 (5%)   | 4          | 96            |
| EGFR tested on or prior to PD-(L)1 inhibitor start  | 543 (21%)  | 171(39%)  | 4684 (68%) | 953        | 115/142 (81%) |
| <b>EGFR status (among those tested)<sup>2,3</sup>:</b>                                    |            |           |            |            |               |
| Mutation positive   |            |           | 305 (7%)   | 68 (11%)   | 6/142 (4%)    |
| Mutation negative   |            |           | 4161 (89%) | 525 (89%)  | 135/142 (95%) |
| Results pending/unknown   |            |           | 60 (1%)    | 358        | 1/142 (1%)    |
| Unsuccessful/indeterminate test   |            |           | 158 (3%)   | 2          | 127           |
| No prior therapy received   | 690 (27%)  | 80 (18%)  | 2074 (30%) | 777 (27%)  | 77 (29%)      |
| <b>Line number of first PD-(L)1 inhibitor in advanced setting:</b>                        |            |           |            |            |               |
| 1   | 690 (27%)  | 80 (18%)  | 2074 (30%) | 777 (27%)  | 77 (29%)      |
| 2   | 1440 (56%) | 205 (47%) | 3357 (49%) | 1414 (49%) | 87 (32%)      |
| 3   | 380 (15%)  | 85 (20%)  | 1012 (15%) | 448 (16%)  | 51 (19%)      |
| 4+  | 85 (3%)    | 65 (15%)  | 481 (7%)   | 221 (8%)   | 54 (20%)      |
| <b>Patients receiving a second PD-(L)1 inhibitor in a subsequent line:</b>                |            |           |            |            |               |
| No  |            | 402 (92%) | 1740 (25%) |            |               |
| No subsequent therapy received  |            |           | 4879 (71%) |            |               |
| Yes   | 93 (4%)    | 33 (8%)   | 305 (4%)   | 112        | 14            |
| <b>Line number of second PD-(L)1 inhibitor in advanced setting:</b>                       |            |           |            |            |               |
| 2   | 28 (30%)   | 11 (33%)  | 99 (33%)   | 9 (8%)     | 5 (36%)       |
| 3   | 45 (48%)   | 10 (30%)  | 134 (44%)  | 51 (46%)   | 4 (29%)       |
| 4+  | 20 (22%)   | 12 (36%)  | 72 (24%)   | 52 (46%)   | 5 (36%)       |
| N/A   |            | 402       |            |            |               |
| Time from advanced diagnosis to first PD-(L)1 inhibitor initiation (months), median [IQR] | 7 [11]     | 8 [11]    | 6 [11]     | 8 [14]     | 7 [12]        |



| Structured follow up time <sup>3</sup>   |         |         |         |         |         |
|--|---------|---------|---------|---------|---------|
| Structured follow-up time from advanced diagnosis (months), median [IQR]           | 18 [18] | 18 [21] | 14 [17] | 18 [20] | 18 [18] |
| Structured follow-up time from PD-(L)1 inhibitor initiation (months), median [IQR] | 8 [13]  | 9 [13]  | 6 [10]  | 8 [11]  | 8 [9]   |

<sup>3</sup> Structured follow-up time is calculated from the relevant time-point for each patient until their last structured activity (i.e., most recent visit or administration)

**Table 2. Median time and 95% confidence interval for real-world extracted endpoints**

| Data Set   | rwOS  | rwTTNT                | rwTTD             | rwTTP              | rwPFS              |
|------------|---|-----------------------|-------------------|--------------------|--------------------|
| Data Set A | 13.50 [12.80, 14.50] <sup>4</sup>                       | 22.50 [NA]            | 7.03 [6.27, 9.97] |                    |                    |
| Data Set B | 15.78 [12.2, 24.59];<br>8.58 [7.56, 10.26] <sup>5</sup> |                       | 3.25[2.76, 3.75]  |                    |                    |
| Data Set C | 8.67 [6.83, 10.02]                                      | 11.60 [8.80, 16.10]   | 4.70 [3.68, 5.52] |                    |                    |
| Data Set D | 9.15 [8.82, 9.51]                                       | 14.03 [ 12.89, 15.15] | 3.21 [3.21, 3.44] | 5.41 [5.18, 5.67]  | 3.28 [3.18, 3.41]  |
| Data Set E | 12.69 [11.7, 13.87]                                     | 12.07 [11.24, 13.48]  | 3.63 [3.40, 3.87] |                    |                    |
| Data Set F | 12.30 [9.61, 16.94]                                     | 12.50 [9.29, NA]      | 4.60 [3.71, 6.32] | 9.37 [7.42, 11.93] | 9.37 [7.42, 11.93] |

*Data sets measured median time for real-world extracted endpoints utilizing a common definition as described in the pilot project methods*

**Table 3. One-year real-world overall survival landmark analysis post PD-(L)1 initiation**

| Data Set   | One-year rwOS Landmark Analysis                      |
|------------|--|
| Data Set A | 0.57 [0.52, 0.57] <sup>4</sup>                       |
| Data Set B | 0.54 [0.48, 0.57];<br>0.41 [0.34, 0.47] <sup>5</sup> |
| Data Set C | 0.40 [0.35, 0.46]                                    |
| Data Set D | 0.42 [0.41, 0.43]                                    |
| Data Set E | 0.51 [0.49, 0.53]                                    |
| Data Set F | 0.40 [0.34, 0.48]                                    |

<sup>4</sup> OS was calculated as days between I/O initiation and disenrollment.

<sup>5</sup> Sites with social security or state death data, censored at estimated earliest date such data should be available if no death was observed

**Table 4. Median times and 95% confidence interval (indexed to initial PD-(L)1 inhibitor line start in advanced setting) segmented by treatment setting and demographic characteristics as described in Table 1**

| Demographics   | Data Set A |   | Data Set B |   | Data Set C |                                   | Data Set D |                                   | Data Set E |                                   | Data Set F |                                   |
|--|------------|---|------------|---|------------|-----------------------------------|------------|-----------------------------------|------------|-----------------------------------|------------|-----------------------------------|
|  | N          | rwTTD (Months)<br>Median [95% CI]             | N          | rwTTD (Months)<br>Median [95% CI]             | N          | rwTTD (Months)<br>Median [95% CI] | N          | rwTTD (Months)<br>Median [95% CI] | N          | rwTTD (Months)<br>Median [95% CI] | N          | rwTTD (Months)<br>Median [95% CI] |
|  |            | rwOS (Months)<br>Median [95% CI] <sup>6</sup> | N          | rwOS (Months)<br>Median [95% CI] <sup>7</sup> |            | rwOS (Months)<br>Median [95% CI]  |            | rwOS (Months)<br>Median [95% CI]  |            | rwOS (Months)<br>Median [95% CI]  |            | rwOS (Months)<br>Median [95% CI]  |
| <b>Age categories at PD-(L)1 inhibitor initiation:</b> |            |   |            |   |            |                                   |            |                                   |            |                                   |            |                                   |
| ≤49 years  | 100        | 5.87 [3.97, 9.80]                             | 24         | 3.97 [1.84, 17.03]                            | 21         | 6.44 [1.28, 16.29]                | 219        | 2.89 [2.30, 3.64]                 | 80         | 2.33 [1.43, 4.40]                 | 8          | 8.52 [2.77, NA]                   |
|  |            | 18.1 [11.87, 21.63]                           | 16         | 9.07 [2.66, NA]                               |            | 12.02 [4.27, NA]                  |            | 9.28 [7.77, 12.07]                |            | 10.20 [7.96, 17.36]               |            | NA                                |
| 50-64 years  | 723        | 6.23 [5.20, 7.27]                             | 251        | 3.62 [2.96, 4.14]                             | 129        | 5.35 [3.35, 9.23]                 | 2047       | 3.21 [2.92, 3.44]                 | 863        | 3.53 [3.17, 4.03]                 | 65         | 6.16 [3.71, 12.26]                |
|  |            | 13.60 [11.83,14.6]                            | 163        | 8.78 [6.81, 11.41]                            |            | 9.33 [6.73, 13.27]                |            | 9.34 [8.43, 10.26]                |            | 13.84 [11.80, 15.32]              |            | 16.94 [9.37, NA]                  |
| 65-74 years  | 728        | 7.40 [6.10, 9.60]                             | 198        | 2.76 [2.27, 3.75]                             | 169        | 4.63 [3.45, 6.44]                 | 2504       | 3.41 [3.21, 3.67]                 | 1047       | 3.77 [3.30, 4.23]                 | 94         | 5.38 [3.42, 6.90]                 |
|  |            | 13.40[12.07,14.93]                            | 135        | 8.52 [6.54, 9.70]                             |            | 8.97 [5.78, 11.14]                |            | 9.34 [8.79, 10.26]                |            | 12.16 [10.49, 14.33]              |            | 12.30 [7.55, 21.58]               |
| 75+ years  | 593        | 7.70 [6.37, 9.37]                             | 84         | 2.47 [1.45, 4.64]                             | 116        | 3.57 [1.84, 5.26]                 | 2153       | 3.25 [3.18, 3.61]                 | 870        | 3.77 [3.30, 4.23]                 | 102        | 3.67 [2.82, 5.74]                 |
|  |            | 13.22 [11.83,14.61]                           | 49         | 10.65 [5.03, NA]                              |            | 6.83 [4.24, 9.13]                 |            | 8.79 [8.23, 9.28]                 |            | 13.02 [10.62, 14.79]              |            | 10.00 [8.71, 15.55]               |
| <b>Gender:</b>   |            |   |            |   |            |                                   |            |                                   |            |                                   |            |                                   |
| Female   | 950        | 6.80 [5.90, 8.23]                             | 276        | 3.53 [2.76, 3.98]                             | 212        | 4.76 [3.45, 7.72]                 | 3751       | 2.98 [2.75, 3.21]                 | 1351       | 3.83 [3.53, 4.27]                 | 125        | 5.03 [3.77,7.03]                  |
|  |            | 13.70 [12.8, 15.23]                           | 188        | 8.94 [6.67, 12.33]                            |            | 9.33 [7.42, 13.44]                |            | 8.43 [7.93, 8.98]                 |            | 14.79 [13.15, 16.87]              |            | 13.23 [8.77, 23.33]               |
| Male   | 1194       | 7.23 [6.10, 8.43]                             | 281        | 3.16 [2.37, 3.81]                             | 222        | 4.62 [3.35, 5.52]                 | 3172       | 3.51 [3.21, 3.70]                 | 1509       | 3.30 [2.87, 3.77]                 | 143        | 3.87 [2.93, 6.40]                 |
|  |            | 13.20 [12.13, 14.5]                           | 175        | 7.92 [6.77, 10.06]                            |            | 7.49 [6.34, 10.02]                |            | 9.84 [9.38, 10.72]                |            | 11.08 [10.16, 12.39]              |            | 11.87 [8.40, 20.40]               |
| <b>CLINICAL CHARACTERISTICS</b>                        |            |   |            |   |            |                                   |            |                                   |            |                                   |            |                                   |
| <b>Group stage at initial diagnosis:</b>               |            |   |            |   |            |                                   |            |                                   |            |                                   |            |                                   |
| Stage 0 / 1  |            |   | 23         | 3.45 [0.92, 5.69]                             |            |                                   | 498        | 4.36 [3.67, 5.28]                 |            |                                   | 18         | 5.18 [3.43, 15.65]                |
|  |            |   | 9          | 5.69 [0.49, 9.70]                             |            |                                   |            | 12.07 [10.69, 14.03]              |            |                                   |            | 7.03 [4.87, NA]                   |

<sup>6</sup> rwOS was calculated as time between I/O initiation and disenrollment

<sup>7</sup> rwOS estimates include sites with social security or state death data available; excluded are sites with only local/EHR death data available

|   |     |                      |     |                    |                      |                     |      |                      |     |                     |
|---|-----|----------------------|-----|--------------------|----------------------|---------------------|------|----------------------|-----|---------------------|
| Stage II  | 22  | 3.68 [1.41, 6.38]    |     | 426                | 3.90 [3.28, 4.95]    |                     | 17   | 4.10 [2.80, NA]      |     |                     |
|   | 13  | 6.38 [1.48, NA]      |     |                    | 11.84 [10.59, 13.28] |                     |      | 13.90 [4.03, NA]     |     |                     |
| Stage III   | 88  | 4.14 [2.76, 3.88]    | 39  | 4.43 [1.38, 10.35] | 1494                 | 3.67 [3.44, 4.13]   | 63   | 5.73 [2.90, 9.67]    |     |                     |
|   | 62  | 9.60 [6.67, NA]      |     | 8.97 [2.73, 13.44] |                      | 9.84 [9.18, 10.79]  |      | 14.87 [9.37, 22.10]  |     |                     |
| Stage IV  | 248 | 3.35 [2.76, 3.88]    | 396 | 4.70 [3.68, 5.52]  | 4334                 | 2.89 [2.75, 3.18]   | 161  | 4.32 [3.19, 6.47]    |     |                     |
|   | 187 | 8.77 [6.77, 10.85]   |     | 8.67 [6.83, 10.02] |                      | 8.26 [7.80, 8.79]   |      | 12.10 [8.30, 20.73]  |     |                     |
| Unknown   | 176 | 5.79 [1.87, 3.85]    |     |                    |                      |                     | 10   | 8.83 [3.74, NA]      |     |                     |
|   | 92  | 7.92 [6.15, 10.55]   |     |                    |                      |                     |      | NA                   |     |                     |
| Histology:  |     |                      |     |                    |                      |                     |      |                      |     |                     |
| Non-squamous cell carcinoma                                 | 370 | 3.16 [2.53, 3.75]    | 320 | 4.76 [3.81, 5.81]  | 4678                 | 3.34 [3.21, 3.51]   | 1981 | 3.60 [3.30, 3.90]    | 194 | 4.15 [3.23, 5.97]   |
|   | 240 | 9.69 [7.56, 12.33]   |     | 8.67 [6.7, 10.18]  |                      | 9.61 [9.11, 10.30]  |      | 14.14 [12.69, 15.81] |     | 11.87 [8.77, 21.80] |
| Squamous cell carcinoma                                     | 147 | 3.25 [2.47, 4.01]    | 73  | 4.14 [1.68, 7.33]  | 1983                 | 3.21 [2.98, 3.54]   | 659  | 3.77 [3.30, 4.30]    | 61  | 7.55 [3.67, 12.97]  |
|   | 93  | 6.80 [4.87, 8.78]    |     | 8.38 [4.8, 12.06]  |                      | 8.66 [7.84, 9.25]   |      | 10.36 [9.24, 11.77]  |     | 13.90 [8.57, NA]    |
| NSCLC histology not otherwise specified (NOS)               | 40  | 3.88 [1.94, 5.10]    | 42  | 3.80 [2.07, 9.13]  |                      |                     | 220  | 3.57 [2.83, 5.17]    | 10  | 5.07 [2.29, NA]     |
|   | 30  | 10.26 [5.13, 13.32]  |     | 7.92 [3.58, 17.15] |                      |                     |      | 11.84 [7.89, 16.08]  |     | 20.40 [6.13, NA]    |
| Smoking status:   |     |                      |     |                    |                      |                     |      |                      |     |                     |
| History of smoking  |     |                      | 340 | 4.53 [3.45, 5.35]  | 6185                 | 3.28 [3.21, 3.48]   | 448  | 5.17 [3.90, 6.53]    | 182 | 4.27 [3.43, 6.19]   |
|   |     |                      |     | 8.67 [6.7, 10.18]  |                      | 9.21 [8.85, 9.64]   |      | 19.17 [14.30, 24.46] |     | 11.43 [8.40, 20.73] |
| No history of smoking                                       |     |                      | 94  | 5.52 [3.22, 9.20]  | 716                  | 2.75 [2.49, 3.11]   | 38   | 3.30 [2.37, 9.07]    | 28  | 6.08 [3.23, 8.71]   |
|   |     |                      |     | 8.34 [6.04, 12.88] |                      | 8.69 [8.03, 9.77]   |      | 14.50 [4.57, NA]     |     | 12.10 [8.71, NA]    |
| Unknown/Not documented                                      |     |                      |     |                    |                      |                     | 2374 | 3.53 [3.30, 3.77]    | 59  | 5.97 [2.77, 10.00]  |
|   |     |                      |     |                    |                      |                     |      | 12.00 [10.82, 13.08] |     | 13.21 [9.37, NA]    |
| PD-L1 expression status (among those tested):               |     |                      |     |                    |                      |                     |      |                      |     |                     |
| PD-L1 positive  |     |                      |     |                    | 512                  | 4.10 [3.38, 4.82]   | 45   | 5.63 [2.83, 18.23]   | 65  | 3.68 [2.82, 5.97]   |
|   |     |                      |     |                    |                      | 10.79 [9.05, 13.28] |      | NA                   |     | 9.63 [7.55, NA]     |
| PD-L1 negative/not detected                                 |     |                      |     |                    | 690                  | 2.75 [2.49, 3.11]   | 45   | 9.63 [NA]            | 29  | 6.32 [4.19, NA]     |
|   |     |                      |     |                    |                      | 8.69 [7.48, 9.84]   |      | NA                   |     | 20.80 [11.43, NA]   |
| Line number of first PD-(L)1 inhibitor in advanced setting: |     |                      |     |                    |                      |                     |      |                      |     |                     |
| 1   | 592 | 9.10 [7.97, 12.40]   | 80  | 5.26 [2.63, 6.64]  | 2074                 | 3.90 [3.67, 4.23]   | 777  | 5.90 [4.93, 6.80]    | 77  | 5.03 [3.67, 8.83]   |
|   |     | 19.83 [17.23, 22.23] |     | 9.17 [5.68, 17.05] |                      | 10.36 [9.48, 11.18] |      | 20.78 [14.79, 25.12] |     | 15.87 [9.87, NA]    |

|    |      |  |     |  |      |   |      |  |    |  |
|----|------|--|-----|--|------|---|------|--|----|--|
| 2  | 1174 | 6.57 [6.03, 7.50]<br>11.7 [10.97, 12.83] | 205 | 4.53 [3.22, 5.81]<br>7.75 [5.75, 10.28]  | 3357 | 3.21 [2.82, 3.21]<br>8.66 [8.13, 9.15]  | 1414 | 3.00 [2.83, 3.30]<br>10.68 [9.83, 11.93] | 87 | 4.81 [2.74, 7.00]<br>9.07 [7.39, 20.40]  |
| 3  | 304  | 4.47 [3.8, 6.1]<br>12.8 [10.7, 14.67]    | 85  | 4.63 [2.30, 8.05]<br>9.33 [6.04, 12.02]  | 1011 | 2.98 [2.75, 3.44]<br>9.02 [7.80, 9.97]  | 448  | 3.53 [3.00, 4.23]<br>10.72 [9.04, 13.87] | 51 | 5.67 [3.43, 8.71]<br>15.29 [9.63, NA]    |
| 4+ | 74   | 3.83 [2.83, 5.47]<br>14.2 [10.1, 17.17]  | 65  | 4.76 [3.94, 11.10]<br>8.67 [5.26, 13.27] | 481  | 2.59 [2.30, 3.18]<br>8.52 [6.89, 10.46] | 221  | 3.30 [2.60, 4.23]<br>12.00 [8.25, 15.58] | 54 | 3.47 [2.77, 7.16]<br>10.43 [6.90, 21.80] |

**Table 5. Correlation between real-world overall survival and real-world extracted endpoints using Spearman’s rank correlation coefficient**

| Data Set   | Comparison     | N    | Correlation (95% CI) |
|------------|----------------|------|----------------------|
| Data Set A | rwOS vs rwTTNT | 83   | 0.36                 |
|            | rwOS vs rwTTD  | 254  | 0.63                 |
| Data Set B | rwOS vs rwTTNT |      |                      |
|            | rwOS vs rwTTD  | 225  | 0.62 (0.54, 0.69)    |
| Data Set C | rwOS vs rwTTNT | 96   | 0.70 (0.58, 0.79)    |
|            | rwOS vs rwTTD  | 295  | 0.89 (0.86, 0.91)    |
| Data Set D | rwOS vs rwTTNT | 1203 | 0.61 (0.57, 0.64)    |
|            | rwOS vs rwTTD  | 4337 | 0.80 (0.79, 0.81)    |
|            | rwOS vs rwPFS  | 4337 | 0.75 (0.74, 0.76)    |
|            | rwOS vs rwTTP  | 2286 | 0.60 (0.57, 0.63)    |
| Data Set E | rwOS vs rwTTNT | 358  | 0.62 (0.54, 0.68)    |
|            | rwOS vs rwTTD  | 1456 | 0.77 (0.75, 0.79)    |
| Data Set F | rwOS vs rwTTNT | 39   | 0.46 (0.33, 0.81)    |
|            | rwOS vs rwTTD  | 142  | 0.80 (0.66, 0.85)    |
|            | rwOS vs rwPFS  | 142  | 0.84 (0.62, 0.86)    |
|            | rwOS vs rwTTP  | 55   | 0.56 (0.21, 0.71)    |

*The correlation analysis is restricted to patients with a death date and documented event as described in the definitions and algorithms*

## Conclusions from Pilot Project Study

1. There is a high level of shared characteristics among the varying data sets despite varying sample sizes, data capture processes, and data sources demonstrating the feasibility of identifying aNSCLC patients treated with immune checkpoint inhibitors from diverse RWD sources.
2. The pilot project demonstrated that several extractable endpoints from EHR and claims data correlate with OS. Further validation is required to determine whether these endpoints are reliable surrogates for OS outside of a traditional clinical trial and whether they can support regulatory and payer decision-making.
3. Survival among patients as assessed through EHR and claims data fall within the range of median OS values observed in several immune checkpoint inhibitor trials.<sup>8</sup>
4. Assessment of extracted endpoints from EHR and claims data demonstrate that efficacy of immune checkpoint inhibitors is relatively consistent across a variety of patient characteristics, such as age and sex.

## Assumptions and Limitations of Pilot Project Data Sets

- Ability to collect reliable data will vary across data providers
- Approaches to analysis may vary even when using a common protocol; A careful review and collaboration is needed to align on a consistent and reliable approach
- Verified diagnosis and diagnosis date, clinical stage and cell type, planned chemotherapy regimen (dose and schedule) and other clinical and socioeconomic factors cannot always be determined from the available EHR and claims data
- Verifying and determining date of death may also prove challenging. Although discharge status and some diagnosis codes may be a source of mortality information, but some data partners rely on linkage to the public SSA death master file (DMF). The public DMF has been shown to under identify deaths<sup>9</sup>
- For claims-based data, some patients with advanced disease may enroll in clinical trials and some or all the care received in a clinical trial setting may not generate insurance claims, thus, data for these patients may not be fully captured or captured at all
- Approaches to the analyses may vary even when using a common protocol and careful review and collaboration is needed to align on a consistent and reliable approach
- Some biomarkers may not routinely be assessed in the real-world setting, but more would have been included in this analysis if a chart review had been conducted or the use of natural language processing (NLP)
- Provider data (EHR) may not identify all chemotherapy as patients may seek care inside and outside a provider group that contributes to the EHR data (e.g., chemotherapy at an academic center then move to a community setting). This may or may not be a source of missing information in the advance NSCLC setting

---

<sup>8</sup> Huang G, Sun X, Liu D, et al. The efficacy and safety of anti-PD-1/PD-L1 antibody therapy versus docetaxel for pretreated advanced NSCLC: a meta-analysis. *Oncotarget*, 4239-4248

<sup>9</sup> Jones B, V. D. (2015, March). Measuring Mortality Information in Clinical Data Warehouses. *AMIA Jt Summits Transl Sci Proc*, 450-5

## Discussion Questions

These questions may help guide the discussion during the meeting:


1. Are there processes to handle challenges associated with the availability and consistency of data across provider types and settings?
2. How to overcome difficulties associated with determining events like death?
3. What opportunities or incentives exist to help improve the format, quality, and validity of RWE?
4. Are there lessons from clinical trials, or registration trials, that need to be considered for RW data?
5. What opportunities exist for FDA decision-making to be supported by RWE?
6. What opportunities exist to expand to other endpoints such as patient reported outcomes (PROs) and patient-generated health data?
7. Are there other extractable endpoints for EHR- or claims-based algorithms that should be validated?
8. What is the role and use of real-world endpoints, such as TTD, TTNT, or PFS, for payer decision-making, particularly in the context of accelerated approval or breakthrough therapy designation?
9. How important is RWE in the development of new payment designs, such as value-based payment, risk-sharing arrangements, and outcomes-based agreements?
10. How timely does the data have to be for regulatory or reimbursement? How quickly must the data be analyzed/reported?
11. For reimbursement/value-based payment/risk sharing, are data from all data sets (A-F) available to payers? Manufacturers?

RESEARCH

Open Access



# The Salford Lung Study protocol: a pragmatic, randomised phase III real-world effectiveness trial in chronic obstructive pulmonary disease

Nawar Diar Bakerly<sup>1</sup>, Ashley Woodcock<sup>2</sup>, John P. New<sup>1</sup>, J. Martin Gibson<sup>1</sup>, Wei Wu<sup>3</sup>, David Leather<sup>4</sup> and Jørgen Vestbo<sup>2,5\*</sup> 

## Abstract

**Background:** New treatments need to be evaluated in real-world clinical practice to account for co-morbidities, adherence and polypharmacy.

**Methods:** Patients with chronic obstructive pulmonary disease (COPD),  $\geq 40$  years old, with exacerbation in the previous 3 years are randomised 1:1 to once-daily fluticasone furoate 100  $\mu\text{g}$ /vilanterol 25  $\mu\text{g}$  in a novel dry-powder inhaler versus continuing their existing therapy. The primary endpoint is the mean annual rate of COPD exacerbations; an electronic medical record allows real-time collection and monitoring of endpoint and safety data.

**Conclusions:** The Salford Lung Study is the world's first pragmatic randomised controlled trial of a pre-licensed medication in COPD.

**Trial registration:** Clinicaltrials.gov identifier NCT01551758.

## Introduction

Double-blind randomised controlled trials (RCTs) in chronic obstructive pulmonary disease (COPD) have indicated that inhaled corticosteroids (ICS) combined with a long-acting  $\beta_2$ -agonist (LABA) are more effective than the individual components in managing stable COPD, reducing exacerbations and improving lung function and health status [1]. However, double-blind RCTs differ from real life in having highly selective eligibility criteria, and enrolling participants who are not representative of patients in clinical practice and have much higher adherence [2]. The once-daily combination of the ICS fluticasone furoate (FF) and the novel LABA vilanterol (VI) (Relvar<sup>®</sup>) in a patient-friendly dry-powder inhaler (DPI) (Ellipta<sup>®</sup>) has the potential for improved adherence over the currently available twice-daily

ICS/LABA combinations, with improved clinical effectiveness in a real-world setting [3].

The Salford Lung Study (SLS) is the world's first pragmatic RCT (pRCT) of an investigational medication. SLS will evaluate the effectiveness and safety of the FF/VI combination compared with existing maintenance therapy in a large, real-world population of patients with COPD in conditions of normal care. The study is being conducted in and around Salford, UK. Salford has a high prevalence of COPD in a community served by a single hospital and an established electronic medical record (EMR), connecting both primary and secondary care. Pharmacies also collaborate to allow patients to collect study medication from their usual community pharmacy.

## Methods

### Study design

SLS is a 12-month, open-label, phase III pRCT, evaluating the effectiveness and safety of FF/VI (Relvar<sup>®</sup>; 100  $\mu\text{g}$ /25  $\mu\text{g}$  once daily, delivered by a novel DPI, Ellipta<sup>®</sup>) in patients with COPD (clinicaltrials.gov identifier NCT01551758). The study is conducted in accordance with the International Conference on

\* Correspondence: jorgen.vestbo@manchester.ac.uk

<sup>2</sup>Institute of Inflammation and Repair, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK

<sup>5</sup>Centre for Respiratory Medicine and Allergy, 2nd Floor Education and Research Centre, University Hospital of South Manchester NHS Foundation Trust, Manchester M23 9LT, UK

Full list of author information is available at the end of the article



Harmonisation, Good Clinical Practice (GCP), all applicable data protection requirements and the ethical principles outlined in the Declaration of Helsinki 2008. The study was approved by the Ethics committee, National Research Ethics Service Committee North West, Greater Manchester South.

**Patients**

All patients with COPD at 66 primary care sites (at the time of manuscript preparation) in and around Salford and South Manchester are identified by their general practitioner (GP) from practice databases and invited to participate in the study.

Eligibility criteria include:

- aged ≥40 years
- documented GP diagnosis of COPD
- regular maintenance inhaler therapy (ICS alone or in combination with long-acting bronchodilator, one or more long-acting bronchodilators, or triple therapy [ICS/LABA plus long-acting muscarinic antagonist])
- at least one COPD exacerbation in the last 3 years.

Minimal exclusion criteria:

- an exacerbation within the previous 2 weeks
- chronic oral corticosteroid use.

At visit 1, patients are offered study participation through written informed consent (Fig. 1). At visit 2 (1–60 days after visit 1), patients are randomised (1:1) to receive either FF/VI or to continue their usual maintenance therapy. Patients randomised to FF/VI are instructed in the use of the Ellipta DPI. Patients randomised to their usual maintenance therapy are re-trained in the correct techniques and dosing. Baseline assessments

are performed at visit 2, including quality of life and disease characteristics (e.g., disease duration, COPD maintenance therapy, smoking status, lung function, medical history). If at months 3, 6 and 9 the patient has had no contact with their GP practice within the previous 8 weeks, the patient is contacted by telephone to assess any serious adverse events (SAEs) or non-serious adverse drug reactions (ADRs) (visits 3, 4 and 5). There is no additional intervention at these assessments. At 12 months, the final visit (6) is a face-to-face meeting with the patient at which the final assessment of outcomes is conducted.

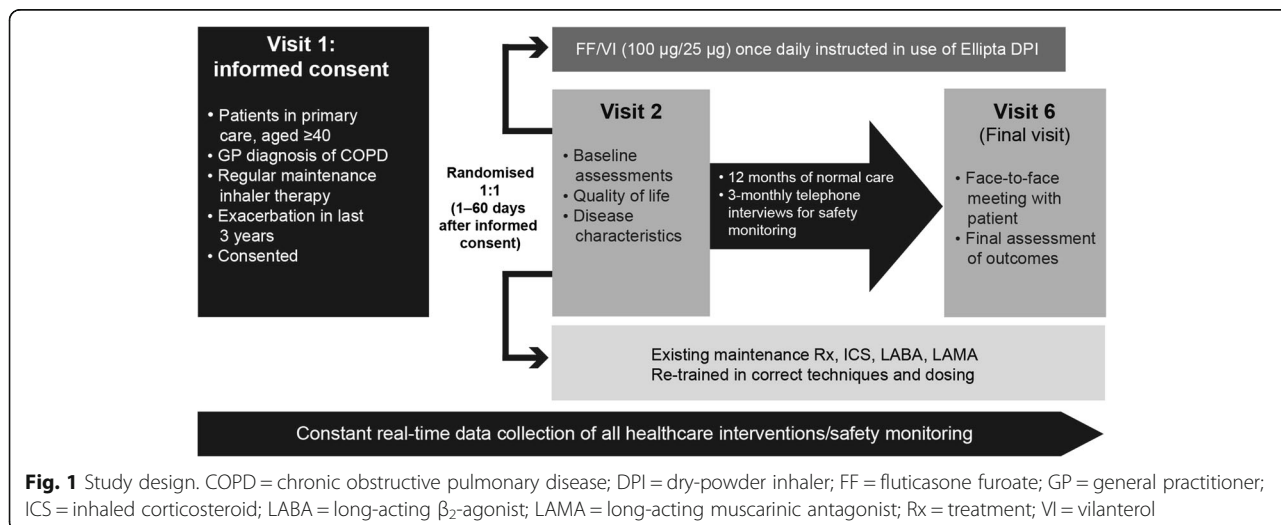
**Participating sites**

**Primary care**

To preserve the real-world nature of the study, the patient experience is as close to routine care as possible. The study’s principal investigators are the GPs. They are ideally placed to facilitate recruitment, identify and report SAEs or serious ADRs and report study endpoints. GPs may make treatment adjustments according to their clinical opinion. Repeat prescriptions of study medication are issued by GPs as usual, and collected by patients from their usual pharmacy. As very few participating GPs had experience of clinical trial participation, all GPs have received training and support in GCP, patient recruitment, study protocol, coding of healthcare issues and general research procedures.

**Pharmacy**

Every pharmacy in Salford and others in South Manchester agreed to participate in the study. As with GPs, very few pharmacists had experience of clinical trial participation. All staff (>500) at participating pharmacies have received training in GCP and safety reporting and standard operating procedures were established. Initially, pharmacies faxed copies of study treatment prescriptions



**Fig. 1** Study design. COPD = chronic obstructive pulmonary disease; DPI = dry-powder inhaler; FF = fluticasone furoate; GP = general practitioner; ICS = inhaled corticosteroid; LABA = long-acting β<sub>2</sub>-agonist; LAMA = long-acting muscarinic antagonist; Rx = treatment; VI = vilanterol

to the study coordination centre, but these are now collected electronically. Prescription collection data are used to assess treatment adherence.

**Hospital**

The majority of admissions are to the local hospitals: Salford Royal Hospital and University Hospital of South Manchester. Admissions are identified electronically and assessed by a separate secondary care team within 48 h.

**Data monitoring**

All hospital admissions, outpatient and emergency department visits are identified from the EMR database (whenever and wherever they occur). From primary care, all healthcare contacts, out-of-hours activity and prescriptions of antibiotics or oral steroids can be identified. These events are reviewed by the study research team and classified as COPD or non-COPD related. Furthermore, the EMRs capture suspected unexpected serious adverse reactions (e.g., reduced kidney function or elevated liver function tests) and, for the purposes of SLS, include data from external sources to identify, e.g., deaths or National Health Service (NHS) hospital admissions outside Salford. NorthWest EHealth ([www.nweh.org.uk](http://www.nweh.org.uk)) manages the EMRs, enabling data on study endpoints and patient safety to be collected continuously and remotely in near-real time, without the need for face-to-face patient contact.

**Endpoints**

**Effectiveness**

The primary endpoint is the mean annual rate of moderate or severe exacerbations. Secondary endpoints include time to first exacerbation and healthcare utilisation. Other endpoints include hospitalisations, use of rescue medication, the COPD Assessment Test (CAT) [4] and EuroQoL-5 dimensions (EQ-5D) questionnaire, listed and defined in Table 1. EMR data for effectiveness endpoints are independently verified by the research team (GP, research nurse, research doctor).

**Safety**

Safety endpoints include death, pneumonia, frequency and type of SAEs, and ADRs. Investigators and site staff are responsible for detecting, documenting and reporting SAEs and ADRs on electronic case report forms (eCRFs), which are continuously monitored by a dedicated clinical safety team.

**Withdrawals**

Patients with COPD exacerbation during the treatment period may remain in the study and continue to take study medication at the discretion of their GP. Severe COPD exacerbations are reported as SAEs. Patients with

**Table 1** Study endpoints

| Endpoint  | Definitions   |
|---|---|
| <b>Primary endpoint</b>   |   |
| Mean annual rate of moderate or severe exacerbations  | <ul style="list-style-type: none"> <li>Moderate exacerbation: patient receiving an exacerbation-related prescription of oral corticosteroids and/or antibiotic (with or without NHS contact) not requiring hospitalisation</li> <li>Severe exacerbation: an exacerbation-related hospitalisation</li> </ul> |
| <b>Secondary endpoints</b>  |   |
| <ul style="list-style-type: none"> <li>COPD-related secondary care contacts</li> <li>COPD-related primary care contacts</li> </ul>  | <ul style="list-style-type: none"> <li>Respiratory-related contacts: contact where the most prominent signs and symptoms with which the patient presents are as a direct result of the patient's COPD</li> </ul>  |
| <ul style="list-style-type: none"> <li>All secondary care contacts</li> <li>All primary care contacts</li> </ul>  | <ul style="list-style-type: none"> <li>All contacts: any interaction between the patient and a doctor or nurse working as part of the NHS (including telephone calls), not including protocol-defined study-related visits</li> </ul>   |
| <ul style="list-style-type: none"> <li>Time to discontinuation of initial therapy</li> <li>Time to addition of a further COPD controller medication</li> <li>Time to first moderate/severe exacerbation</li> <li>Time to first severe exacerbation (i.e., hospitalisation)</li> </ul>   |   |
| <b>Other endpoints</b>  |   |
| <ul style="list-style-type: none"> <li>Number of hospitalisations</li> <li>Number of days in hospital</li> <li>Total number of respiratory-related home visits (including out-of-hours calls) and telephone consultations</li> <li>CAT: disease management, quality of life</li> <li>EQ-5D</li> <li>Adherence to study medication</li> <li>Number of salbutamol inhalers collected by the patients from study-enrolled community pharmacies over the 12-month treatment period</li> </ul> | <ul style="list-style-type: none"> <li>Adherence is assessed based on analysis of medications (prescribed, dispensed and collected) and use of the MARS-A at visit 2 and visit 6/early withdrawal visit</li> </ul>  |

CAT COPD Assessment Test, COPD chronic obstructive pulmonary disease, EQ-5D EuroQoL Questionnaire, MARS-A Medication Adherence Report Scale for Asthma, NHS National Health Service

worsening COPD status while on study treatment may have their medication adjusted at the GP's discretion or receive other permitted COPD therapies. If, in the investigator's opinion, the frequency or severity of exacerbations prevents ongoing participation, the patient can be withdrawn. The reason for withdrawal is recorded in the eCRF and patients are followed up for 12 months following randomisation, with their consent.

### Statistical analysis and rationale

The primary efficacy analysis population is intent-to-treat, defined as all randomised patients who have received at least one prescription of study medication.

Sample size calculations are based on the primary endpoint (mean annual rate of moderate and severe exacerbations) and primary efficacy analysis population. A total of 2238 patients (1119 patients per treatment group) are needed. The study has 80 % power to detect a relative reduction of 12 % in the mean annual moderate or severe exacerbation rate, assuming a mean exacerbation rate of 2.3 for the control group [5]. Calculations are based on a negative binomial regression with a dispersion rate of 0.7 and use a two-sided 5 % significance level.

To account for variation in treatment response between patient subgroups, randomisation is stratified by baseline maintenance therapy and by history of COPD exacerbation in the previous 12 months (yes/no) to ensure treatment balance for the primary efficacy analysis population. Analyses based on subgroups are also planned; subgroups will be defined on baseline medication, lung function, comorbidities and other factors.

### Discussion

SLS is a unique pRCT and, to our knowledge, the first prospective real-world comparative effectiveness study of an investigational medicine, which commenced in March 2012, prior to UK regulatory approval (launch date January 2014). The pragmatic inclusion criteria in SLS represent the broad definition of a patient eligible for COPD maintenance therapy in the real world, irrespective of co-morbidities. Study accessibility is maximised by employing minimal exclusion criteria and requirements for additional GP visits. Medicine prescription and supply is achieved as usual, through the patient's own GP and pharmacy.

Real-world outcomes can be assessed by observational studies that provide high external validity but in contrast have low internal validity [6]. With the limitations in observational studies and those in double-blind RCTs [2] such studies alone may not fully reflect the true impact and value of treatments for COPD. As such, well-designed pRCTs may offer complementary data to these standard types of studies, representing true real-world effectiveness.

Performing a study of a pre-licence drug in a real-world setting has posed many new challenges in study design, operational planning and study support. Patient safety is a priority in studying a pre-licence medicine. Patient safety in SLS is monitored in almost real-time by a combination of remote surveillance of EMRs and clinical monitoring. This sets a new standard, in which safety signals can be seen more quickly than in

conventional RCTs. The major challenge has been managing and assessing the relevance and importance of safety signals within the huge volume of healthcare data being generated.

The SLS has limitations. Patients are not blinded and although patients are far less selected than in a usual efficacy trial, some selection bias cannot be precluded. Also, the fact that patients are recruited on the basis of a diagnosis from an electronic medical record and not from a specialist clinic could raise concerns. However, our approach mirrors the real world and a recent study found that registered data had satisfactory validity [7].

### Conclusions

SLS is an innovative project with the aim of evaluating the safety and effectiveness of an investigational medicine in a real-world setting. Data from SLS will allow a better understanding of the risk/benefit profile of the FF/VI combination in the wider COPD community. The study will likely be a role model for future evaluation of effectiveness of new therapies.

### Abbreviations

ADR: Adverse drug reaction; CAT: COPD Assessment Test; COPD: Chronic obstructive pulmonary disease; DPI: Dry-powder inhaler; eCRF: Electronic case report form; EMR: Electronic medical record; EQ-5D: EuroQoL-5 dimensions (EQ-5D) questionnaire; FF: Fluticasone furoate; GCP: Good Clinical Practice; GP: General practitioner; ICS: Inhaled corticosteroid; LABA: Long-acting  $\beta_2$ -agonist; LAMA: Long-acting muscarinic antagonist; MARS-A: Medication Adherence Report Scale for Asthma; NHS: National Health Service; pRCT: Pragmatic randomised controlled trial; RCT: Randomised controlled trial; Rx: Treatment; SAE: Serious adverse event; SLS: Salford Lung Study; VI: Vilanterol.

### Competing interests

NDB's employing organisation provides IT support to GlaxoSmithKline. He has received educational grants and speaker's fees from GlaxoSmithKline and Novartis, and support for attending educational conferences from Boehringer Ingelheim, GlaxoSmithKline and Novartis. AW has acted on advisory boards and provided consultancy for Ammiral, Chiesi, Cytos and GlaxoSmithKline. He has received travel support to speak at international meetings from Boehringer Ingelheim and GlaxoSmithKline. He is an investigator on cough and asthma studies for Afferent and GlaxoSmithKline. JPN has received consulting and speaker's fees, and an educational grant from GlaxoSmithKline. JMG's institution has received funding from GlaxoSmithKline as the SLS study sponsor. WW is an employee of, and holds shares/stock options in, GlaxoSmithKline. DL is an employee of, and holds shares/stock options in, GlaxoSmithKline. JV has received travel support and consultancy fees from GlaxoSmithKline (related to the SLS study); in addition, he has received consultancy fees from Ammiral, AstraZeneca, Bioxydyn, Boehringer-Ingelheim, Chiesi, GlaxoSmithKline (outside the SLS study), Novartis, Syntaxin and Takeda (Nycomed), and speaker's fees from AstraZeneca, Boehringer Ingelheim, Chiesi, GlaxoSmithKline, Novartis and Takeda (Nycomed). His wife has previously worked for AstraZeneca, Ferring and GlaxoSmithKline (until 2009).

### Authors' contributions

All authors are involved in the design and implementation of the Salford Lung Study and contributed equally to the preparation of this paper, including development of the outline, review of all drafts, final approval and decision to submit the manuscript to *Respiratory Research*.

### Acknowledgements

The authors thank Kerry Acheson of iMed Comms, Macclesfield, UK, who provided medical writing support, which was funded by GlaxoSmithKline.

### Funding support

GlaxoSmithKline are sponsors of the Salford Lung Study and provided scientific support for the study design, protocol writing, data collection, analysis, interpretation of study data, and funded medical writing services.

### Author details

<sup>1</sup>Salford Royal NHS Foundation Trust, Salford, UK. <sup>2</sup>Institute of Inflammation and Repair, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK. <sup>3</sup>GlaxoSmithKline, Research Triangle Park, Durham, NC, USA. <sup>4</sup>GSK Respiratory Centre of Excellence, GlaxoSmithKline UK Ltd, Uxbridge, UK. <sup>5</sup>Centre for Respiratory Medicine and Allergy, 2nd Floor Education and Research Centre, University Hospital of South Manchester NHS Foundation Trust, Manchester M23 9LT, UK.

Received: 2 July 2015 Accepted: 26 August 2015

Published online: 04 September 2015

### References

1. Calverley PM, Rennard SI. What have we learned from large drug treatment trials in COPD? *Lancet*. 2007;370:774–85.
2. Roche N, Reddel H, Martin R, Brusselle G, Papi A, Thomas M, et al. Quality standards for real-world research. Focus on observational database studies of comparative effectiveness. *Ann Am Thorac Soc*. 2014;11 Suppl 2:S99–104.
3. Claxton AJ, Cramer J, Pierce C. A systematic review of the associations between dose regimens and medication compliance. *Clin Ther*. 2001;23:1296–310.
4. Jones P, Harding G, Wiklund I, Berry P, Leidy N. Improving the process and outcome of care in COPD: development of a standardised assessment tool. *Prim Care Respir J*. 2009;18:208–15.
5. New JP, Delderfield MR, Stein ND, Austin S, Vestbo J, Woodcock AA. Assessing the burden of asthma and COPD in Salford UK: retrospective analysis using a whole population electronic medical record. *Eur Respir J*. 2011;38 Suppl 55:732S. Abstract 4012.
6. Price D, Chisholm A, van der Molen T, Roche N, Hillyer EV, Bousquet J. Reassessing the evidence hierarchy in asthma: evaluating comparative effectiveness. *Curr Allergy Asthma Rep*. 2011;11:526–38.
7. Elkhenini HF, Davis KJ, Stein ND, New JP, Delderfield MR, Gibson M, et al. Using an electronic medical record (EMR) to conduct clinical trials: Salford Lung Study feasibility. *BMC Med Inform Dec Making*. 2015;15:8. doi:10.1186/s12911-015-0132-z.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)



## ORIGINAL ARTICLE

# Effectiveness of Fluticasone Furoate–Vilanterol for COPD in Clinical Practice

Jørgen Vestbo, D.M.Sc., David Leather, M.B., Ch.B., Nawar Diar Bakerly, M.D., John New, M.B., B.S., J. Martin Gibson, Ph.D., Sheila McCorkindale, M.B., Ch.B., Susan Collier, M.B., Ch.B., Jodie Crawford, M.Sc., Lucy Frith, M.Sc., Catherine Harvey, D.Phil., Henrik Svedsater, Ph.D., and Ashley Woodcock, M.D., for the Salford Lung Study Investigators\*

## ABSTRACT

**BACKGROUND**

Evidence for the management of chronic obstructive pulmonary disease (COPD) comes from closely monitored efficacy trials involving groups of patients who were selected on the basis of restricted entry criteria. There is a need for randomized trials to be conducted in conditions that are closer to usual clinical practice.

**METHODS**

In a controlled effectiveness trial conducted in 75 general practices, we randomly assigned 2799 patients with COPD to a once-daily inhaled combination of fluticasone furoate at a dose of 100  $\mu\text{g}$  and vilanterol at a dose of 25  $\mu\text{g}$  (the fluticasone furoate–vilanterol group) or to usual care (the usual-care group). The primary outcome was the rate of moderate or severe exacerbations among patients who had had an exacerbation within 1 year before the trial. Secondary outcomes were the rates of primary care contact (contact with a general practitioner, nurse, or other health care professional) and secondary care contact (inpatient admission, outpatient visit with a specialist, or visit to the emergency department), modification of the initial trial treatment for COPD, and the rate of exacerbations among patients who had had an exacerbation within 3 years before the trial, as assessed in a time-to-event analysis.

**RESULTS**

The rate of moderate or severe exacerbations was significantly lower, by 8.4% (95% confidence interval, 1.1 to 15.2), with fluticasone furoate–vilanterol therapy than with usual care ( $P=0.02$ ). There was no significant difference in the annual rate of COPD-related contacts to primary or secondary care. There were no significant between-group differences in the rates of the first moderate or severe exacerbation and the first severe exacerbation in the time-to-event analyses. There were no excess serious adverse events of pneumonia in the fluticasone furoate–vilanterol group. The numbers of other serious adverse events were similar in the two groups.

**CONCLUSIONS**

In patients with COPD and a history of exacerbations, a once-daily treatment regimen of combined fluticasone furoate and vilanterol was associated with a lower rate of exacerbations than usual care, without a greater risk of serious adverse events. (Funded by GlaxoSmithKline; Salford Lung Study ClinicalTrials.gov number, NCT01551758.)

From the Centre for Respiratory Medicine and Allergy, Manchester Academic Health Sciences Centre, University of Manchester and University Hospital of South Manchester NHS Foundation Trust (J.V., A.W.), Manchester Academic Health Sciences Centre, University of Manchester and Salford Royal NHS Foundation Trust (J.M.G.), and NIHR Clinical Research Network Greater Manchester (S.M.), Manchester; Global Respiratory Franchise (D.L.) and Respiratory Research and Development (S.C., J.C., L.F., H.S.), GlaxoSmithKline UK, Brentford; Salford Royal NHS Foundation Trust (N.D.B., J.N., J.M.G.), NorthWest EHealth (J.N., J.M.G.), and NHS Salford Clinical Commissioning Group (S.M.), Salford; and Global Clinical Safety and Pharmacovigilance, Safety Evaluation and Risk Management, GlaxoSmithKline UK, Uxbridge (C.H.) — all in the United Kingdom. Address reprint requests to Dr. Vestbo at the Centre for Respiratory Medicine and Allergy, 2nd Fl., Education and Research Centre, University Hospital of South Manchester, Southmoor Rd., Manchester M23 9LT, United Kingdom, or at jorgen.vestbo@manchester.ac.uk.

\*A complete list of the investigators in the Salford Lung Study is provided in the Supplementary Appendix, available at NEJM.org.

This article was published on September 4, 2016, at NEJM.org.

N Engl J Med 2016;375:1253-60.

DOI: 10.1056/NEJMoa1608033

Copyright © 2016 Massachusetts Medical Society.



**G**UIDELINES ON THE MANAGEMENT OF chronic obstructive pulmonary disease (COPD) are based on numerous randomized, controlled trials of efficacy, which are usually generated for registration purposes.<sup>1</sup> However, these trials have included patients who were selected with the use of strict criteria and were closely monitored, and therefore the results have limited relevance to everyday clinical practice.<sup>2</sup> To counter this, it has been proposed that integrated comparative effectiveness trials involve more representative patients and be conducted in much less restricted environments.<sup>3-5</sup>

The Salford Lung Study was designed to evaluate the effectiveness and safety of the once-daily inhaled combination of fluticasone furoate and vilanterol (fluticasone furoate–vilanterol) as compared with existing maintenance therapy (usual care) in a large, real-world population of patients with COPD in conditions of normal care. The trial was initiated before the approval of fluticasone furoate–vilanterol in the United Kingdom and was conducted in and around Salford, United Kingdom, a community served mainly by a single hospital with an established electronic health record (EHR) system that connects primary and secondary care. This setting permits the unobtrusive observation of patients for effectiveness and safety monitoring, blended into routine clinical care.<sup>6</sup>

## METHODS

### PATIENTS

Between March 13, 2012, and October 23, 2014, we recruited patients who were 40 years of age or older, had received a documented diagnosis of COPD from a general practitioner, and had had one or more COPD exacerbations in the previous 3 years. Patients had to be taking regular maintenance inhaler therapy, defined as the use of one or more long-acting bronchodilators; inhaled glucocorticoids, alone or in combination with a long-acting bronchodilator; or a combination of inhaled glucocorticoids, a long-acting beta-agonist (LABA), and a long-acting muscarinic antagonist (LAMA). There were no restrictions regarding smoking history or spirometric values. Among the few exclusion criteria were an exacerbation within the previous 2 weeks and long-term use of oral glucocorticoids. Details of the trial design and the analysis approach have been published previously.<sup>7,8</sup>

Patients were recruited in primary care practices by the health care professionals who provided their normal, everyday care. All the patients provided written informed consent. The trial was conducted in accordance with the International Conference on Harmonisation Good Clinical Practice guidelines and the provisions of the 2008 Declaration of Helsinki. The trial protocol was approved by the National Research Ethics Service Committee North West, Greater Manchester South. The protocol, including the statistical analysis plan, is available with the full text of this article at NEJM.org.

### TRIAL DESIGN

This prospective, 12-month, open-label, parallel-group, randomized trial was conducted in 75 general practices in Salford and South Manchester, United Kingdom. Randomization was performed by means of a centralized randomization service, with stratification according to baseline maintenance therapy and presence or absence of a COPD exacerbation in the previous 12 months. Participants were assigned, in a 1:1 ratio, to receive one of two treatments: combination therapy with 100  $\mu\text{g}$  of fluticasone furoate and 25  $\mu\text{g}$  of vilanterol (Relvar [in Europe] or Breo [in the United States], GlaxoSmithKline), administered once daily as a dry powder through an inhaler (Ellipta, GlaxoSmithKline) (the fluticasone furoate–vilanterol group); or the continuation of usual care as determined by the general practitioner (the usual-care group). Patients who were randomly assigned to fluticasone furoate–vilanterol and had been previously treated with two long-acting bronchodilators and an inhaled glucocorticoid were allowed to continue taking a LAMA in addition to fluticasone furoate–vilanterol.

At the first trial visit, patients were offered participation and provided written informed consent. Within 1 to 60 days after the first visit, patients were randomly assigned to receive fluticasone furoate–vilanterol or to continue their usual maintenance therapy. (The 2 full months was the result of being able to use planned appointments in order to make the trial as close to normal practice as possible.) Trial staff trained the patients in each treatment group in the correct inhaler techniques and dosing, obtained baseline information on disease duration, smoking status, lung function, and concomitant medical history, and performed baseline assessments

of COPD symptoms with the use of the COPD Assessment Test (CAT)<sup>9</sup> and of quality of life with the use of the European Quality of Life-5 Dimensions (EQ-5D) questionnaire.<sup>10</sup> Spirometric findings were evaluated according to the Global Initiative for Chronic Obstructive Lung Disease (GOLD), with airflow limitation present when the ratio of the forced expiratory volume in 1 second (FEV<sub>1</sub>) to forced vital capacity was less than 0.7. Severity was graded according to the level of FEV<sub>1</sub>.

If at months 3, 6, and 9 patients had had no contact with their general practice within the previous 8 weeks, they were contacted by telephone by a trial team member to assess any serious adverse events or nonserious adverse drug reactions; there was no additional intervention at these assessments. At 12 months, trial staff met the patients to make a final assessment of outcomes. Thus, most patients had contact with trial staff only at recruitment, at the baseline visit, and at 12 months.

To preserve the real-world nature of the trial, the patients' experience was as close to everyday clinical practice care as possible. The key investigators in the trial were the general practitioners, who could choose the appropriate therapy according to their clinical opinion, and treatments were dispensed by community pharmacies in the usual way. Patients could switch from fluticasone furoate-vilanterol to usual care, but patients in the usual-care group were not permitted to switch to the fluticasone furoate-vilanterol group. All the general practitioners and pharmacy staff received training regarding Good Clinical Practice guidelines as well as training in trial procedures and trial medications as appropriate to their roles.<sup>8</sup>

#### OUTCOME MEASUREMENTS

The primary outcome was the mean annual rate of moderate or severe exacerbations, defined as any worsening of respiratory symptoms that led to treatment with antibiotic agents or systemic glucocorticoids (or both), to hospital admission, or to scheduled or unscheduled hospital visits. The primary outcome was assessed in the primary effectiveness analysis population, which was a subgroup of the entire trial population that included patients who had undergone randomization, received a prescription of the trial medication (e.g., fluticasone furoate-vilanterol or, in the usual-care group, a COPD-controller medica-

tion), and had had one or more exacerbations in the preceding year. All the secondary outcomes were analyzed in the entire trial population (i.e., all the patients who underwent randomization and received a prescription of the trial medication) and included the rate of first exacerbation, as assessed in a time-to-event analysis, and the annual rates of primary and secondary health care contacts. Other outcomes included the CAT score and the EQ-5D questionnaire results. Except for exacerbations, modification to trial medication, CAT score, EQ-5D questionnaire, and demographic characteristics, data were collected in real time with the use of an integrated primary and secondary care EHR that was developed by NorthWest EHealth (NWEH). EHR data for the primary outcome were independently verified by the research team (general practitioner, research nurse, or research doctor).

#### SAFETY EVALUATION

Safety outcomes included serious adverse events of pneumonia (defined as pneumonia, which was prespecified as an adverse event of special interest), the frequency and type of other serious adverse events, and adverse drug reactions. Adverse events of special interest were defined a priori as groups of events of interest that were considered to be possibly related to inhaled glucocorticoids or LABAs. Safety monitoring was performed by means of continuous real-time monitoring of the patients' EHRs with the use of the linked NWEH database system and by means of telephone contact every 3 months (unless another contact occurred). Investigators reported serious adverse events and adverse drug reactions on electronic case-report forms, which were continuously monitored by near-real-time data monitoring and a dedicated clinical safety team. Cause of death was not adjudicated but was assigned by the primary investigator for all fatal events.

#### TRIAL OVERSIGHT

The Salford Lung Study team sought scientific advice by means of a joint consultation process with the National Institute for Health and Care Excellence and the Medicines and Healthcare Products Regulatory Agency. Informal advice was sought from the National Research Ethics Service Committee North West, Greater Manchester South, United Kingdom, before the formal ethics application.

The trial was designed by the sponsor and the academic partners. The sponsor and NWEH collected the data. Statistical analyses were performed by a contract research organization on behalf of, and with oversight by, employees of the sponsor. All the authors had full access to the data and vouch for the accuracy and completeness of all the data and analyses and for the fidelity of the trial to the protocol. The first draft of the manuscript was written jointly by the primary academic and senior authors, and all the authors worked collaboratively to prepare the final content and made the decision to submit the manuscript for publication. Editorial support was provided by a medical writer, paid by the sponsor.

#### STATISTICAL ANALYSIS

Sample-size calculations were based on the primary outcome (mean annual rate of moderate or severe exacerbations). We calculated that 2238 patients would need to be enrolled for the trial to have 80% power to detect a relative change of 12% in the mean annual rate of moderate or severe exacerbations between the fluticasone furoate–vilanterol group and the usual-care group, assuming a mean rate of 2.3 exacerbations in the usual-care group, as estimated on the basis of a retrospective analysis of historical data from patients from the Salford area who underwent randomization at the time of the protocol amendment<sup>11</sup> that were collected from the linked NWEH database. Calculations were based on a negative binomial regression, with a dispersion rate of 0.7, and used a two-sided 5% significance level. All the analyses were conducted according to the intention-to-treat principle (see the Supplementary Appendix, available at NEJM.org).

## RESULTS

#### TRIAL POPULATION

Of 3161 patients with COPD who were screened, 2802 underwent randomization (see the Supplementary Appendix). Three patients in the fluticasone furoate–vilanterol group never took the trial medication, so the overall trial population consisted of 2799 patients. Of these, 2269 patients (81%) had one or more moderate or severe exacerbations in the year before the trial and made up the primary effectiveness analysis population (Table 1). In the overall trial popula-

tion, 1291 patients in the fluticasone furoate–vilanterol group and 1309 in the usual-care group completed the trial; in the primary effectiveness analysis population, 1051 patients in the fluticasone furoate–vilanterol group and 1056 in the usual-care group completed the trial.

In the primary effectiveness analysis population, 276 patients (12%) were taking a LABA, a LAMA, or both (35 patients were taking both) at the time of randomization. A total of 762 patients (34%) were receiving inhaled glucocorticoids, a combination of inhaled glucocorticoids and a LABA, or a combination of inhaled glucocorticoids and a LAMA; 119 of these patients were using inhaled glucocorticoids as monotherapy. A total of 1231 patients (54%) were receiving combination triple therapy with inhaled glucocorticoids, a LABA, and a LAMA.

Overall, 47% of the patients reported having had two or more moderate COPD exacerbations in the year before entry; 7% reported having had one or more severe exacerbations. A total of 22% of the patients had a diagnosis of asthma recorded. More than three quarters of the patients (77%) had coexisting conditions (Table 1).

In the fluticasone furoate–vilanterol group, 342 patients (24%) modified their medication regimen; 302 of these patients (22%) switched back to their previous care, of whom 54 (4%) switched back because of a need for better control. In the usual-care group, 160 patients (11%) modified their medication regimen, including 114 (8%) who had a need for better control.

#### PRIMARY OUTCOME

In the primary effectiveness analysis population, the rate of moderate or severe exacerbations was 1.74 exacerbations per year in the fluticasone furoate–vilanterol group, as compared with 1.90 per year in the usual-care group, indicating an 8.4% (95% confidence interval [CI], 1.1 to 15.2) lower rate in the fluticasone furoate–vilanterol group ( $P=0.02$ ) (Fig. 1A). This finding was confirmed in the entire trial population, in which the rate of moderate or severe exacerbations was 1.50 exacerbations per year in the fluticasone furoate–vilanterol group, as compared with 1.64 per year in the usual-care group, indicating an 8.4% (95% CI, 1.4 to 14.9) lower rate in the fluticasone furoate–vilanterol group ( $P=0.02$ ). In patients with COPD of GOLD grade 1 or 2 at baseline (GOLD grade 1, indicating mild disease, is



**Table 1. Characteristics of the Participants at Baseline.\***

| Characteristic   | Entire Trial Population (N = 2799) | Usual Care (N = 1403) | Fluticasone Furoate-Vilanterol (N = 1396) | Primary Effectiveness Analysis Population (N = 2269) |
|--|------------------------------------|-----------------------|---|--|
| Age — yr   | 67±10                              | 67±10                 | 67±10                                     | 67±10  |
| Female sex — no. (%)                                       | 1369 (49)                          | 671 (48)              | 698 (50)                                  | 1122 (49)  |
| Body-mass index†   | 28±6                               | 28±6                  | 28±7                                      | 28±6   |
| Current smoking — no. (%)                                  | 1289 (46)                          | 666 (47)              | 623 (45)                                  | 1046 (46)  |
| Postbronchodilator FEV <sub>1</sub> — liters               | 1.62±0.64                          | 1.62±0.65             | 1.62±0.64                                 | 1.59±0.64  |
| No. of exacerbations during the 12 mo before randomization | 2.01±1.99                          | 2.04±2.08             | 1.98±1.90                                 | 2.48±1.93  |
| Coexisting condition — no. (%)                             |                                    |                       |   |  |
| Any  | 2145 (77)                          | 1076 (77)             | 1069 (77)                                 | 1758 (77)  |
| Cardiac condition  | 720 (26)                           | 367 (26)              | 353 (25)                                  | 588 (26)   |
| Vascular condition   | 1363 (49)                          | 675 (48)              | 688 (49)                                  | 1095 (48)  |
| Asthma   | 609 (22)                           | 293 (21)              | 316 (23)                                  | 512 (23)   |
| Diabetes   | 438 (16)                           | 208 (15)              | 230 (16)                                  | 353 (16)   |

\* Plus-minus values are means ±SD. There were no significant differences between the treatment groups in any of the baseline characteristics. The primary effectiveness analysis population was a subgroup of the entire trial population that included patients who had undergone randomization, received a prescription of the trial medication, and had had one or more exacerbations in the preceding year. Additional details on the baseline characteristics are provided in Table S1 in the Supplementary Appendix. FEV<sub>1</sub> denotes forced expiratory volume in 1 second.

† The body-mass index is the weight in kilograms divided by the square of the height in meters.

defined as an FEV<sub>1</sub> ≥80% of the predicted value, and GOLD grade 2, indicating moderate disease, as an FEV<sub>1</sub> ≥50% and <80% of the predicted value, both in the presence of a ratio of FEV<sub>1</sub> to forced vital capacity of <0.7), the rate of exacerbations was 1.50 exacerbations per year in the fluticasone furoate-vilanterol group, as compared with 1.71 per year in the usual-care group, indicating a 12.1% (95% CI, 1.0 to 21.9) lower rate in the fluticasone furoate-vilanterol group.

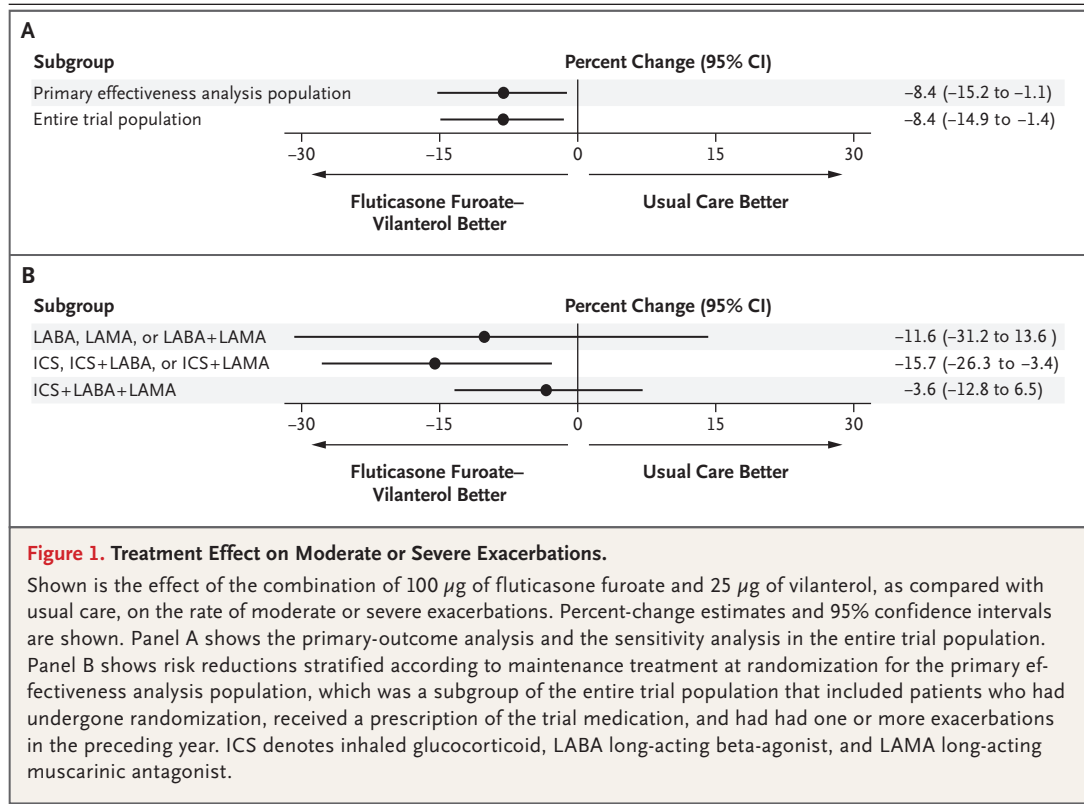
Fig. 1B shows the percent change in the rate of moderate or severe exacerbations between the groups in the primary effectiveness analysis population, stratified according to prerandomization treatment; the interaction of treatment with strata was not significant (P=0.29). The treatment difference was significant among patients in the primary effectiveness population whose randomization stratum and prerandomization treatment included an inhaled glucocorticoid and a LABA (1.87 exacerbations per year among 927 patients in the fluticasone furoate-vilanterol group vs. 2.03 exacerbations per year among 908 patients in the usual-care group), with

an 8.0% (95% CI, 0.11 to 15.4) lower rate in the fluticasone furoate-vilanterol group (P=0.047).

#### SECONDARY OUTCOMES

There was no significant difference in the rate of first moderate or severe exacerbation in the time-to-event analysis in the entire trial population (hazard ratio with fluticasone furoate-vilanterol vs. usual care, 0.93; 95% CI, 0.85 to 1.02). Similarly, there was no significant difference in the rate of severe exacerbations between the fluticasone furoate-vilanterol group and the usual-care group (0.09 and 0.08 exacerbations per year, respectively; the rate with fluticasone furoate-vilanterol was higher by 9.7% [95% CI, -16.9 to 44.7]; P=0.52) or in the rate of first severe exacerbation in the time-to-event analysis (hazard ratio, 1.27; 95% CI, 0.98 to 1.66; P=0.08).

There was no significant difference between the fluticasone furoate-vilanterol group and the usual-care group in the annual rate of COPD-related contact with primary care; the rate was 1.7% (95% CI, -5.1 to 8.0) lower in the fluticasone furoate-vilanterol group than in the usual-



care group. The annual rate of all primary care contacts was slightly higher (12.3%; 95% CI, 5.4 to 19.6) in the fluticasone furoate–vilanterol group than in the usual-care group. There were no significant differences in the rate of secondary health care contacts.

In an analysis that was based on the entire trial population, 596 of 1317 patients (45%) in the fluticasone furoate–vilanterol group had a decrease in their CAT score by 2 or more points (indicating an improvement in COPD-related health status), as compared with 481 of 1325 patients (36%) in the usual-care group (odds ratio in favor of fluticasone furoate–vilanterol, 1.51; 95% CI, 1.28 to 1.77;  $P < 0.001$ ). There was no significant between-group difference in the change from baseline in the EQ-5D score. Results in the primary effectiveness analysis population were similar to those that were based on the entire trial population.

#### SAFETY

The incidence of serious adverse events during treatment was similar in the fluticasone furoate–vilanterol group and the usual-care group (with

events occurring in 404 patients [29%] and 383 patients [27%], respectively). There was no notable difference between the two groups with regard to any adverse event of special interest (Table 2). A total of 94 patients (7%) in the fluticasone furoate–vilanterol group had one or more serious adverse events listed as pneumonia, as compared with 83 (6%) in the usual-care group (incidence ratio, 1.1; 95% CI, 0.9 to 1.5). For the comparison of the fluticasone furoate–vilanterol group with the usual-care group, there was a trend toward a higher mean number of serious adverse events of pneumonia in the stratum receiving a treatment regimen without an inhaled glucocorticoid at randomization (mean annual rate, 3.01 hospitalizations; 95% CI, 0.97 to 9.33) than in the strata receiving an inhaled glucocorticoid at randomization ( $P = 0.10$  for the interaction of treatment with baseline maintenance therapy in the analysis across the three strata). A total of 13 patients (1%) in each group had an event of pneumonia (adverse event of special interest) with a fatal outcome. A total of 45 patients in the fluticasone furoate–vilanterol group and 30 in the usual-care group died dur-

ing the trial; reported causes of fatal events are listed in the Supplementary Appendix. One patient in each group died from a serious adverse event that was recorded as being related to the trial medication (pneumonia in 1 patient in the usual-care group, and pulmonary embolism and deep-vein thrombosis in 1 in the fluticasone furoate–vilanterol group). No subgroups with a higher risk of a serious adverse event of pneumonia in the fluticasone furoate–vilanterol group than in the usual-care group were identified.

## DISCUSSION

The Salford Lung Study on COPD was a large, randomized, comparative effectiveness trial conducted in a population that was intended to represent that seen in everyday clinical practice. We found that a simple, once-daily treatment with an inhaled combination of fluticasone furoate and vilanterol was superior to usual care by the patient's general practitioner with regard to the frequency of moderate or severe exacerbations and was not associated with a significantly higher risk of serious adverse events.

The combination of fluticasone furoate and vilanterol has been shown previously to result in lower rates of exacerbations of COPD than vilanterol alone in conventional randomized, controlled trials of efficacy.<sup>12</sup> However, this trial shows that broad populations of patients with COPD benefit from treatment with fluticasone furoate–vilanterol, and the findings differ from those of efficacy trials in which fluticasone furoate–vilanterol was associated with outcomes that were similar to those with the twice-daily combination of fluticasone propionate and salmeterol.<sup>13</sup> We found no excess number of serious adverse events of pneumonia in the overall comparisons, but as expected, we found a trend toward a greater number of serious adverse events of pneumonia with fluticasone furoate–vilanterol than with a treatment regimen consisting of bronchodilators only.<sup>14</sup> Also, we cannot rule out a higher incidence of mild pneumonia with fluticasone furoate–vilanterol than with usual care.

The strength of the trial derives from its innovative design. It took place in a single urban area, with primary and secondary care connected through an EHR, integrated with a new data recording system to enable the collection of a trial-relevant data set that contained more than

**Table 2. Serious Adverse Events of Special Interest during Treatment in the Entire Trial Population.\***

| Event   | Usual Care<br>(N=1403) | Fluticasone<br>Furoate–<br>Vilanterol<br>(N=1396) |
|---|------------------------|---|
|   | number (percent)       |   |
| Cardiovascular event                                      |                        |   |
| Any event   | 107 (8)                | 108 (8)   |
| Cardiac arrhythmia  | 54 (4)                 | 52 (4)  |
| Cardiac failure   | 28 (2)                 | 28 (2)  |
| Cardiac ischemia  | 33 (2)                 | 34 (2)  |
| Hypertension  | 1 (<1)                 | 0   |
| Stroke  | 25 (2)                 | 21 (2)  |
| Pneumonia   | 83 (6)                 | 94 (7)  |
| Lower respiratory tract infection,<br>excluding pneumonia | 58 (4)                 | 64 (5)  |
| Decreased bone mineral density<br>and associated fracture | 45 (3)                 | 45 (3)  |
| Effects on glucose level                                  | 16 (1)                 | 23 (2)  |
| Hypersensitivity  | 10 (1)                 | 10 (1)  |
| Effects on potassium level                                | 2 (<1)                 | 2 (<1)  |
| Glucocorticoid-associated eye disease                     | 2 (<1)                 | 2 (<1)  |
| Local effects from glucocorticoids                        | 1 (<1)                 | 0   |

\* Serious adverse events of special interest during treatment that were associated with the known pharmacologic action of inhaled glucocorticoids or long-acting beta-agonists were identified with the use of standardized *Medical Dictionary for Regulatory Activities* (MedDRA), version 18.1, queries (SMQs) and sponsor-defined special interest terms when no SMQ was available. The grouping of events was defined according to standard MedDRA groups, if available.

3 million lines of data for all the effectiveness and safety outcomes. After randomization, a patient was contacted by telephone only as a safety check on three occasions over a period of 12 months, and only then if there had been no health care contact within a 3-month period. All treatment was carried out by the usual caregivers, while patients were simultaneously monitored remotely with the use of the EHR for the early detection of safety events.

This comparative effectiveness trial needs careful interpretation. Although randomized, the trial was an open-label trial, which could potentially have introduced bias, although we made all efforts to have the treatment experience be similar for all the patients, by giving them similar initial training on the use of the inhaler, having them obtain their prescriptions from the general prac-

tioner, having them collect the medication at their usual pharmacy, and so forth. However, the unblinded trial is the likely reason for the larger degree of switching of treatment over the first 3 months of the trial in the fluticasone furoate-vilanterol group than in the usual-care group. Patients switched to familiar treatment, despite fewer changes that were due to treatment failure in the fluticasone furoate-vilanterol group than in the usual-care group (i.e., need for better control). It should be noted that approximately 50% of the patients were taking triple therapy despite well-preserved lung function. A considerable proportion of patients had a diagnosis of asthma recorded. We do not believe that all these patients had an asthma-COPD overlap syndrome<sup>15</sup>; instead, the finding could indicate that some patients with COPD received a diagnosis of asthma early in the course of their COPD. This situation would usually have led to exclusion from COPD efficacy studies. Most of the general practitioners also took part in the Salford Lung Study involving patients with asthma<sup>16</sup> and thus had no incentive to include patients with current asthma in this trial.

Our findings challenge the automatic transfer of findings from efficacy studies to clinical guidelines or everyday clinical practice. For any

new treatment, safety and efficacy randomized, controlled trials are essential, but they are carried out in carefully selected populations, from which patients with coexisting conditions are excluded, and represent less than 10% of patients with COPD.<sup>2</sup> Frequent face-to-face monitoring ensures high adherence to therapy and good inhaler technique. This comparative effectiveness trial that was conducted in a population of patients with COPD was largely unsupervised over the yearlong period, which allowed important factors in usual clinical care, such as adherence, frequency of dosing, and persistence of good inhaler technique, to come into play.

In conclusion, patients in general practice who had a diagnosis of COPD and a heightened risk of exacerbations had a benefit with simple, once-daily inhaled combination treatment with fluticasone furoate and vilanterol, without an additional risk of serious adverse events. Future effectiveness studies are likely to influence clinical guidelines, not only for COPD but for many other chronic diseases.

Supported by GlaxoSmithKline.

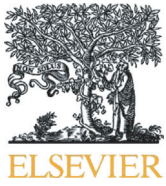
Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank Charlotte Kennerley, Ph.D., of Gardiner-Caldwell Communications, for assistance with the preparation of an earlier version of the figure.

## REFERENCES

- Vestbo J, Hurd SS, Agustí AG, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2013;187:347-65.
- Herland K, Akselsen JP, Skjønberg OH, Bjermer L. How representative are clinical study patients with asthma or COPD for a larger "real life" population of patients with obstructive lung disease? *Respir Med* 2005;99:11-9.
- Berger ML, Dreyer N, Anderson F, Towse A, Sedrakyan A, Normand SL. Prospective observational studies to assess comparative effectiveness: the ISPOR Good Research Practices Task Force report. *Value Health* 2012;15:217-30.
- Chalkidou K, Tunis S, Whicher D, Fowler R, Zwarenstein M. The role for pragmatic randomized controlled trials (pRCTs) in comparative effectiveness research. *Clin Trials* 2012;9:436-46.
- NorthWest EHealth home page (<http://nweh.co.uk/>).
- Fiore LD, Lavori PW. Integrating randomized comparative effectiveness research with patient care. *N Engl J Med* 2016;374:2152-8.
- New JP, Bakerly ND, Leather D, Woodcock A. Obtaining real-world evidence: the Salford Lung Study. *Thorax* 2014;69:1152-4.
- Bakerly ND, Woodcock A, New JP, et al. The Salford Lung Study protocol: a pragmatic, randomised phase III real-world effectiveness trial in chronic obstructive pulmonary disease. *Respir Res* 2015;16:101.
- Jones PW. COPD assessment test — rationale, development, validation and performance. *COPD* 2013;10:269-71.
- EuroQol Group. EuroQol — a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199-208.
- Elkhenini HE, Davis KJ, Stein ND, et al. Using an electronic medical record (EMR) to conduct clinical trials: Salford Lung Study feasibility. *BMC Med Inform Decis Mak* 2015;15:8.
- Dransfield MT, Bourbeau J, Jones PW, et al. Once-daily inhaled fluticasone furoate and vilanterol versus vilanterol only for prevention of exacerbations of COPD: two replicate double-blind, parallel-group, randomised controlled trials. *Lancet Respir Med* 2013;1:210-23.
- Agustí A, de Teresa L, De Backer W, et al. A comparison of the efficacy and safety of once-daily fluticasone furoate/vilanterol with twice-daily fluticasone propionate/salmeterol in moderate to very severe COPD. *Eur Respir J* 2014;43:763-72.
- Wedzicha JA, Banerji D, Chapman KR, et al. Indacaterol-glycopyrronium versus salmeterol-fluticasone for COPD. *N Engl J Med* 2016;374:2222-34.
- Postma DS, Rabe KF. The asthma-COPD overlap syndrome. *N Engl J Med* 2015;373:1241-9.
- Woodcock A, Bakerly ND, New JP, et al. The Salford Lung Study protocol: a pragmatic, randomised phase III real-world effectiveness trial in asthma. *BMC Pulm Med* 2015;15:160.

Copyright © 2016 Massachusetts Medical Society.



## Trial Designs

## High-dose influenza vaccine to reduce clinical outcomes in high-risk cardiovascular patients: Rationale and design of the INVESTED trial



Orly Vardeny, PharmD MS<sup>a,\*</sup>, Jacob A. Udell, MD, MPH<sup>b,1</sup>, Jacob Joseph, MD<sup>c</sup>, Michael E. Farkouh, MD, MSc<sup>d</sup>, Adrian F. Hernandez, MD<sup>e</sup>, Alison J. McGeer, MD<sup>f</sup>, H. Keipp Talbot, MD MPH<sup>g</sup>, Deepak L. Bhatt, MD<sup>h</sup>, Christopher P. Cannon, MD<sup>h</sup>, Shaun G. Goodman, MD<sup>i</sup>, Inder Anand, MD<sup>j</sup>, David L. DeMets, PhD<sup>k</sup>, Jon Temte, MD, PhD<sup>l</sup>, Janet Wittes, PhD<sup>m</sup>, Kristin Nichol, MD<sup>a</sup>, Clyde W. Yancy, MD<sup>n</sup>, J. Michael Gaziano, MD<sup>c</sup>, Lawton S. Cooper, MD, MPH<sup>o</sup>, KyungMann Kim, PhD<sup>k</sup>, Scott D. Solomon, MD<sup>c</sup>

<sup>a</sup> Center for Chronic Disease Outcomes Research, Minneapolis VA Health Care System, and University of Minnesota, Minneapolis, MN

<sup>b</sup> Cardiovascular Division, Department of Medicine, Peter Munk Cardiac Centre, Toronto General Hospital and Women's College Hospital, University of Toronto, Toronto, Ontario, Canada

<sup>c</sup> Boston VA Health Care System, and Cardiovascular Division, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

<sup>d</sup> Cardiovascular Division, Department of Medicine, Peter Munk Cardiac Centre, University of Toronto, Toronto, Ontario, Canada

<sup>e</sup> Cardiology Division, Department of Medicine, Duke University School of Medicine, Durham, NC

<sup>f</sup> Department of Microbiology, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada

<sup>g</sup> Division of Infectious Diseases, Department of Medicine, Vanderbilt University, Nashville, TN

<sup>h</sup> Brigham and Women's Hospital Heart & Vascular Center, and Harvard Medical School, Boston, MA

<sup>i</sup> Division of Cardiology, Department of Medicine, St Michael's Hospital, University of Toronto, Toronto, Ontario, Canada

<sup>j</sup> San Diego and Minneapolis VA Health Care Systems and University of Minnesota, San Diego, CA, and Minneapolis, MN

<sup>k</sup> Department of Statistics and Biostatistics, University of Wisconsin-Madison, Madison, WI

<sup>l</sup> Department of Family Medicine and Community Health, University of Wisconsin-Madison, Madison, WI

<sup>m</sup> Statistics Collaborative, Inc., Washington, DC

<sup>n</sup> Cardiovascular Division, Northwestern University Feinberg School of Medicine, Chicago, IL

<sup>o</sup> National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD

Declarations of interest: O. V. has received research funding from Novartis and Sanofi-Pasteur. J. A. U. has received research funding from AstraZeneca, Novartis, and Sanofi Aventis; consultancy fees from Amgen, Boehringer Ingelheim (BI), Janssen, Merck, Novartis, and Sanofi-Pasteur; and honoraria for lectures from BI and Janssen. S. D. S. has received research funding from Sanofi-Pasteur. H. K. T. has received research funding from Sanofi Pasteur, MedImmune, and Gilead and serves as a safety consultant for VaxInnate and Seqiris. A. J. M. has received research support from GlaxoSmithKline and Sanofi Pasteur. A. F. H. has received consultancy fees from Amgen, AstraZeneca, Bayer, BI, Boston Scientific, Merck, Novartis, and Sanofi Aventis and research support from AstraZeneca, GlaxoSmithKline, Luitpold, Merck, and Novartis. Dr D. L. B. discloses the following: Advisory Board: Cardax, Elsevier Practice Update Cardiology, Medscape Cardiology, Regado Biosciences; Board of Directors: Boston VA Research Institute, Society of Cardiovascular Patient Care; Chair: American Heart Association Quality Oversight Committee; Data Monitoring Committees: Baim Institute for Clinical Research (formerly Harvard Clinical Research Institute), Cleveland Clinic, Duke Clinical Research Institute, Mayo Clinic, Mount Sinai School of Medicine, Population Health Research Institute; Honoraria: American College of Cardiology (Senior Associate Editor, *Clinical Trials and News*, ACC.org; Vice-Chair, ACC Accreditation Committee), Baim Institute for Clinical Research (formerly Harvard Clinical Research Institute; clinical trial steering committee), Belvoir Publications (Editor in Chief, *Harvard Heart Letter*), Duke Clinical Research Institute (clinical trial steering committees), HMP Global (Editor in Chief, *Journal of Invasive Cardiology*), *Journal of the American College of Cardiology* (Guest Editor; Associate Editor), Population Health Research Institute (clinical trial steering committee), Slack Publications (Chief Medical Editor, *Cardiology Today's Intervention*), Society of Cardiovascular Patient Care (Secretary/Treasurer), WebMD (CME steering committees); Other: *Clinical Cardiology* (Deputy Editor), NCDR-ACTION Registry Steering Committee (Chair), VA CART Research and Publications Committee (Chair); Research Funding: Abbott, Amgen, AstraZeneca, Bristol-Myers Squibb (BMS), Chiesi, Eisai, Ethicon, Forest Laboratories, Idorsia, Ironwood, Ischemix, Lilly, Medtronic, PhaseBio, Pfizer, Regeneron, Roche, Sanofi Aventis, Synaptic, The Medicines Company; Royalties: Elsevier (Editor, *Cardiovascular Intervention: A Companion to Braunwald's Heart Disease*); Site Co-Investigator: Biotronik, Boston Scientific, St Jude Medical (now Abbott), Svelte; Trustee: American College of Cardiology; Unfunded Research: FlowCo, Merck, Plx Pharma, Takeda. C. P. C. receives research grants from Amgen, Arisaph, BI, BMS, Daiichi Sankyo, Janssen, Merck, and Takeda; consulting fees from Alnylam, Amgen, Arisaph, Astra Zeneca, BI, BMS, Eisai, GlaxoSmithKline, Kowa, Lipimedix, Merck, Pfizer, Regeneron, Sanofi, and Takeda. S. G. G. receives research grant support and speaker/consulting honoraria from Amgen, AstraZeneca, Bayer, BI, Bristol Myers Squibb, CSL Behring, Daiichi-Sankyo, Eli Lilly, Fenix Group International, Ferring Pharmaceuticals, GlaxoSmithKline, Janssen/Johnson & Johnson, Luitpold Pharmaceuticals, Matrizyme, Merck, Novartis, Pfizer, Regeneron, Sanofi, Servier, Tenax Therapeutics, Heart and Stroke Foundation of Ontario/University of Toronto, Canadian Heart Research Centre and MD Primer, Canadian VIGOUR Centre, and Duke Clinical Research Institute. I. A. receives consultant fees from Amgen, Arca, AstraZeneca, BI, Cyberonics, Novartis, and Zensun. D. L. D. serves as a consultant to the National Institutes of Health and the Food and Drug Administration, and serves as an independent biostatistician serving as a member of a trial steering committee or data and safety monitoring boards for Actelion, Amgen, Astra-Zeneca, Bayer, BI, GlaxoSmithKline, Merck, Novartis, Otsuka, Pfizer, Roche, Sanofi-Aventis, Duke Clinical Research Institute, Harvard Clinical Research Institute, and McMaster University Population Health Research Institute. J. T. served as a consultant for Sanofi Pasteur and received in-kind research support from Quidel Corporation. J. W. has served on the data and safety monitoring board for Sanofi-Pasteur, and the company for which J. W. works and is part-owner has contracts with many companies that produce drugs, biologics, and devices, including Sanofi Pasteur for development of influenza vaccines. K. N. and C. W. Y. report no disclosures.

RCT# NCT02787044.

\* Reprint requests: Orly Vardeny, PharmD, MS, Minneapolis VA Health Care System and University of Minnesota, 1 Veterans Dr, Minneapolis, MN 55417.

E-mail address: [ovardeny@umn.edu](mailto:ovardeny@umn.edu) (O. Vardeny).

<sup>1</sup> The first 2 authors contributed equally to the manuscript.



## ARTICLE INFO

## Article history:

Received 10 April 2018

Accepted 18 May 2018

Available online xxxx

## ABSTRACT

**Background:** Influenza leads to significant cardiopulmonary morbidity and mortality—particularly in patients with cardiovascular disease—that may be prevented with a standard influenza vaccine. However, patients with cardiovascular conditions have a reduced immune response to influenza vaccine, potentially resulting in reduced effectiveness for preventing clinical events. High-dose vaccine augments immune response in cardiac patients, suggesting that a high-dose influenza vaccination strategy may further reduce morbidity and mortality. Alternatively, broader coverage with an influenza vaccine containing an increased number of viral strains is an alternative strategy without direct evaluation.

**Research design and methods:** INfluenza Vaccine to Effectively Stop Cardio Thoracic Events and Decompensated heart failure (INVESTED) is a pragmatic, randomized, double-blind, parallel-group, active-controlled trial comparing the effectiveness of an annual vaccination strategy of high-dose trivalent versus standard-dose quadrivalent influenza vaccine in patients with a history of recent heart failure or myocardial infarction hospitalization. The trial will enroll approximately 9,300 patients over 4 influenza seasons. The primary hypothesis is that high-dose influenza vaccine will reduce the composite outcome of all-cause mortality and hospitalization from a cardiovascular or pulmonary cause compared with standard-dose influenza vaccine within each enrolling season. Approximately 1,300 primary outcome events will provide >90% power to detect an 18% relative risk reduction at a 2-sided  $\alpha$  level of .05.

**Conclusion:** INVESTED is the largest and longest study to assess whether high-dose influenza vaccine is superior to standard-dose influenza vaccine in reducing cardiopulmonary events in a high-risk cardiovascular population (ClinicalTrials.gov Identifier: NCT02787044).

Published by Elsevier Inc.

Influenza leads to significant morbidity and mortality, particularly in patients with cardiovascular disease.<sup>1</sup> Influenza infection has been temporally associated with acute cardiovascular events, such as acute coronary syndrome and acute heart failure (HF).<sup>2–4</sup> Because of the increased risk for influenza-related complications, annual influenza immunization is recommended by the Centers for Disease Control and Prevention (CDC) and cardiovascular professional societies.<sup>5–7</sup> Moreover, influenza vaccination has been associated with reduced cardiac-related hospital admissions, acute exacerbations of HF, and winter mortality.<sup>8,9</sup> In a meta-analysis of clinical trials testing the efficacy of influenza vaccination in patients at cardiovascular risk, annual vaccination reduced the risk for major adverse cardiovascular events by 36%.<sup>10</sup> Numerous vaccine formulations are available, differing on number and dose of viral antigens, preparation (egg-based versus recombinant), and presence of adjuvant. Vaccine antigen composition changes annually in an effort to harmonize with circulating strains, and each year, virulence of influenza varies, as does the match between vaccine antigens and circulating strains.

Several lines of evidence suggest that a strategy of using high-dose influenza vaccine in high-risk cardiovascular patients might reduce more morbidity and mortality than the standard-dose vaccine. Immune response to influenza vaccine varies with age and concomitant medical conditions and is referred to as *immunosenescence*. Immunosenescence is present in patients with HF as evidenced by lower antibody titers after standard influenza vaccination compared with healthy controls.<sup>11</sup> In a randomized trial, we demonstrated that antibody responses in patients with HF were augmented by a higher dose of influenza vaccine.<sup>12</sup> A large randomized trial of high-dose versus standard-dose influenza vaccine in medically stable patients older than 65 years showed a 24% risk reduction in laboratory-confirmed influenza and influenza-like illness (ILI) without substantial adverse events.<sup>13</sup> Additionally, results supported the potential for reduced cardiopulmonary hospitalizations among those assigned to high-dose vaccine.<sup>14</sup> Incorporating these results with a meta-analysis of trials comparing high-dose and standard-dose influenza vaccine, higher-dose influenza vaccination was associated with a 27% reduced risk for major adverse cardiovascular events compared to standard-dose vaccine.<sup>10</sup> High-dose influenza vaccine is Food and Drug Administration–approved for prevention of influenza in medically stable adults older than 65 years but is not currently indicated for patients younger than 65 years, and there are limited data in those with unstable, high-risk medical conditions. On the other hand, high-dose vaccine is only currently available in trivalent (3 viral strains) presentation, whereas standard-dose vaccine is also offered as a

quadrivalent (4 viral strains, containing an additional B-lineage strain) presentation. The Advisory Committee on Immunization Practices, which informs the CDC vaccine guidelines, does not preferentially recommend one influenza vaccine formulation over another, and the vaccine formulation that offers the most clinical protection in these high-risk patients is unknown.<sup>15</sup>

The high morbidity and health care costs among patients with high-risk cardiovascular disease along with the reduced immune responses to standard-dose influenza vaccines in patients with heart disease provide a compelling rationale to investigate alternative influenza vaccination strategies in this group. Accordingly, we designed an outcomes study to test the hypothesis that in patients with recent acute myocardial infarction (AMI) or HF hospitalization, a trivalent influenza vaccine with 4 times the dose of hemagglutinin antigen will reduce major cardiovascular and pulmonary-related morbidity and mortality compared with standard-dose quadrivalent vaccine. The comparative efficacy and safety of these influenza vaccines will be assessed over the course of 4 individual seasons in an amalgamated fashion, as well as each year independently, to accommodate for seasonal variation in influenza virulence and vaccine effectiveness.

### Trial design and methods

INVESTED is a randomized, double-blind, parallel-group, active-controlled, 2-arm study comparing the effectiveness of high-dose versus standard-dose influenza vaccine in reducing all-cause mortality or cardiopulmonary hospitalizations in high-risk cardiovascular patients. The trial was designed by members of the Executive and Steering Committees in collaboration with the National Heart, Lung, and Blood Institute (NHLBI). The trial has been registered on ClinicalTrials.gov (NCT02787044).

### Patients

The eligibility criteria are summarized in Table 1. Briefly, eligible patients are 18 years of age or older with a documented history of either a hospitalization for spontaneous (type 1) or secondary (type 2) MI within a year of the study baseline visit, or a history of hospitalization for HF within 2 years of the baseline visit. In addition, patients need to fulfill at least 1 additional enrichment criterion (Table 1). Enrichment criteria were selected in consideration that they select for patients at high risk for the primary end point as well as for immunosenescence with standard-dose influenza vaccine. (See Fig. 1.)

**Table I**  
Eligibility criteria

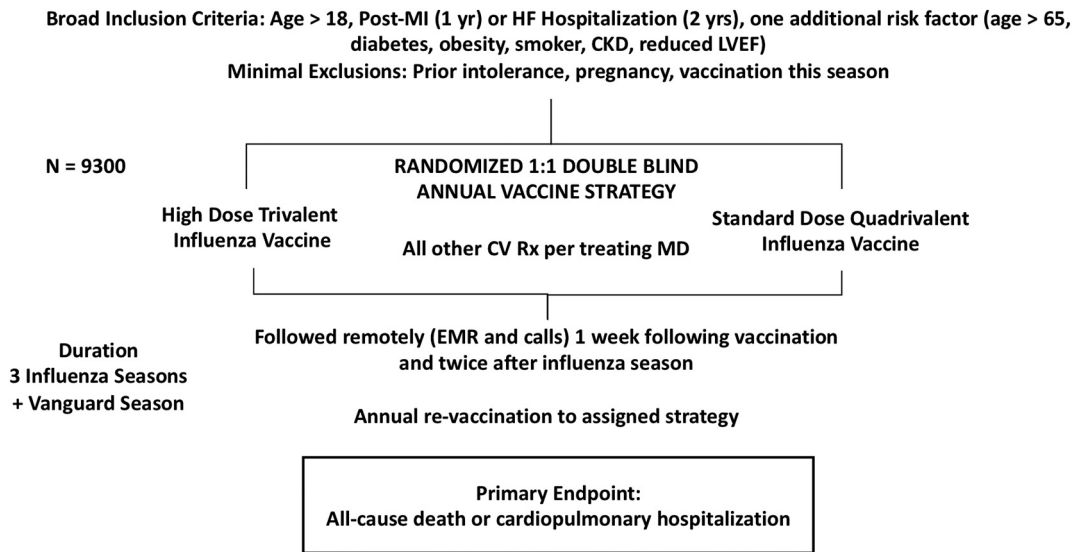
**Inclusion criteria**

1. Willing to give written informed consent and able and willing to adhere to follow-up schedules
2. At least 18 y of age
3. Documented history of at least 1 of the below CV events:
  - a. Hospitalization for spontaneous MI (type 1 or type 2 event) (within 1 y of baseline visit)
  - b. Hospitalization for HF (within 2 y of baseline visit) but not currently acutely decompensated.
4. Fulfills at least 1 of the following additional risk factors:
  - a. Prior MI hospitalization (for participants qualifying on HF hospitalization or a second MI hospitalization for those qualifying based on MI)
  - b. Prior HF hospitalization (for participants qualifying based on MI hospitalization or a second HF hospitalization for those qualifying based on HF)
  - c. Age ≥65 y
  - d. Current or historical LVEF <40%
  - e. Documented diagnosis (via ICD-9 code) of type 1 or type 2 diabetes mellitus (laboratory findings, eg, elevated A1C, FPG, plasma glucose in the absence of a clinical diagnosis is not sufficient)
  - f. Current BMI ≥30
  - g. Documented history of renal impairment (eGFR ≤60 for at least 2 readings in the past year)
  - h. Documented history of ischemic stroke
  - i. Documented history of peripheral artery disease
  - j. Current tobacco smoker (smokes 1 or more cigarettes daily)

**Exclusion criteria**

1. Known allergy, hypersensitivity (anaphylaxis), or Guillain-Barré syndrome within 6 wk after influenza vaccine, or severe allergy to egg protein
2. Any noncardiac condition that, in the opinion of the investigator, would lead to life expectancy <9 m
3. Receipt of influenza vaccine during current influenza season
4. Any acute infection requiring antibiotics within 14 d of influenza vaccination (prophylactic antibiotics prior to dental or other procedures, or scheduled use of antibiotics for other types of prophylaxis does not exclude the subject). If an acute course of antibiotics is required, the patient may still participate in INVESTED 14 d after completing antibiotics.
5. Known fever over 100°F or 38°C within 7 d of influenza vaccination
6. Women who are pregnant or breast-feeding
7. Not suitable for study participation due to other reasons at the discretion of the investigator

ICD-9, International Classification of Diseases, Ninth Revision; FPG, fasting plasma glucose.



**Figure 1.** Study schematic.

Key exclusion criteria (Table I) include known allergy or hypersensitivity to influenza vaccine, history of serious adverse reaction to influenza vaccine, any condition that would lead to life expectancy of <9 months, prior receipt of influenza vaccine for the upcoming influenza season, infection requiring antibiotics in the 14 days prior to randomization, known fever within 7 days of randomization, pregnancy, or lactation. Enrollment in INVESTED began on September 21, 2016, following protocol approval by the study's Protocol Review Committee and an Institutional Review Board affiliated with each investigative site. The study will include approximately 200 sites in the United States and Canada. The study is being conducted in accordance with Good Clinical Practice and the Declaration of Helsinki 2002.

*Study objectives*

The primary objective of this study is to compare high-dose trivalent inactivated influenza vaccine (IIV3-HD) with standard-dose quadrivalent inactivated influenza vaccine (IIV4) on time to first occurrence of death or cardiopulmonary hospitalization within each enrolling season (Table II). Secondary objectives are to compare the effect of high-dose influenza vaccine versus standard-dose vaccine on total (first and recurrent) cardiopulmonary hospitalizations or death, time to first occurrence of cardiovascular death or cardiovascular hospitalization within each enrolling season, time to first occurrence of death or cardiopulmonary hospitalization across all enrolling seasons, and the individual

**Table II**  
Study objectives

|   |
|---|
| <p>Primary objective</p> <p>To compare the effects of high-dose influenza trivalent vaccine to standard-dose quadrivalent vaccine on time to first occurrence of death or cardiopulmonary hospitalization within each enrolling season</p> <p>Secondary objectives</p> <p>To compare the effects of high-dose influenza trivalent vaccine to standard-dose quadrivalent vaccine on:</p> <ul style="list-style-type: none"> <li>• Total (first and recurrent) cardiopulmonary hospitalizations or death</li> <li>• Time to first occurrence of cardiovascular death or cardiovascular hospitalization within each enrolling season</li> <li>• Time to first occurrence of death or cardiopulmonary hospitalization across all enrolled influenza seasons</li> <li>• Time to first occurrence of the individual components of the primary end point</li> </ul> <p>Exploratory objectives</p> <p>To compare the effects of high-dose influenza trivalent vaccine to standard-dose quadrivalent vaccine on:</p> <ul style="list-style-type: none"> <li>• Time to first occurrence of all-cause death or cardiopulmonary hospitalization according to effectiveness of vaccine relative to virulence of influenza strain and the quality of the match between influenza strain and vaccine within individual seasons</li> <li>• Time to first occurrence of cardiovascular death or HF hospitalization</li> <li>• Time to first occurrence of cardiovascular death, nonfatal MI, or nonfatal stroke</li> <li>• Time to first occurrence of all-cause death and cardiopulmonary hospitalizations</li> </ul> |
|---|

components of the primary end point. Exploratory objectives are listed in Table II.

### Study design

#### Identification of patients and enrollment

The study is comprised of several networks of performance sites: a Canadian network, a network of Veterans Administration sites, a network of other US non-Veterans Administration sites, and a network of sites from PCORnet (the National Patient-Centered Clinical Research Network). Several recruitment strategies will be used, and sites within each network may use a combination of methods depending on their capabilities. Networks and sites with electronic health record abilities may query electronic health records based on study enrolment criteria and create screening lists for individual site principal investigators, which will be forwarded to site research personnel in the early summer months prior to each enrolling season, and may include electronic contact of potential participants. Participants can be enrolled prior to discharge from a hospitalization for acute HF or myocardial infarction once no longer acutely decompensated. Screening can also occur anytime as part of an inpatient assessment or outpatient visit in a cardiology or primary care clinic, cardiac rehabilitation visit, or other clinical setting. Enrollment and randomization nevertheless will be timed to coincide with study vaccine availability with confirmation of participant eligibility at the baseline visit. Individual sites may use additional strategies for which IRB approval will be obtained prior to implementation.

Participants can be enrolled for up to 3 influenza seasons and will be vaccinated with the same vaccine strategy (high dose or standard dose) to which they were randomized during their first enrollment season using each year's World Health Organization-recommended composition of viral antigens.

#### Vaccination and randomized double-blind treatment period

Participants will be assigned to receive 1 of 2 formulations of influenza vaccine: IIV3-HD or IIV4. IIV3-HD is currently the only available higher-dose formulation. Nevertheless, we chose to use IIV4 as the comparator because this vaccine was projected to potentially become standard of care in the regions the trial was being conducted and because of the potential theoretical advantages of the additional B-lineage coverage in the quadrivalent vaccine. Thus, an IIV-4 represented a comparator for which there remained equipoise to determine which strategy was superior. Vaccine will be administered intramuscularly once at randomization and yearly thereafter. Both vaccines are licensed in the United States and

Canada. IIV4 is indicated for active immunization of persons 6 months of age and older against influenza disease caused by influenza virus subtypes A and type B. A single injectable sterile suspension 0.5-mL dose contains 15 µg of hemagglutinin from each of 4 viral strains for a total of 60 µg in 1 dose. IIV3-HD is indicated for active immunization of persons 65 years of age and older against influenza disease caused by influenza virus subtypes A and type B. A single injectable sterile suspension 0.5-mL dose contains 60 µg of hemagglutinin from 3 viral strains for a total of 180 µg in 1 dose. Both inactivated influenza vaccines are prepared from influenza viruses propagated in embryonated chicken eggs.

Sanofi Pasteur provides both formulations of vaccine as 0.5-mL single-dose, prefilled syringes. Vaccine syringes are subsequently blinded and labeled by a third-party vendor and shipped to investigator sites with a temperature-monitoring device to verify maintenance of the cold chain during transit.

#### Monitoring for safety

Following vaccine administration, participants will be monitored by site personnel for acute vaccine-related adverse events for at least 20 minutes. Participants will be provided a symptom diary to track vaccine-related events at home for 7 days. One week postvaccination, participants will be contacted by a member of the study team by phone to assess potential vaccine-related local and systemic adverse events, including allergic reactions. (see Table III)

#### Monitoring for cardiopulmonary events

Surveillance for hospitalization or death will include 1 telephone call completed by site personnel during influenza season and another phone call during the summer following influenza season. Participants will also be asked to inform local site personnel of hospitalizations at any time they occur.

#### Biomarkers, immune response, and genetic analyses

Blood will be collected in a subset of up to 3,000 consenting participants and banked for future studies. Analyses will examine associations of biomarkers that reflect immunity, inflammation, thrombosis, metabolism, and vascular or hemodynamic risk with influenza vaccine response and cardiovascular disease. One planned substudy examines postvaccination hemagglutination inhibition antibody titers in response to influenza vaccine antigens, which will be measured in participants who consent to blood draws as described above. Blood will be collected during the vaccination visit (baseline) and again 4 weeks postvaccination to test the hypothesis that a higher influenza vaccine dose will result in a more pronounced humoral immune response, evidenced by higher geometric mean titers postvaccination and greater antibody titer increases from baseline, and to test the hypothesis that higher antibody titers are associated with a reduced rate of the composite of all-cause death and cardiopulmonary hospitalization. Other key objectives include exploring the effects of each vaccination strategy on circulating biomarker levels over time and assessing the utility of incorporating biomarker levels into risk prediction models that identify patients that particularly benefit from high-dose influenza vaccine. Blood will be stored for future investigations of genetic contributors to cardiopulmonary risk and patient responsiveness to influenza vaccine.

#### Measures to minimize biases

##### Randomization

After informed consent is obtained and eligibility assessed, participants are randomized in a 1:1 ratio to IIV3-HD or IIV4 using permuted blocks of random block sizes, stratified naturally by influenza season, but no other stratification factors. Patients will receive the same dose for subsequent influenza seasons.



**Table III**  
Schedule of time and events

| Measurement                             | Screening*<br>visit | Baseline visit<br>(August-January) | Week 1 phone<br>call ( $\pm 4$ d) | Week 2-4 visit<br>( $\pm 4$ d) <sup>‡</sup> | During influenza season<br>phone call | Summer<br>phone call | Years 2 & 3 baseline <sup>§</sup><br>(August-December) |
|---|---------------------|------------------------------------|-----------------------------------|---|---------------------------------------|----------------------|--|
| Informed consent                        | X                   |                                    |                                   |   |                                       |                      |  |
| Demographics & history <sup>†</sup>     | X                   | X                                  |                                   |   |                                       |                      | X  |
| Inclusion/exclusion                     | X                   | X                                  |                                   |   |                                       |                      | X  |
| Current medications                     | X                   | X                                  |                                   |   |                                       |                      | X  |
| Blood draw <sup>†</sup>                 |                     | X                                  |                                   | X   |                                       |                      |  |
| Vaccine administration                  |                     | X                                  |                                   |   |                                       |                      | X  |
| Assessment of vaccine-related reactions |                     | X                                  | X                                 |   |                                       |                      | X  |
| Cardiopulmonary event assessment        |                     |                                    |                                   | X   | X                                     | X                    |  |
| Year 2 & 3 visit scheduling             |                     |                                    |                                   |   |                                       | X                    |  |

\* Screening and baseline procedures may be completed at 1 visit, followed by randomization and vaccine administration.

<sup>†</sup> History includes previous vaccinations.

<sup>‡</sup> Baseline and week 2-4 blood draw for immune end points (eg, geometric mean titers postvaccination, change in antibody titers at 4 weeks postvaccination, seroconversion, seroprotection, and B-type vaccine antigens 2-4 weeks postvaccination), biomarkers, and genetic markers will be assessed in a subset of up to 3,000 participants at participating sites [will be implemented after the Vanguard year], including an end point assessment at week 4.

<sup>§</sup> Years 2 and 3 follow year 1 procedures after baseline visit.

### Masking

In an effort to minimize crossover related to perceived benefit of one vaccine formulation over another, participants, site investigators, study personnel, persons performing follow-up surveillance, and study statisticians will remain masked to the identity of the treatment from the time of randomization until database lock, except for the statisticians supporting the Data and Safety Monitoring Board.

### Study management and committees

INVESTED is conducted under a cooperative agreement to the Clinical Coordinating Center and the Data Coordinating Center from the NHLBI, under the guidance and leadership of the Executive Committee which is comprised of academic members and the NHLBI project officer. An academic steering committee also advises the Executive Committee regularly. An independent, external Data and Safety Monitoring Board appointed by the NHLBI oversees the safety of the patients in the trial and reviews the results of the interim efficacy analysis. A Clinical Endpoints Committee is responsible for classifying all deaths and for adjudicating all nonfatal events.

### Statistical considerations

The primary efficacy analysis will be performed according to a modified intention-to-treat principle for the primary end point of the time to first occurrence of all-cause death or cardiopulmonary hospitalization during each enrolling season, defined as beginning 2 weeks following receipt of influenza vaccine and continuing until July 31 of the following calendar year using standard survival analysis methods. As such, participants can contribute primary end point events during multiple enrolling seasons. The primary efficacy analysis will be based on a 2-sided log-rank test at a significance level of .05, stratified by influenza season.<sup>16</sup> The Kaplan-Meier method will be used to estimate the survival distribution for the time to first occurrence of all-cause death or cardiopulmonary hospitalization within each enrolling season.<sup>17</sup> An unadjusted estimate of the hazard ratio and CI will be obtained using a Cox proportional-hazards model with only treatment as a model term, stratified by influenza season.<sup>18</sup>

To test the hypothesis that a strategy of high-dose influenza vaccine over multiple seasons will be superior to standard-dose vaccine, one of the secondary analyses will be a standard ITT analysis from the time from randomization until final subject censoring, which will occur following the final season the subject receives study vaccination.

A sensitivity analysis is planned to account for potential differential survivorship bias and bias due to differential dropout after the initial randomization, during which we will use principal stratification,

matching based on propensity score, or inverse probability of treatment weighting for adjusted Kaplan-Meier estimator and log-rank test.<sup>19-21</sup>

Prespecified subgroups will be analyzed using Cox proportional-hazards models with age (<65 or  $\geq 65$  years old), baseline cardiovascular risk group (AMI or HF), and treatment (high-dose or standard-dose influenza vaccination) as model terms, stratified by influenza season, to obtain an adjusted hazard ratio with CIs, while adjusting for the following covariates: past vaccination history to adjust for the theoretical possibility of interference between successive vaccinations, and match between vaccine and circulating influenza strains, and the interaction between treatment and match for circulating B (Victoria)-lineage that is included only in the standard-dose IIV4 (binary), based on influenza typing and subtyping data from Canada and the United States to account for the differences in B vaccine antigens present only in the IIV4.

A secondary “in season” analysis will also be undertaken, limited to an evaluation of efficacy during the formally delineated influenza season with start and end of season defined according to the CDC and Public Health Agency of Canada surveillance system. For example, we will use information provided in the CDC’s Flu View Report which is updated on a weekly basis (<http://www.cdc.gov/flu/weekly/>). For each US state, we will use the point at which influenza transitions from “sporadic” to “local” on the graphic “Geographic Spread of Influenza as Assessed by State and Territorial Epidemiologists” or by using the point of transition from “minimal” to “low” activity on the “ILI Net State Activity Indicator Map.” We will adopt a similar approach for each Fluwatch region in Canada (<https://www.canada.ca/en/public-health/services/diseases/flu-influenza/influenza-surveillance/weekly-influenza-reports.html>) using the transition from “sporadic” to “local” on the map of ILI activity for each region.

To assess the independence of the primary end points from year to year in individuals receiving influenza vaccines more than once, the frailty model version of the Cox proportional-hazards regression will be evaluated.<sup>22</sup> In case the independence assumption is not tenable, we will estimate intrasubject correlation from year to year using the method of Prentice and Cai.<sup>23</sup>

### Analysis of secondary end points

Secondary end points consist of total (first and recurrent) cardiopulmonary hospitalizations or all-cause death during the subject’s entire study participation duration, the composite of cardiovascular death or cardiopulmonary hospitalization within each enrolling season, the composite of all-cause death or cardiopulmonary hospitalization across all enrolling seasons, and individual components of the primary end point, including time to all-cause death and time to first occurrence of cardiopulmonary hospitalization. Time to composite end points and times to individual components of the composite end points will be analyzed similarly as the primary end point with individual components of

the composite end points that are nonterminating events analyzed using methods for competing risks.<sup>24</sup> Recurrent events analysis will be performed for recurrent nonterminating events across all enrolling seasons.<sup>25–27</sup> For all analyses, 2-sided *P* values <.05 will be considered statistically significant. In addition, the rate of cardiopulmonary hospitalization with death as competing risk will be analyzed using nonparametric and semiparametric analyses based on the *mean frequency function* defined as the marginal mean of the cumulative number of cardiopulmonary hospitalizations over time subject to a terminal event of death.<sup>28,29</sup>

#### Sample size and power

The enrollment target is approximately 4,650 participants per treatment arm, for a total of 9,300 participants. This is based on an estimated treatment effect size of IIV3-HD versus IIV4 of 18% risk reduction, that is, a hazard ratio of 0.82, in all-cause death or cardiovascular hospitalizations, with an anticipated similar magnitude of benefit for all-cause death or cardiopulmonary hospitalizations. This estimate is derived from our meta-analysis of randomized trials of relatively healthy outpatients comparing these 2 active vaccination treatments, using an estimated risk reduction of 27% for the composite end point, reduced by 35% for dilution of the treatment effect among those with active heart disease. Based on data from contemporary clinical trials of patients with coronary heart disease or HF, the event rate for the primary end point is estimated to be 9% during the subject's first enrolling season following randomization for each subject, 8% during the second enrolling season, and 7% during her third enrolling season after vaccination. The primary composite end point events are assumed to be 30% deaths and 70% cardiopulmonary hospitalizations. Considering a follow-up to the end of enrolling season (before the next influenza season) and a conservative 30% rate of not being vaccinated in a subsequent influenza season, a trial of 9,300 participants over a pilot season (n=500) during 2016/2017 and 3 subsequent influenza seasons in 2017–2018, 2018–2019, and 2019–2020 is projected to result in 45, 291, 440, and 519 primary end point events by the end of the 2019–2020 enrolling season, for a total of 1,296 events, with each patient possibly contributing primary end point events over multiple seasons. Assuming 2 interim analyses for efficacy using the O'Brien-Fleming group sequential method at the end of 2017–2018 and 2018–2019 enrolling seasons,<sup>30</sup> the trial will have power of 0.94 to detect an 18% risk reduction at a 2-sided significance level of .05.

#### Discussion

Influenza infection is associated with substantial morbidity and mortality in patients with cardiovascular disease. Although influenza vaccination is recommended in patients with cardiovascular conditions, the effectiveness may be limited because of relative immunosenescence in patients with cardiovascular conditions, and data from several trials and a meta-analysis suggest that a more effective vaccination strategy could potentially mitigate the reduced immune response. INVESTED will directly test the hypothesis that high-dose influenza vaccine reduces all-cause mortality and cardiopulmonary hospitalizations in high-risk cardiovascular patients compared with standard-dose vaccine. This trial has the potential to inform guidelines and public policy regarding use of influenza vaccine in high-risk patients.

Several elements in the design of INVESTED are worthy of consideration. We are using an active control rather than placebo because influenza vaccination is considered standard of care for influenza prevention in the United States and Canada, although a significant proportion of patients with heart disease may not get vaccinated.<sup>31</sup> INVESTED is enrolling participants over multiple consecutive influenza seasons. This strategy allows for accrued evaluation of efficacy and safety in the context of the unpredictable nature of variability in influenza severity and vaccine effectiveness due to influenza's mutagenicity. Whereas in other trials subject recruitment can be accomplished during all months of a given year, recruitment for influenza vaccine studies is truncated to

just a few months, corresponding to the timing of seasonal vaccination. A passive recruitment approach of waiting to encounter potentially eligible participants is inadequate, as ideally participants are engaged prior to receipt of their standard-of-care influenza vaccine. This strategy requires identification of potential participants in the months prior to influenza vaccine becoming commercially available, which may be as early as August. At that time, patients may seek early vaccination in accordance with CDC and Health Canada recommendations to receive vaccine once it is available. However, this challenge also presents an opportunity to explore pragmatic approaches to participant recruitment, including use of a computable phenotype based on enrolment criteria and *International Classification of Diseases, 9th/10th Revisions*, codes to identify potential participants, and using electronic health record systems to invite potentially eligible participants to participate. Lastly, INVESTED has few exclusion criteria, coinciding with known safety of influenza vaccine in adults, allowing this to select patients highly representative of the intended cardiac population.

Many known or suspected respiratory virus infections have been associated with acute onset of MI and other cardiovascular events.<sup>2,4</sup> However, influenza A and influenza B have shown the most consistent association<sup>4</sup> and have a safe vaccine option for prevention, albeit incomplete and inconsistent year to year. In recent years, when vaccine effectiveness rates were reported at 10%–30%, notable disappointment resulted by public health officials and the lay public. However, evidence-based cardiovascular therapies offers comparable relative risk reductions for hard end points.<sup>32,33</sup> Thus, from the perspective of cardioprotection, we consider a 10%–30% risk reduction with influenza vaccination of high clinical value. As a safe and cost-effective intervention, vaccination is a worthwhile strategy for cardiopulmonary illness prevention.

A number of proposed mechanisms support a potential causal association between influenza infection and cardiovascular risk, either indirectly or directly. Indirect mechanisms include increased metabolic demand in the setting of influenza infection. When complemented by hypoxemia, influenza may exacerbate underlying cardiovascular disease because of increased sympathetic tone, potential volume overload, increased risk for plaque rupture, and arrhythmia.<sup>3</sup> Influenza infection predisposes patients to develop opportunistic infections such as pneumonia, which in itself is associated with increased cardiovascular events.<sup>34,35</sup> More directly, influenza infection has also been associated with myocardial depression,<sup>36</sup> which has been ascribed to an increase in proinflammatory cytokines,<sup>37,38</sup> and autopsy series have documented histologic evidence of myocardial injury, myocarditis, and myocyte necrosis following influenza-related deaths.<sup>39</sup> Moreover, influenza can stimulate a potent acute inflammatory response, which is a known trigger of acute plaque rupture. This mechanism is supported by observational data showing a temporal relationship between influenza infection spikes and myocardial infarctions.<sup>2,4</sup>

A potential limitation of the INVESTED trial is that we are not ascertaining symptoms of ILI nor are we pursuing confirmatory diagnoses of influenza infection. Symptoms of ILI have been temporally linked to influenza infection when measured in close proximity to the event. However, because we are ascertaining events at the end of influenza season, it could be months after the respiratory infection, in which case the recall bias for ILI would be substantial and unlikely to provide information relevant to the trial's hypothesis. As INVESTED is a large, simple trial, it is logistically difficult and costly to collect specimens from individuals with respiratory illnesses in real time to confirm and subtype influenza. To account for the effect of antigen match on vaccine effectiveness, we will interpret results in the context of the match between vaccine and annually changing circulating influenza strains by using prospectively collected influenza typing and subtyping data from the CDC and Public Health Agency of Canada. Another noteworthy challenge for this influenza vaccine trial is the use of a surrogate end point for vaccine effectiveness, which can dilute the impact of vaccine, particularly during seasons when activity of viruses other than influenza, such as

respiratory syncytial virus, parainfluenza virus, and human metapneumovirus, is high. INVESTED is comparing 2 vaccination strategies without a placebo control group; therefore, we cannot definitively determine the benefit of either strategy of influenza vaccination for cardioprotection over no vaccination. Although in the United States and Canada there is no longer equipoise to address this hypothesis in a randomized trial, our study will determine whether further cardioprotection can be realized from a more effective vaccine strategy, similar to rigorously tested intensive strategies of lipid-lowering therapy. Moreover, there are at least 2 ongoing international placebo-controlled trials testing the cardioprotective efficacy of standard influenza vaccination in patients with either MI (IAMI; NCT02831608)<sup>40</sup> or HF (RCT-IVVE; NCT02762851) with which we can indirectly compare results via network meta-analysis. Lastly, it is possible that differences between vaccine doses may vary based on the index enrollment event of MI or HF. We have prespecified examining results by enrollment subgroup; however, we are limited in power for the interaction test of index event by treatment; as such, any potential response differences will be interpreted with caution. It is also possible that the benefit of one vaccine strategy over another may be driven by pulmonary events, which are a component of our primary end point, over cardiac events.

In summary, INVESTED will examine whether high-dose compared with standard-dose influenza vaccine will reduce all-cause mortality and cardiopulmonary hospitalizations in high-risk cardiovascular patients who are particularly vulnerable to influenza and may derive inadequate immunity from standard-dose vaccination. INVESTED is the largest and longest study to assess whether vaccination is effective for secondary prevention in patients following recent presentation with HF or myocardial infarction.

## Funding

INVESTED is funded by the NHLBI (U01 HL130163 and U01 HL130204). Additional funds for site payments and vaccine are provided by Sanofi Pasteur, which has no scientific role in the INVESTED trial. Additional funds for the biological specimen acquisition are provided by grants from Heart and Stroke Foundation of Canada (ERLI024) and the Mount Sinai Hospital/University Health Network AMO Innovation Fund (21502002).

## Disclaimer

The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the NHLBI; the National Institutes of Health; the US Departments of Veterans Affairs, and Health and Human Services; or the US Government.

## References

- Reichert TA, Simonsen L, Sharma A, et al. Influenza and the winter increase in mortality in the United States, 1959–1999. *Am J Epidemiol* 2004;160(5):492–502.
- Smeeth L, Thomas SL, Hall AJ, et al. Risk of myocardial infarction and stroke after acute infection or vaccination. *N Engl J Med* 2004;351(25):2611–8.
- Corrales-Medina VF, Madjid M, Musher DM. Role of acute infection in triggering acute coronary syndromes. *Lancet Infect Dis* 2010;10(2):83–92.
- Kwong JC, Schwartz KL, Campitelli MA, et al. Acute myocardial infarction after laboratory-confirmed influenza infection. *N Engl J Med* 2018;378(4):345–53.
- Yancy CW, Jessup M, Bozkurt B, et al. ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Circulation* 2013;2013.
- Davis MM, Taubert K, Benin AL, et al. Influenza vaccination as secondary prevention for cardiovascular disease: a science advisory from the American Heart Association/American College of Cardiology. *Circulation* 2006;114(14):1549–53.
- Piepoli MF, Hoes AW, Agewall S, et al. 2016 European guidelines on cardiovascular disease prevention in clinical practice: the Sixth Joint Task Force of the European Society of Cardiology and other societies on cardiovascular disease prevention in clinical practice (constituted by representatives of 10 societies and by invited experts) developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). *Eur Heart J* 2016;37(29):2315–81.
- Nichol KL, Nordin J, Mullooly J, et al. Influenza vaccination and reduction in hospitalizations for cardiac disease and stroke among the elderly. *N Engl J Med* 2003;348(14):1322–32.
- de Diego C, Vila-Corcoles A, Ochoa O, et al. Effects of annual influenza vaccination on winter mortality in elderly people with chronic heart disease. *Eur Heart J* 2009;30(2):209–16.
- Udell JA, Zawi R, Bhatt DL, et al. Association between influenza vaccination and cardiovascular outcomes in high-risk patients: a meta-analysis. *JAMA* 2013;310(16):1711–20.
- Vardeny O, Sweitzer NK, Detry MA, et al. Decreased immune responses to influenza vaccination in patients with heart failure. *J Card Fail* 2009;15(4):368–73.
- Van Ermen A, Hermanson MP, Moran JM, et al. Double dose vs. standard dose influenza vaccination in patients with heart failure: a pilot study. *Eur J Heart Fail* 2013;15(5):560–4.
- DiazGranados CA, Dunning AJ, Kimmel M, et al. Efficacy of high-dose versus standard-dose influenza vaccine in older adults. *N Engl J Med* 2014;371(7):635–45.
- DiazGranados CA, Robertson CA, Talbot HK, et al. Prevention of serious events in adults 65 years of age or older: a comparison between high-dose and standard-dose inactivated influenza vaccines. *Vaccine* 2015;33(38):4988–93.
- Grohskopf LA, Sokolow LZ, Broder KR, et al. Prevention and control of seasonal influenza with vaccines: recommendations of the Advisory Committee on Immunization Practices—United States, 2017–18 influenza season. *MMWR Recomm Rep* 2017;66(2):1–20.
- Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep* 1966;50(3):163–70.
- Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *Biometrics* 1958;282(457–481).
- Cox DR. Regression models and life-tables. *J R Stat Soc* 1972;34(2):187–220.
- Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics* 2002;58(1):21–9.
- Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70:41–55.
- Xie J, Liu C. Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Stat Med* 2005;24(20):3089–110.
- Oakes D. In: Klein JP, Goel PK, eds. Frailty models for multiple event times. Netherlands: eSASotAK; 1992.
- Prentice RL, Cai J. Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika* 1992;79:495–512.
- Fine JP, Gray RJ. A proportional hazards model for the redistribution of a competing risk. *J Am Stat Assoc* 1999;94:496–509.
- Anderson PK, Gill RD. Cox's regression model for counting processes: a large sample study. *Ann Stat* 1982;10:1100–20.
- Prentice RL, Williams B, Peterson AV. On the regression analysis of multivariate failure time data. *Biometrika* 1981;68:373–9.
- Wei LJ, Lin DY. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J Am Stat Assoc* 1989;84:1065–73.
- Ghosh D, Lin DY. Nonparametric analysis of recurrent events and death. *Biometrics* 2000;56(2):554–62.
- Ghosh D, Lin DY. Semiparametric analysis of recurrent events data in the presence of dependent censoring. *Biometrics* 2003;59(4):877–85.
- O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics* 1979;35(3):549–56.
- Vardeny O, Claggett B, Udell JA, et al. Influenza vaccination in patients with chronic heart failure: the PARADIGM-HF trial. *JACC Heart Fail* 2016;4(2):152–8.
- Amsterdam EA, Wenger NK, Brindis RG, et al. 2014 AHA/ACC guideline for the management of patients with non-ST-elevation acute coronary syndromes: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation* 2014;130(25):e344–426.
- Stone NJ, Robinson JG, Lichtenstein AH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation* 2014;129(25 Suppl 2):S1–45.
- Yende S, D'Angelo G, Mayr F, et al. Elevated hemostasis markers after pneumonia increases one-year risk of all-cause and cardiovascular deaths. *PLoS One* 2011;6(8):e22847.
- Corrales-Medina VF, Alvarez KN, Weissfeld LA, et al. Association between hospitalization for pneumonia and subsequent risk of cardiovascular disease. *JAMA* 2015;313(3):264–74.
- Fagnou D, Pasquier P, Bodson L, et al. Myocardial dysfunction during H1N1 influenza infection. *J Crit Care* 2013;28(4):321–7.
- Takano T, Tajiri H, Kashiwagi Y, et al. Cytokine and chemokine response in children with the 2009 pandemic influenza A (H1N1) virus infection. *Eur J Clin Microbiol Infect Dis* 2011;30(1):117–20.
- Bermejo-Martin JF, Ortiz de Lejarazu R, Pumarola T, et al. Th1 and Th17 hypercytokinemia as early host response signature in severe pandemic influenza. *Crit Care* 2009;13(6):R201.
- Paddock CD, Liu L, Denison AM, et al. Myocardial injury and bacterial pneumonia contribute to the pathogenesis of fatal influenza B virus infection. *J Infect Dis* 2012;205(6):895–905.
- Frobert O, Gotberg M, Angeras O, et al. Design and rationale for the Influenza vaccination After Myocardial Infarction (IAMI) trial. A registry-based randomized clinical trial. *Am Heart J* 2017;189:94–102.

# Association of Initiation of Basal Insulin Analogs vs Neutral Protamine Hagedorn Insulin With Hypoglycemia-Related Emergency Department Visits or Hospital Admissions and With Glycemic Control in Patients With Type 2 Diabetes

Kasia J. Lipska, MD, MHS; Melissa M. Parker, MS; Howard H. Moffet, MPH; Elbert S. Huang, MD, MPH; Andrew J. Karter, PhD

**IMPORTANCE** In clinical trials of patients with type 2 diabetes, long-acting insulin analogs modestly reduced the risk of nocturnal hypoglycemia compared with human neutral protamine Hagedorn (NPH) insulin, but cost 2 to 10 times more. Outcomes in clinical practice may differ from trial results.

**OBJECTIVE** To compare the rates of hypoglycemia-related emergency department (ED) visits or hospital admissions associated with initiation of long-acting insulin analogs vs human NPH insulin in patients with type 2 diabetes.

**DESIGN, SETTING, AND PARTICIPANTS** A retrospective observational study using data from Kaiser Permanente of Northern California from January 1, 2006, through September 30, 2015. Patients with type 2 diabetes who initiated a long-acting insulin analog or NPH insulin were included and censored at death, loss of health plan coverage, change in insulin treatment, or study end on September 30, 2015.

**EXPOSURE** Initiation of basal insulin analogs (glargine or detemir) vs NPH insulin.

**MAIN OUTCOMES AND MEASURES** The primary outcome was the time to a hypoglycemia-related ED visit or hospital admission and the secondary outcome was the change in hemoglobin A<sub>1c</sub> level within 1 year of insulin initiation.

**RESULTS** There were 25 489 patients with type 2 diabetes who initiated basal insulin therapy (mean age, 60.2 [SD, 11.8] years; 51.9% white; 46.8% female). During a mean follow-up of 1.7 years, there were 39 hypoglycemia-related ED visits or hospital admissions among 1928 patients who initiated insulin analogs (11.9 events [95% CI, 8.1 to 15.6] per 1000 person-years) compared with 354 hypoglycemia-related ED visits or hospital admissions among 23 561 patients who initiated NPH insulin (8.8 events [95% CI, 7.9 to 9.8] per 1000 person-years) (between-group difference, 3.1 events [95% CI, -1.5 to 7.7] per 1000 person-years;  $P = .07$ ). Among 4428 patients matched by propensity score, the adjusted hazard ratio was 1.16 (95% CI, 0.71 to 1.78) for hypoglycemia-related ED visits or hospital admissions associated with insulin analog use. Within 1 year of insulin initiation, hemoglobin A<sub>1c</sub> level decreased from 9.4% (95% CI, 9.3% to 9.5%) to 8.2% (95% CI, 8.1% to 8.2%) after initiation of insulin analogs and from 9.4% (95% CI, 9.3% to 9.5%) to 7.9% (95% CI, 7.9% to 8.0%) after initiation of NPH insulin (adjusted difference-in-differences for glycemic control, -0.22% [95% CI, -0.09% to -0.37%]).

**CONCLUSIONS AND RELEVANCE** Among patients with type 2 diabetes, initiation of a basal insulin analog compared with NPH insulin was not associated with a reduced risk of hypoglycemia-related ED visits or hospital admissions or with improved glycemic control. These findings suggest that the use of basal insulin analogs in usual practice settings may not be associated with clinical advantages for these outcomes.

JAMA. 2018;320(1):53-62. doi:10.1001/jama.2018.7993  
Published online June 23, 2018.

← Editorial page 38

+ Author Audio Interview

**Author Affiliations:** Section of Endocrinology, Department of Internal Medicine, Yale University School of Medicine, New Haven, Connecticut (Lipska); Division of Research, Kaiser Permanente of Northern California, Oakland (Parker, Moffet, Karter); Department of Medicine, University of Chicago, Chicago, Illinois (Huang); Department of General Internal Medicine, University of California, San Francisco (Karter); Department of Epidemiology, University of Washington, Seattle (Karter); Department of Health Services, University of Washington, Seattle (Karter).

**Corresponding Author:** Kasia J. Lipska, MD, MHS, Yale School of Medicine, PO Box 208020, New Haven, CT 06520 (kasia.lipska@yale.edu).



Treatment of type 2 diabetes typically begins with lifestyle modification and initiation of metformin; however, 14% to 25% of patients eventually require initiation of insulin to reach recommended glycemic targets.<sup>1,2</sup> The mainstay of insulin treatment has long been human synthetic insulin; however, insulin analogs have become increasingly popular in clinical practice during the past decade.<sup>3,4</sup> Insulin analogs are molecularly altered forms of insulin that more closely mimic the pharmacokinetic profile of endogenous insulin.

In clinical trials, long-acting insulin analogs modestly reduce the risk of nocturnal hypoglycemia compared with human insulin, but have not been shown to reduce the risk of severe hypoglycemia or to improve glycemic control among patients with type 2 diabetes.<sup>5</sup> Discrepancies between trial results and outcomes in clinical practice are common and highlight the importance of gathering additional evidence from usual care settings.<sup>6</sup>

Although human insulin products are still used preferentially within Kaiser Permanente of Northern California (KPNC), prior work demonstrated widespread adoption of insulin analogs among US patients during the past 2 decades.<sup>3,4,7</sup> At the same time, the prices of insulin analogs have increased dramatically,<sup>8,9</sup> Medicaid payments for insulin have increased substantially,<sup>10</sup> and patients' out-of-pocket spending on insulin analogs has doubled.<sup>4</sup> In this setting, it is imperative to understand the differences in health outcomes associated with the use of the more expensive insulin analogs vs the more affordable human insulin products.

This study investigated the rates of hypoglycemia-related emergency department (ED) visits or hospital admissions and changes in levels of glycemic control after initiation of long-acting insulin analogs (glargine or detemir) compared with human neutral protamine Hagedorn (NPH) insulin among patients with type 2 diabetes in clinical practice.

## Methods

### Study Source

The institutional review boards of the Kaiser Foundation Research Institute and the University of Chicago approved the study. Participant informed consent was waived. A large, integrated health care delivery system, KPNC provides care for approximately 30% of the residents in the Northern California service area. The KPNC diabetes registry has been maintained since 1993. The registry now includes more than 350 000 adults with diabetes and is updated annually by identifying all health plan members with diabetes.

The identification of clinically recognized diabetes among health plan members is based on multiple sources of data including pharmacy use; laboratory results; and outpatient, emergency department, and hospitalization diagnoses of diabetes detailed further in a published algorithm.<sup>11</sup> Race/ethnicity was measured because prior studies suggest it is associated with both hypoglycemia and glycemic control.<sup>12,13</sup> Determination of race/ethnicity was based on self-reported race/ethnicity captured in the electronic

### Key Points

**Question** Is initiation of a basal insulin analog compared with human neutral protamine Hagedorn (NPH) insulin associated with a reduced risk of hypoglycemia-related emergency department (ED) visits or hospital admissions in patients with type 2 diabetes?

**Findings** In this retrospective observational study of 25 489 patients with type 2 diabetes, initiation of basal insulin analogs compared with NPH insulin was not associated with a significant difference in hypoglycemia-related ED visits or hospital admissions among a propensity-score matched cohort of 4428 patients (hazard ratio, 1.16).

**Meaning** Among patients with type 2 diabetes, the use of basal insulin analogs compared with NPH insulin was not associated with a reduced risk of hypoglycemia-related ED visits or hospital admissions.

medical record according to fixed categories. The study methods and a validation study of the KPNC diabetes registry (99% sensitivity for diabetes based on chart review registration) have been published.<sup>14</sup>

### Study Population

Using electronic medical records from KPNC, 49 190 adults (aged  $\geq 19$  years) with diabetes were identified. Each patient had full health plan and prescription coverage for 24 months prior to initiating insulin between January 1, 2006, and December 31, 2014. Patients with type 1 diabetes were excluded ( $n = 1838$ ) based on a validated algorithm that uses self-report or age of diabetes onset and drug treatment history to determine diabetes type.<sup>15</sup> Clinicians within KPNC can prescribe either NPH insulin or insulin analogs to patients with type 2 diabetes without obtaining prior approval; however, clinicians are encouraged to start with NPH insulin.

The analytic cohort consisted of patients who initiated basal insulin therapy and had no insulin prescription fills during the prior 12 months (Figure 1). Patients started with either NPH insulin or the insulin analog glargine or detemir. Patients using prandial insulin at baseline were excluded from the study. Patients who initiated prandial insulin during the study were censored at that time.

### Study Outcomes

The primary outcome was the time to hypoglycemia-related ED visit or hospital admission after initiation of insulin therapy based on a primary or principal discharge diagnosis of hypoglycemia using a validated algorithm (any of the following *International Classification of Diseases, Ninth Revision* codes: 251.0, 251.1, 251.2, 962.3, or 250.8 modified by 259.8, 272.7, 681, 682, 686.9, 707.1-707.9, 709.3, 730.0-730.2, or 731.8).<sup>16</sup>

The secondary outcome was the change in hemoglobin A<sub>1c</sub> level, which is a marker for the clinical effectiveness of insulin. For the baseline hemoglobin A<sub>1c</sub> level, the last measure during the 12 months prior to insulin initiation was used. The change from baseline to the last hemoglobin A<sub>1c</sub> level was

assessed prior to censoring and within 3 to 12 months after insulin initiation. A change in hemoglobin A<sub>1c</sub> level of 0.5% or greater is typically considered to be clinically significant.<sup>17</sup>

### Statistical Analysis

The analysis involved multiple steps. During the first step, a propensity score model was developed, predicting the binary outcome of initiating treatment with basal insulin analogs (compared with NPH insulin) using a flexible, data-adaptive model selection procedure called the deletion, substitution, and addition algorithm by Neugebauer and Bullard (available in R version 3.1.4; R Foundation for Statistical Computing).<sup>18</sup> The deletion, substitution, and addition procedure made use of training and test data sets to select the estimator with the lowest cross-validated risk among a list of candidate estimators developed via machine learning (ie, deletion, substitution, and addition of potential covariates as well as interactions and higher-order parameters).

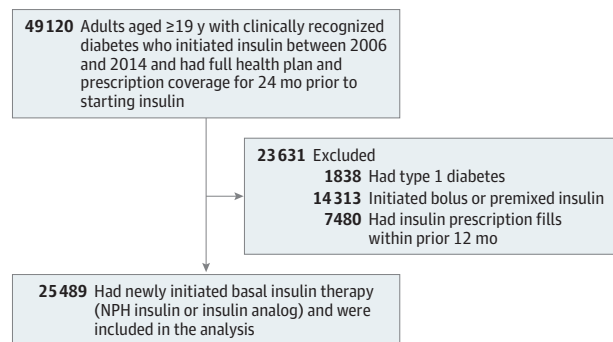
Potential covariates included: demographics, index year, clinical and comorbid characteristics, clinician specialty (primary care, endocrinology, or other specialty), KPNC service area, Charlson comorbidity index, chronic kidney disease stage, chronic liver disease, visual impairment, history of diabetic ketoacidosis, history of depression, glycemic control, the number of hypoglycemia-related ED visits or hospital admissions during the year prior to baseline, the number of ED visits or inpatient stays (for any reason) during the year prior to baseline, medication nonadherence (continuous measure of medication gaps<sup>19,20</sup>), outpatient medical visits (ie, the number of face-to-face visits with a clinician) during the 2 years prior to baseline, the patient co-pay for index insulin dispensed, and indicators of prevalent use for each of the diabetes therapeutic drug classes, statins, angiotensin-converting enzyme inhibitors, and  $\beta$ -blockers.

Missing data for continuous variables were imputed based on the within-group mean. Missing data for categorical variables were treated as a separate category. The C statistic (area under the receiver operating characteristic curve) for this model was 0.81, suggesting good discrimination.

During the second step, the predicted probability (ie, propensity score) of initiating treatment with long-acting insulin analogs was calculated for each patient. Quintiles of the propensity score were created based on the distribution of the propensity scores among the exposed patients (ie, patients who initiated insulin analogs). Using frequency matching (random sampling with replacement), 500 reference patients who initiated NPH insulin were selected from each of the quintiles defined by the exposed group.

This frequency matching created a population in which the distribution of covariates in the NPH insulin cohort was similar to those in the insulin analog cohort, thus minimizing observed confounders. Balance in the covariate distribution in each cohort was assessed by visually inspecting box plots and cumulative probability distributions of the propensity scores between exposed and reference patients and quantitatively through the calculation of the standardized difference, which compares the difference in means or prevalence of baseline covariates in units of the pooled SDs. A standardized difference

Figure 1. Derivation of the Study Cohort



Adults with type 2 diabetes and full health plan and prescription coverage were included if they began basal insulin therapy (neutral protamine Hagedorn [NPH] or insulin analog) between January 1, 2006, and December 31, 2014.

with the absolute value of less than or equal to 0.1 indicates a negligible difference in the mean or prevalence of a covariate between groups.<sup>21</sup>

During the third step, a survival analysis was conducted for the outcome of hypoglycemia-related ED visits or hospital admissions. This approach examined time to first event of hypoglycemia-related ED visit or hospital admission. Patients were censored at the earliest event: death, end of KPNC membership, end of prescription drug benefits, discontinuation of NPH insulin or long-acting insulin, addition of any other insulin subtype, or end of follow-up (September 30, 2015). The hazard ratios (HRs) and 95% CIs were calculated from the results of the Cox proportional hazards analyses on 1000 bootstrap samples with replacement, and were created using the methods described above.

The proportional hazard assumption was tested by assessing independence between the Schoenfeld residuals and follow-up time. The primary analysis included the HR after adjusting for baseline covariates that remained unbalanced after propensity score matching (ie, those with the absolute value of the standardized difference >0.1), as well as additional adjustments for prior hypoglycemia-related ED visits or hospital admissions and for time-dependent indicators of diabetes medication use. The use of sulfonylureas, metformin, or thiazolidinediones was based on dispensing of a given medication within 6 months prior to the start of insulin; thereafter, it was based on monthly fills and days' supply dispensed.

In a sensitivity analysis, the HR was additionally calculated using traditional regression adjustment for covariates that were significantly different at baseline for prior hypoglycemia-related ED visits or hospital admissions and for time-dependent indicators of diabetes medication use. Based on a post hoc estimate with a sample size of 25 489 patients, the study had 80% power to detect a HR of 2.1 or greater or of 0.5 or less for the outcome of hypoglycemia-related ED visits or hospital admissions associated with the initiation of insulin analogs vs NPH insulin.

During the fourth step, the change in hemoglobin A<sub>1c</sub> level following insulin initiation was estimated using a difference-in-differences approach. This approach measured the change

in glycemic control associated with the initiation of long-acting insulin analogs (first difference) after subtracting the background change (second difference [eg, due to secular trends]) among patients who initiated NPH insulin.<sup>22</sup> This model was based on the counterfactual assumption that if patients who initiated insulin analogs had instead initiated NPH insulin, their changes in hemoglobin A<sub>1c</sub> level would be similar to the changes observed in the NPH insulin reference group, who were frequency matched based on the propensity score quintile. The model was adjusted for baseline covariates that remained unbalanced after propensity score matching.

In the main secondary outcome analysis, participants with missing data for hemoglobin A<sub>1c</sub> level at baseline and those who were censored within 90 days of baseline were excluded. In a sensitivity analysis, patients also were excluded if the use of any class of diabetes medications changed from baseline until they were censored or until 12 months after initiation of insulin, whichever occurred first. The purpose of this analysis was to isolate the relationship between insulin initiation and change in hemoglobin A<sub>1c</sub> levels.

The difference-in-differences estimates and 95% CIs were calculated from the results of a least-squares regression analysis on 1000 bootstrap samples with replacement.<sup>23</sup> We used R version 3.3.1 and SAS version 9.3 (SAS Institute Inc) statistical software for all analyses. A *P* value < .05 was considered statistically significant and all testing was 2-sided.

## Results

### Patient Characteristics at Baseline

Between 2006 and 2014, a total of 25 489 patients with type 2 diabetes initiated basal insulin therapy (Table 1). The mean age was 60.2 years (SD, 11.8 years) and 46.8% were female. The racial/ethnic makeup of the cohort consisted of 51.9% who were white, 9.2% who were black, 17.6% who were Hispanic, and 15.3% who were Asian. The Charlson comorbidity index value was 0 among 28.1%, 1 among 28.5%, 2 among 11.3%, and 3 or greater among 32.1%.

In this cohort, data were missing for race/ethnicity (n = 280), chronic kidney disease stage (n = 213), duration of diabetes (n = 6641), age at diabetes onset (n = 6641), body mass index (n = 1429), elevated serum creatinine level (n = 33), neighborhood deprivation index (n = 242), hemoglobin A<sub>1c</sub> level (n = 402), KPNC service area (n = 61), and medication non-adherence (n = 5474).

Among the patients who initiated insulin, 23 561 (92%) started with NPH insulin and 1928 (8%) started with insulin analogs. Patients who initiated insulin analogs were more likely to have a greater number of comorbid conditions and had more ED or hospital use events (for any cause) within the prior year, but the magnitude of the differences was small (Table 1). One substantive difference was that the median co-payments for insulin analogs (\$20) were significantly higher than for NPH insulin (\$10). The mean baseline hemoglobin A<sub>1c</sub> levels for the 2 groups were 9.41% [SD, 2.0%] among patients who started insulin analogs and 9.40% [SD, 1.8%] among patients who started NPH insulin.

In the propensity score-matched cohort (n = 4428), the differences in the characteristics of patients who initiated insulin analog vs NPH insulin were minimized; however, statistical differences persisted for outpatient medical visits, KPNC service area, and year of index prescription. These differences were not substantive.

### Primary Outcome

Among patients who initiated insulin analogs (n = 1928; 3289.8 person-years), there were 32 ED visits and 7 hospital admissions related to hypoglycemia (11.9 events [95% CI, 8.1 to 15.6] per 1000 person-years) during a mean follow-up of 1.71 years (95% CI, 1.62 to 1.79) and a median follow-up of 1.03 years (interquartile range, 0.36 to 2.37). Among patients who initiated NPH insulin (n = 23 561; 40 060.0 person-years), there were 309 ED visits and 45 hospital admissions related to hypoglycemia (8.8 events [95% CI, 7.9 to 9.8] per 1000 person-years) during a mean follow-up of 1.70 years (95% CI, 1.68 to 1.72) and a median follow-up of 1.09 years (interquartile range, 0.41 to 2.38). The between-group difference was 3.1 events (95% CI, -1.5 to 7.7) per 1000 person-years (*P* = .07).

The Kaplan-Meier curve appears in Figure 2. Among all censoring events, 2.8% were due to death, 31.9% were due to discontinuation of insulin, and 31.6% were due to initiation of an additional type of insulin. The proportional hazard assumption was met because the Schoenfeld residuals for the exposure were independent of time (Pearson correlation coefficient, 0.06; *P* = .20).

After frequency matching the patients who initiated insulin analogs with those who initiated NPH insulin, and after additional adjustment for unbalanced covariates, prior hypoglycemia-related ED visits or hospital admissions, and time-dependent indicators of diabetes medication use, there was no significant difference in hypoglycemia-related ED visits or hospital admissions (HR, 1.16 [95% CI, 0.71 to 1.78]; Table 2).

### Secondary Outcome

In the main secondary outcome analysis of change in glycemic control, participants with missing data for hemoglobin A<sub>1c</sub> level at baseline (n = 402) and those who were censored within 90 days of baseline (n = 3665) were excluded (n = 4067). Within 1 year of initiation of insulin analogs, hemoglobin A<sub>1c</sub> level decreased by 1.26 percentage points (95% CI, 1.16 to 1.36 percentage points) from 9.41% (95% CI, 9.34% to 9.50%) to 8.16% (95% CI, 8.09% to 8.24%).

Within 1 year of initiation of NPH insulin, hemoglobin A<sub>1c</sub> level decreased by 1.48 percentage points (95% CI, 1.39 to 1.57 percentage points) from 9.39% (95% CI, 9.32% to 9.47%) to 7.92% (95% CI, 7.85% to 7.99%). Between the baseline and post-baseline measures, the mean number of days was 298 (SD, 103 days) among patients who initiated insulin analogs and 288 days (SD, 98 days) among patients who initiated NPH (standardized difference, 0.10). After adjustment, the difference-in-differences for glycemic control was -0.22% (95% CI, -0.09% to -0.37%), indicating that the use of NPH insulin was associated with a statistically significant greater decrease in hemoglobin A<sub>1c</sub> level (Table 3). However, this difference is not considered clinically significant.<sup>17</sup>

Table 1. Baseline Characteristics of 25 489 Patients With Type 2 Diabetes

| Characteristic   | Insulin Analog<br>(n = 1928) | Before Frequency Matching   |   | After Frequency Matching <sup>a</sup>  |   |
|--|------------------------------|-----------------------------|---|--|---|
|  |                              | NPH Insulin<br>(n = 23 561) | Standardized<br>Difference <sup>b</sup> | NPH Insulin<br>(n = 2500) <sup>c</sup> | Standardized<br>Difference <sup>b</sup> |
| Age, mean (SD), y  | 60.6 (12.8)                  | 60.2 (11.8)                 | 0.04                                    | 60.8 (11.8)                            | -0.01                                   |
| Female sex, No. (%)  | 912 (47)                     | 11 105 (47)                 | 0.01                                    | 1140 (46)                              | 0.03                                    |
| Race/ethnicity, No. (%) <sup>d</sup>   |                              |                             |   |  |   |
| Asian  | 332 (17)                     | 3534 (15)                   | 0.06                                    | 383 (15)                               | 0.05                                    |
| Black  | 214 (11)                     | 2109 (9)                    | 0.07                                    | 231 (9)                                | 0.06                                    |
| White  | 957 (50)                     | 12 136 (52)                 | -0.04                                   | 1265 (51)                              | -0.02                                   |
| Hispanic   | 293 (15)                     | 4130 (18)                   | -0.06                                   | 446 (18)                               | -0.07                                   |
| Other  | 114 (6)                      | 1390 (6)                    | -0.04                                   | 133 (5)                                | -0.04                                   |
| Neighborhood deprivation index by quartile, No. (%) <sup>d,e</sup>   |                              |                             |   |  |   |
| First (least deprived)   | 374 (20)                     | 4643 (20)                   | -0.01                                   | 486 (19)                               | -0.002                                  |
| Second   | 538 (28)                     | 6695 (29)                   | -0.01                                   | 702 (28)                               | -0.004                                  |
| Third  | 572 (30)                     | 7030 (30)                   | -0.004                                  | 760 (30)                               | -0.02                                   |
| Fourth (most deprived)   | 423 (22)                     | 4972 (21)                   | 0.02                                    | 532 (21)                               | 0.02                                    |
| Comorbidities, No. (%)   |                              |                             |   |  |   |
| Charlson comorbidity index <sup>f</sup>  |                              |                             |   |  |   |
| 0  | 501 (26)                     | 6654 (28)                   | -0.05                                   | 690 (28)                               | -0.04                                   |
| 1  | 533 (28)                     | 6736 (29)                   | -0.02                                   | 735 (29)                               | -0.04                                   |
| 2  | 228 (12)                     | 2652 (11)                   | 0.02                                    | 256 (10)                               | 0.05                                    |
| ≥3   | 666 (35)                     | 7519 (32)                   | 0.06                                    | 819 (33)                               | 0.04                                    |
| Chronic kidney disease stage <sup>d</sup>  |                              |                             |   |  |   |
| 0  | 202 (11)                     | 3121 (13)                   | -0.09                                   | 337 (14)                               | -0.09                                   |
| 1  | 468 (25)                     | 6024 (26)                   | -0.03                                   | 597 (24)                               | 0.01                                    |
| 2  | 656 (35)                     | 8348 (36)                   | -0.03                                   | 883 (35)                               | -0.03                                   |
| 3A   | 297 (15)                     | 3064 (13)                   | 0.04                                    | 316 (13)                               | 0.05                                    |
| 3B   | 179 (9)                      | 2088 (9)                    | 0.01                                    | 237 (9)                                | -0.01                                   |
| 4  | 77 (4)                       | 627 (3)                     | 0.07                                    | 82 (3)                                 | 0.04                                    |
| 5 or dialysis  | 28 (1)                       | 115 (1)                     | 0.10                                    | 19 (1)                                 | 0.07                                    |
| Elevated serum creatinine level, No <sup>d,g</sup>   | 266 (14)                     | 2664 (11)                   | 0.08                                    | 334 (13)                               | 0.01                                    |
| Chronic liver disease  | 103 (5)                      | 1392 (6)                    | -0.02                                   | 141 (6)                                | -0.01                                   |
| Depression   | 395 (20)                     | 5266 (22)                   | -0.05                                   | 527 (21)                               | -0.01                                   |
| Visual impairment or blindness   | 95 (5)                       | 618 (3)                     | 0.12                                    | 93 (4)                                 | 0.06                                    |
| Health Care Use, No. (%)   |                              |                             |   |  |   |
| Emergency department visit for any cause in prior year   | 649 (34)                     | 6822 (29)                   | 0.10                                    | 780 (31)                               | 0.05                                    |
| Inpatient hospitalization for any cause in prior year  | 379 (20)                     | 3069 (13)                   | 0.18                                    | 421 (17)                               | 0.07                                    |
| No. of outpatient medical visits in prior 2 y by quartile  |                              |                             |   |  |   |
| 0-6  | 423 (22)                     | 5931 (25)                   | -0.08                                   | 613 (25)                               | -0.06                                   |
| 7-11   | 435 (23)                     | 6148 (26)                   | -0.08                                   | 609 (24)                               | -0.04                                   |
| 12-19  | 480 (25)                     | 5769 (24)                   | 0.01                                    | 631 (25)                               | -0.01                                   |
| ≥20  | 590 (31)                     | 5713 (24)                   | -0.14                                   | 647 (26)                               | 0.11                                    |
| Diabetic ketoacidosis in prior year  | 31 (2)                       | 206 (1)                     | 0.07                                    | 46 (2)                                 | -0.02                                   |
| Emergency department or inpatient hospitalization for hypoglycemia within prior year                       | 16 (1)                       | 115 (1)                     | 0.04                                    | 22 (1)                                 | -0.01                                   |
| No. of hypoglycemic events resulting in emergency department or inpatient stay in prior year, median (IQR) | 0 (0 to 2)                   | 0 (0 to 3)                  | 0.04                                    | 0 (0 to 3)                             | -0.0002                                 |
| Kaiser Permanente of Northern California service area <sup>d</sup>   |                              |                             |   |  |   |
| A  | 114 (6)                      | 1989 (8)                    | -0.10                                   | 196 (8)                                | -0.08                                   |
| B  | 209 (11)                     | 2392 (10)                   | 0.02                                    | 224 (9)                                | 0.06                                    |
| C  | 123 (6)                      | 1631 (7)                    | -0.02                                   | 121 (5)                                | 0.07                                    |
| D  | 144 (7)                      | 740 (3)                     | 0.19                                    | 172 (7)                                | 0.02                                    |
| E  | 128 (7)                      | 1870 (8)                    | -0.05                                   | 164 (7)                                | 0.004                                   |
| F  | 71 (4)                       | 1513 (6)                    | -0.13                                   | 123 (5)                                | -0.06                                   |
| G  | 253 (13)                     | 2080 (9)                    | -0.14                                   | 272 (11)                               | 0.07                                    |
| H  | 139 (7)                      | 3810 (16)                   | -0.28                                   | 244 (10)                               | -0.09                                   |
| I  | 97 (5)                       | 1981 (8)                    | -0.14                                   | 166 (7)                                | -0.07                                   |
| J  | 143 (7)                      | 581 (2)                     | 0.23                                    | 135 (5)                                | 0.08                                    |
| K  | 65 (3)                       | 1831 (8)                    | -0.19                                   | 175 (7)                                | -0.16                                   |
| L  | 139 (7)                      | 1600 (7)                    | 0.02                                    | 129 (5)                                | 0.09                                    |
| M  | 272 (14)                     | 1513 (6)                    | 0.26                                    | 354 (14)                               | -0.001                                  |

(continued)



Table 1. Baseline Characteristics of 25 489 Patients With Type 2 Diabetes (continued)

| Characteristic   | Insulin Analog<br>(n = 1928) | Before Frequency Matching   |   | After Frequency Matching <sup>a</sup>  |   |
|--|------------------------------|-----------------------------|---|--|---|
|  |                              | NPH Insulin<br>(n = 23 561) | Standardized<br>Difference <sup>b</sup> | NPH Insulin<br>(n = 2500) <sup>c</sup> | Standardized<br>Difference <sup>b</sup> |
| Prescribing clinician specialty                              |                              |                             |   |  |   |
| Primary care   | 1631 (85)                    | 21 595 (92)                 | -0.22                                   | 2120 (85)                              | -0.002                                  |
| Endocrinologist  | 74 (4)                       | 667 (3)                     | 0.05                                    | 90 (4)                                 | 0.01                                    |
| Other specialist   | 223 (12)                     | 1299 (6)                    | 0.22                                    | 290 (12)                               | -0.01                                   |
| Clinical Characteristics of Diabetes                         |                              |                             |   |  |   |
| Duration of diabetes, mean (SD), y <sup>d</sup>              | 11.6 (7.9)                   | 10.6 (6.4)                  | 0.18                                    | 11.7 (7.4)                             | -0.01                                   |
| Age at diabetes onset, mean (SD), y <sup>d</sup>             | 49.2 (11.2)                  | 50.0 (10.8)                 | -0.08                                   | 49.0 (9.3)                             | 0.03                                    |
| Body mass index, mean (SD) <sup>d</sup>                      | 32.2 (7.5)                   | 33.3 (7.5)                  | -0.15                                   | 32.7 (7.2)                             | -0.06                                   |
| Hemoglobin A <sub>1c</sub> level, mean (SD), % <sup>d</sup>  | 9.41 (2.0)                   | 9.40 (1.8)                  | 0.01                                    | 9.39 (1.8)                             | 0.02                                    |
| Type of diabetes medication, No. (%)                         |                              |                             |   |  |   |
| None   | 166 (9)                      | 1142 (5)                    | 0.15                                    | 172 (7)                                | 0.06                                    |
| Metformin  | 1330 (69)                    | 17 915 (76)                 | -0.16                                   | 1805 (72)                              | -0.07                                   |
| Sulfonylurea   | 1590 (82)                    | 20 648 (88)                 | -0.15                                   | 2142 (86)                              | -0.09                                   |
| Thiazolidinedione  | 540 (28)                     | 5533 (23)                   | 0.10                                    | 668 (27)                               | 0.03                                    |
| Dipeptidyl peptidase 4 inhibitors                            | 38 (2)                       | 248 (1)                     | 0.08                                    | 40 (2)                                 | 0.03                                    |
| Glucagon-like peptide 1 receptor agonists                    | 23 (1)                       | 71 (<1)                     | 0.10                                    | 12 (1)                                 | 0.07                                    |
| Other <sup>h</sup>   | 54 (3)                       | 322 (1)                     | 0.10                                    | 44 (2)                                 | 0.07                                    |
| Types of cardiometabolic medications, No. (%)                |                              |                             |   |  |   |
| Statins  | 1409 (73)                    | 18 553 (79)                 | -0.13                                   | 1922 (77)                              | -0.09                                   |
| Angiotensin-converting enzyme inhibitors                     | 912 (47)                     | 11 185 (47)                 | -0.003                                  | 1192 (48)                              | -0.01                                   |
| β-Blockers   | 828 (43)                     | 9951 (42)                   | 0.01                                    | 1056 (42)                              | 0.01                                    |
| Medication nonadherence, % <sup>d,i</sup>                    |                              |                             |   |  |   |
|  | 432 (22)                     | 5473 (23)                   | 0.15                                    | 591 (24)                               | -0.03                                   |
| Year of index insulin prescription, No. (%)                  |                              |                             |   |  |   |
| 2006   | 289 (15)                     | 1683 (7)                    | 0.25                                    | 313 (13)                               | 0.07                                    |
| 2007   | 310 (16)                     | 3277 (14)                   | 0.06                                    | 354 (14)                               | 0.06                                    |
| 2008   | 280 (15)                     | 2357 (10)                   | 0.14                                    | 373 (15)                               | -0.01                                   |
| 2009   | 243 (13)                     | 1947 (8)                    | 0.14                                    | 294 (12)                               | 0.03                                    |
| 2010   | 104 (5)                      | 2072 (9)                    | -0.13                                   | 176 (7)                                | -0.07                                   |
| 2011   | 169 (9)                      | 2667 (11)                   | -0.09                                   | 212 (8)                                | 0.01                                    |
| 2012   | 211 (11)                     | 3120 (13)                   | -0.07                                   | 289 (12)                               | -0.02                                   |
| 2013   | 214 (11)                     | 3227 (14)                   | -0.08                                   | 247 (10)                               | 0.04                                    |
| 2014   | 108 (6)                      | 3211 (14)                   | -0.27                                   | 242 (10)                               | -0.15                                   |
| Patient co-pay for index insulin dispensed, median (IQR), \$ | 20 (10 to 35)                | 10 (5 to 10)                | 0.67                                    | 15 (10 to 45)                          | 0.05                                    |

Abbreviations: IQR, interquartile range; NPH, neutral protamine Hagedorn.

<sup>a</sup> Patients who initiated NPH insulin were frequency matched with patients initiating insulin analogs based on propensity score quintile.

<sup>b</sup> Compares characteristics for patients who initiated insulin analogs vs NPH insulin. An absolute value  $\leq 0.1$  indicates a negligible difference in the mean or prevalence of a covariate between groups.<sup>21</sup>

<sup>c</sup> The number of patients in each category for the bootstrapped analysis (1000 samples of 2500 each) calculated based on the distributions in the 2.5 million observations and then applied to the 2500. Numbers reflect average across the 1000 samples.

<sup>d</sup> Missing data: race/ethnicity (n = 280), neighborhood deprivation index (n = 242), chronic kidney disease stage (n = 213), elevated serum creatinine (n = 33), Kaiser Permanente of Northern California service area (n = 61), duration of diabetes (n = 6641), age at diabetes onset (n = 6641), body mass index (n = 1429), hemoglobin A<sub>1c</sub> (n = 402), and medication nonadherence (n = 5474).

<sup>e</sup> Created by a principal components analysis of 8 census-derived variables at the census tract level (% of men in management and professional occupations, living in crowded housing, households in poverty, female-headed households with dependents, households receiving public assistance, households earning <\$30 000/year, individuals with less than a high school education, and unemployment).<sup>24,25</sup> Negative scores = less deprivation.

<sup>f</sup> Based on the modified version of the Deyo Charlson Score.<sup>26</sup> Possible scores ranged from 0-17 and represent the number of selected comorbid conditions (other than diabetes) during the 2 years prior to baseline.

<sup>g</sup> Defined as >1.5 mg/dL in men and >1.4 mg/dL in women.

<sup>h</sup> Including α-glucosidase inhibitors, amylin, and meglitinides.

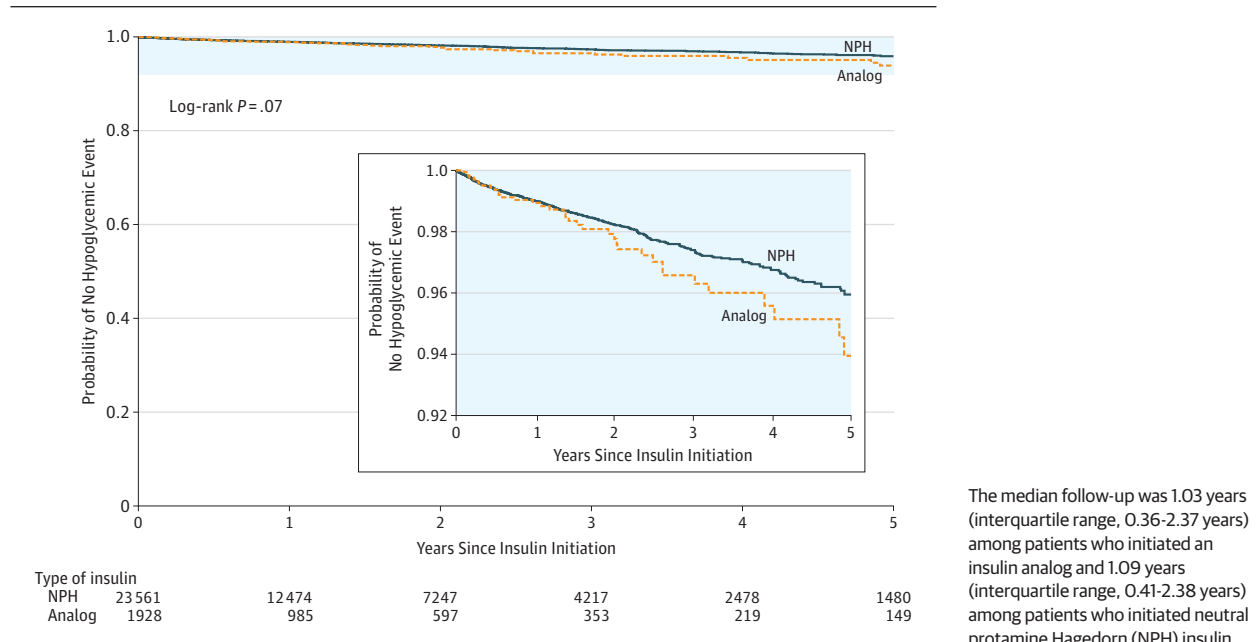
<sup>i</sup> Based on the continuous measure of medication gaps.<sup>19,20</sup> Possible values ranged from 0-100 and represent the percentage of time a patient did not have sufficient medication supply based on electronic pharmacy dispensing data. Medication nonadherence was defined as lacking medication >20% of the time within the year prior to baseline.

### Sensitivity Analyses

For hypoglycemia-related ED visits or hospital admissions after initiation of insulin analogs vs NPH insulin, the HR was 1.22 (95% CI, 0.86 to 1.75) using traditional regression adjustment for covariates that were significantly different at baseline, for prior hypoglycemia-related ED visits or hospital admissions, and for time-dependent indicators of diabetes medication use (Table 2).

Within 1 year of insulin initiation, hemoglobin A<sub>1c</sub> levels decreased by 1.27 percentage points (95% CI, 1.14 to 1.39 percentage points) among patients who started insulin analogs and by 1.43 percentage points (95% CI, 1.33 to 1.53 percentage points) among patients who started NPH insulin after excluding patients with changes in the use of diabetes medications (n = 5199). The difference-in-differences in hemoglobin A<sub>1c</sub>

**Figure 2. Time to First Hypoglycemia-Related Emergency Department Visit or Hospital Admission Among Patients With Type 2 Diabetes**



**Table 2. Survival Analysis**

|  | Hazard Ratio (95% CI) <sup>a</sup> |
|--|------------------------------------|
| <b>Full Cohort (N = 25 489)</b>  |                                    |
| Unadjusted   | 1.35 (0.97-1.88)                   |
| Model 1: adjusts for prior severe hypoglycemia-related ED visits or hospital admissions  | 1.33 (0.95-1.84)                   |
| Model 2: adjusts for prior severe hypoglycemia-related ED visits or hospital admissions and time-dependent indicators for oral diabetes therapy use  | 1.31 (0.94-1.82)                   |
| Model 3: adjusts for prior severe hypoglycemia-related ED visits or hospital admissions, time-dependent indicators for oral diabetes therapy use, and additional unbalanced baseline covariates <sup>b</sup> | 1.22 (0.86-1.75)                   |
| <b>Propensity Score-Matched Sample (n = 4428)</b>  |                                    |
| Unadjusted   | 1.17 (0.73-1.75)                   |
| Model 1: adjusts for prior severe hypoglycemia-related ED visits or hospital admissions  | 1.18 (0.74-1.78)                   |
| Model 2: adjusts for prior severe hypoglycemia-related ED visits or hospital admissions and time-dependent indicators for oral diabetes therapy use  | 1.21 (0.75-1.84)                   |
| Model 3: adjusts for prior severe hypoglycemia-related ED visits or hospital admissions, time-dependent indicators for oral diabetes therapy use, and additional unbalanced baseline covariates <sup>c</sup> | 1.16 (0.71-1.78)                   |

Abbreviation: ED, emergency department.

<sup>a</sup> The hazard ratios in the full cohort used traditional regression adjustment. The hazard ratios in the frequency-matched sample used 1000 bootstrap regressions. Risk conferred by initiating long-acting insulin analog vs neutral protamine Hagedorn insulin. Hazard ratio >1 favors NP insulin.

<sup>b</sup> The covariates were baseline diabetes treatment regimen, statin use, visual

impairment, hospital use, outpatient medical visits, duration of diabetes, body mass index, year of index prescription, patient insulin co-pay, Kaiser Permanente of Northern California (KPNC) service area, prescribing clinician specialty, and medication nonadherence.

<sup>c</sup> The covariates were outpatient medical visits, KPNC service area, and year of index prescription.

levels was  $-0.16\%$  (95% CI,  $-0.005\%$  to  $0.34\%$ ), which was not a clinically significant change.

## Discussion

In this observational study of patients with type 2 diabetes in a large integrated health care system, initiation of basal insulin analogs compared with NPH insulin was not associated with a lower rate of ED visits or hospital admissions related to hypoglycemia. Moreover, initiation of NPH insulin was associ-

ated with a slightly greater, but not clinically meaningful, decline in hemoglobin A<sub>1c</sub> level from baseline. These results suggest that the use of basal insulin analogs among patients with type 2 diabetes in usual practice settings may not be associated with clinical advantages with respect to these outcomes compared with NPH insulin.

Randomized clinical trials suggest that long-acting insulin analogs may offer some advantages for patients with type 2 diabetes, but the benefits appear relatively modest.<sup>5</sup> In several meta-analyses,<sup>5,27,28</sup> benefits in terms of reduced hypoglycemia risk have been reported for both types of insulin

Table 3. Changes in Glycemic Control

|  | Hemoglobin A <sub>1c</sub> Level, Mean (95% CI), % <sup>a</sup> |                     |
|--|---|---------------------|
|  | Insulin Analog  | NPH Insulin         |
| Baseline <sup>b</sup>  | 9.41 (9.34 to 9.50)   | 9.39 (9.32 to 9.47) |
| Postbaseline <sup>c</sup>  | 8.16 (8.09 to 8.24)   | 7.92 (7.85 to 7.99) |
| No. of days between baseline and postbaseline measure, mean (SD) | 298 (103)   | 288 (98)            |
| Pre-post change  | 1.26 (1.16 to 1.36)   | 1.48 (1.39 to 1.57) |
| Unadjusted difference-in-differences estimate <sup>d</sup>       | -0.22 (-0.09 to -0.37)  |                     |
| Adjusted difference-in-differences estimate <sup>e</sup>         | -0.22 (-0.09 to -0.37)  |                     |

Abbreviation: NPH, neutral protamine Hagedorn.

<sup>a</sup> The results are from 1000 bootstrap samples of the frequency-matched cohort representing 21 422 of the 24 489 patients in the cohort who met the inclusion criteria for the hemoglobin A<sub>1c</sub> analysis (prebaseline hemoglobin A<sub>1c</sub> level collected within the 12 months prior to insulin initiation and at least 90 days of follow-up after insulin initiation).

<sup>b</sup> Indicates the last measure within the 12 months prior to insulin initiation.

<sup>c</sup> The last hemoglobin A<sub>1c</sub> level was assessed prior to censoring and within 3 to 12 months after insulin initiation.

<sup>d</sup> Calculated as the pre-post change in hemoglobin A<sub>1c</sub> level for the insulin analog group minus the pre-post change in hemoglobin A<sub>1c</sub> level for the NPH insulin group.

<sup>e</sup> Adjusted for outpatient medical visits, Kaiser Permanente of Northern California service area, and year of index prescription.

analog (glargine and detemir) compared with NPH insulin, but only with respect to nocturnal hypoglycemia. Benefits with respect to severe hypoglycemia (defined as requiring assistance from another person to administer carbohydrates, glucagon, or other resuscitative treatment) were not significant.

One meta-analysis<sup>29</sup> found that glargine use in combination with oral agents was associated with a higher probability of reaching the target hemoglobin A<sub>1c</sub> level without nocturnal hypoglycemia compared with the use of NPH insulin in combination with oral agents. This difference was largely driven by a reduction in nocturnal hypoglycemia with the use of glargine. In a patient-level meta-analysis,<sup>30</sup> the incidence of overall and nocturnal hypoglycemia was modestly lower in patients with type 2 diabetes treated with glargine compared with NPH insulin. Differences in glycemic control between NPH insulin and basal insulin analogs in the trials were minimal and there were not consistently greater decreases in hemoglobin A<sub>1c</sub> levels associated with insulin analog use in patients with type 2 diabetes.<sup>30</sup>

In 1 meta-analysis,<sup>28</sup> compared with NPH insulin, the difference in hemoglobin A<sub>1c</sub> level was -0.05% (95% CI, -0.13% to 0.04%) for insulin glargine and 0.13% (95% CI, 0.03% to 0.22%) for insulin detemir. Because allocation to insulin analogs vs NPH insulin was not concealed in most trials, the potential for ascertainment bias exists, especially for subjective outcomes such as patient-reported hypoglycemia. In addition, clinical trials typically include specific algorithms to achieve strict glycemic control targets and, as a result, trial participants achieve tighter glycemic control compared with patients encountered in clinical practice.

Prior observational studies have shown reduced hypoglycemia risk and improved glycemic control with the use of insulin analogs compared with NPH insulin, but they did not adequately account for confounding factors. For example, 2 large observational studies<sup>31,32</sup> using national registries in Finland showed a significantly increased risk of hospitalization related to severe hypoglycemia with the use of NPH insulin compared with either the insulin analog detemir or glargine, but the studies lacked information about hemoglobin A<sub>1c</sub> level and major risk factors for hypoglycemia.

In a related study also conducted in Finland, NPH insulin use was associated with increased mortality compared with basal insulin analogs, but again, the study did not adjust for important confounders.<sup>33</sup> Similarly, other studies did not use the more rigorous techniques for balancing covariates, such as propensity score matching.<sup>32,34</sup>

One study from the US Department of Veterans Affairs compared insulin analogs with NPH insulin using clinician practice pattern as an instrumental variable to address confounding by indication.<sup>35</sup> This study examined hospitalizations for ambulatory care-sensitive conditions and mortality, and found no consistent difference in these outcomes when comparing use of long-acting insulin analogs and NPH insulin.<sup>35</sup>

The findings of the present study support the use of NPH insulin in many patients with type 2 diabetes to reduce the costs of care. Insulin prices in the United States have increased 3-fold between 2002 and 2013, particularly for insulin analogs.<sup>8</sup> In 2013, the estimated per-patient expenditures on insulin were greater than for all other diabetes medications combined.<sup>8</sup> A prior study found an increase in the use of insulin analogs for the management of type 2 diabetes with a resultant increase in inflation-adjusted out-of-pocket costs.<sup>4</sup>

The rising cost of insulin may directly affect the health outcomes of patients with diabetes because the associated increased cost share is known to contribute to nonadherence.<sup>36-39</sup> In contrast to insulin analogs, NPH insulin can be purchased for as little as \$25 per vial, about one-tenth the price of either insulin analog glargine or detemir.<sup>40</sup> It is likely that only select patients with type 2 diabetes benefit from insulin analogs vs human insulin preparations. To contain health care costs, decisions to use more expensive insulin should be made by informed patients and clinicians, and driven by convincing data about the benefits, harms, and tradeoffs.

### Limitations

This study has several limitations. First, this was an observational study and is thus subject to confounding. Specifically, patients with type 2 diabetes who initiated insulin analogs may

be different from those who started NPH insulin. Despite matching on propensity score quintiles, some substantive differences between the 2 groups remained (standardized difference >0.1).

However, the results did not change after additional adjustment for these unbalanced factors. Nonetheless, it is possible that the study did not completely account for confounding by indication due to unmeasured or missing confounders.

Second, the primary outcome was based on ED or hospital use related to hypoglycemia. Therefore, differences in nocturnal and self-reported hypoglycemia, or adverse events treated by emergency medical services but not transported to the ED could not be examined.

Third, the 95% CIs ranged from 0.71 to 1.78 for the HR for hypoglycemia-related ED visits or hospital admissions, which may include a clinically important difference. Therefore, this study may have been underpowered to detect a benefit or harm of that magnitude.

Fourth, these findings come from an integrated health care delivery system and may not necessarily be generalizable to other types of health care settings.

Fifth, the comparisons between NPH insulin and basal insulin analogs did not include convenience, number of injections required, or mode of delivery (vial vs pen). It is possible that basal insulin analogs may confer these and other advantages to patients with type 2 diabetes.

## Conclusions

Among patients with type 2 diabetes, initiation of a basal insulin analog compared with NPH insulin was not associated with a reduced risk of hypoglycemia-related ED visits or hospital admissions or with improved glycemic control. These findings suggest that the use of basal insulin analogs in usual practice settings may not be associated with clinical advantages for these outcomes.

### ARTICLE INFORMATION

**Accepted for Publication:** May 22, 2018.

**Published Online:** June 23, 2018.

doi:10.1001/jama.2018.7993

**Author Contributions:** Dr Karter had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

**Concept and design:** Lipska, Parker, Karter.

**Acquisition, analysis, or interpretation of data:** All authors.

**Drafting of the manuscript:** Lipska.

**Critical revision of the manuscript for important intellectual content:** All authors.

**Statistical analysis:** Parker.

**Obtained funding:** Karter.

**Administrative, technical, or material support:** Moffet.

**Supervision:** Huang, Karter.

**Conflict of Interest Disclosures:** The authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest.

Dr Lipska reported receiving support from the Centers for Medicare & Medicaid Services to develop and maintain publicly reported quality measures. Mr Moffet reported receiving grant funding from Eli Lilly, AstraZeneca, and Regeneron/Sanofi. Dr Karter reported receiving grant funding from AstraZeneca. No other disclosures were reported.

**Funding/Support:** This project was supported by grants R01DK103721 and R01DK081796 from the National Institutes of Health. Dr Lipska receives support from the National Institute on Aging and the American Federation of Aging Research through Paul Beeson Career Development award K23AG048359 and award P30AGO21342 from the Yale Claude D. Pepper Older Americans Independence Center. Drs Huang and Karter are supported by awards P30 DK092924 and P30 DK092949, respectively, from the Centers for Diabetes Translational Research within the National Institute of Diabetes and Digestive and Kidney Diseases. Dr Huang was also supported by grant K24 DK105340 from the National Institute of Diabetes and Digestive and Kidney Diseases.

**Role of the Sponsor:** The study sponsors had no role in design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Meeting Presentation:** Presented in part at the American Diabetes Association's 78th Scientific Sessions; June 23, 2018; Orlando, Florida.

### REFERENCES

- Best JD, Drury PL, Davis TM, et al; Fenofibrate Intervention and Event Lowering in Diabetes Study Investigators. Glycemic control over 5 years in 4,900 people with type 2 diabetes: real-world diabetes therapy in a clinical trial cohort. *Diabetes Care*. 2012;35(5):1165-1170. doi:10.2337/dc11-1307
- Lipska KJ, Yao X, Herrin J, et al. Trends in drug utilization, glycemic control, and rates of severe hypoglycemia, 2006-2013. *Diabetes Care*. 2017;40(4):468-475. doi:10.2337/dc16-0985
- Landon BE, Zaslavsky AM, Souza J, Ayanian JZ. Trends in diabetes treatment and monitoring among Medicare beneficiaries. *J Gen Intern Med*. 2018;33(4):471-480. doi:10.1007/s11606-018-4310-4
- Lipska KJ, Ross JS, Van Houten HK, Beran D, Yudkin JS, Shah ND. Use and out-of-pocket costs of insulin for type 2 diabetes mellitus from 2000 through 2010. *JAMA*. 2014;311(22):2331-2333. doi:10.1001/jama.2014.6316
- Horvath K, Jeitler K, Berghold A, et al. Long-acting insulin analogues versus NPH insulin (human isophane insulin) for type 2 diabetes mellitus. *Cochrane Database Syst Rev*. 2007;(2):CD005613.
- Edelman SV, Polonsky WH. Type 2 diabetes in the real world: the elusive nature of glycemic control. *Diabetes Care*. 2017;40(11):1425-1432. doi:10.2337/dc16-1974
- Turner LW, Nartey D, Stafford RS, Singh S, Alexander GC. Ambulatory treatment of type 2 diabetes in the US, 1997-2012. *Diabetes Care*. 2014;37(4):985-992. doi:10.2337/dc13-2097
- Hua X, Carvalho N, Tew M, Huang ES, Herman WH, Clarke P. Expenditures and prices of antihyperglycemic medications in the United States: 2002-2013. *JAMA*. 2016;315(13):1400-1402. doi:10.1001/jama.2016.0126
- Cefalu WT, Dawes DE, Gavlak G, et al; Insulin Access and Affordability Working Group. Insulin Access and Affordability Working Group: Conclusions and Recommendations. *Diabetes Care*. 2018;41(6):1299-1311. doi:10.2337/dci18-0019
- Luo J, Avorn J, Kesselheim AS. Trends in Medicaid reimbursements for insulin from 1991 through 2014. *JAMA Intern Med*. 2015;175(10):1681-1686. doi:10.1001/jamainternmed.2015.4338
- Karter AJ, Schillinger D, Adams AS, et al. Elevated rates of diabetes in Pacific Islanders and Asian subgroups: the Diabetes Study of Northern California (DISTANCE). *Diabetes Care*. 2013;36(3):574-579. doi:10.2337/dc12-0722
- Karter AJ, Lipska KJ, O'Connor PJ, et al; SUPREME-DM Study Group. High rates of severe hypoglycemia among African American patients with diabetes: the Surveillance, Prevention, and Management of Diabetes Mellitus (SUPREME-DM) network. *J Diabetes Complications*. 2017;31(5):869-873. doi:10.1016/j.jdiacomp.2017.02.009
- Kirk JK, D'Agostino RB Jr, Bell RA, et al. Disparities in HbA1c levels between African-American and non-Hispanic white adults with diabetes: a meta-analysis. *Diabetes Care*. 2006;29(9):2130-2136. doi:10.2337/dc05-1973
- Karter AJ, Ferrara A, Liu JY, Moffet HH, Ackerson LM, Selby JV. Ethnic disparities in diabetic complications in an insured population. *JAMA*. 2002;287(19):2519-2527. doi:10.1001/jama.287.19.2519
- Huang ES, Liu JY, Moffet HH, John PM, Karter AJ. Glycemic control, complications, and death in older diabetic patients: the diabetes and aging study. *Diabetes Care*. 2011;34(6):1329-1336. doi:10.2337/dci10-2377
- Ginde AA, Blanc PG, Lieberman RM, Camargo CA Jr. Validation of ICD-9-CM coding algorithm for improved identification of

- hypoglycemia visits. *BMC Endocr Disord*. 2008;8:4. doi:10.1186/1472-6823-8-4
17. Little RR, Rohlfing CL, Sacks DB; National Glycohemoglobin Standardization Program (NGSP) Steering Committee. Status of hemoglobin A1c measurement and goals for improvement: from chaos to order for improving diabetes care. *Clin Chem*. 2011;57(2):205-214. doi:10.1373/clinchem.2010.148841
18. Sinisi SE, van der Laan MJ. Deletion/substitution/addition algorithm in learning with applications in genomics. *Stat Appl Genet Mol Biol*. 2004;3:e18. doi:10.2202/1544-6115.1069
19. Steiner JF, Koepsell TD, Fihn SD, Inui TS. A general method of compliance assessment using centralized pharmacy records: description and validation. *Med Care*. 1988;26(8):814-823. doi:10.1097/00005650-198808000-00007
20. Steiner JF, Prochazka AV. The assessment of refill compliance using pharmacy records: methods, validity, and applications. *J Clin Epidemiol*. 1997;50(1):105-116. doi:10.1016/S0895-4356(96)00268-5
21. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res*. 2011;46(3):399-424. doi:10.1080/00273171.2011.568786
22. Meyer BD. Natural and quasi-experiments in economics. *J Bus Econ Stat*. 1995;13(2):151-161.
23. Cheung YB. A modified least-squares regression approach to the estimation of risk difference. *Am J Epidemiol*. 2007;166(11):1337-1344. doi:10.1093/aje/kwm223
24. Messer LC, Laraia BA, Kaufman JS, et al. The development of a standardized neighborhood deprivation index. *J Urban Health*. 2006;83(6):1041-1062. doi:10.1007/s11524-006-9094-x
25. Laraia BA, Karter AJ, Warton EM, Schillinger D, Moffet HH, Adler N. Place matters: neighborhood deprivation and cardiometabolic risk factors in the Diabetes Study of Northern California (DISTANCE). *Soc Sci Med*. 2012;74(7):1082-1090. doi:10.1016/j.socscimed.2011.11.036
26. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol*. 1992;45(6):613-619. doi:10.1016/0895-4356(92)90133-8
27. Freemantle N, Chou E, Frois C, et al. Safety and efficacy of insulin glargine 300 u/mL compared with other basal insulin therapies in patients with type 2 diabetes mellitus: a network meta-analysis. *BMJ Open*. 2016;6(2):e009421. doi:10.1136/bmjopen-2015-009421
28. Singh SR, Ahmad F, Lal A, Yu C, Bai Z, Bennett H. Efficacy and safety of insulin analogues for the management of diabetes mellitus: a meta-analysis. *CMAJ*. 2009;180(4):385-397. doi:10.1503/cmaj.081041
29. Rys P, Wojciechowski P, Rogoz-Sitek A, et al. Systematic review and meta-analysis of randomized clinical trials comparing efficacy and safety outcomes of insulin glargine with NPH insulin, premixed insulin preparations or with insulin detemir in type 2 diabetes mellitus. *Acta Diabetol*. 2015;52(4):649-662. doi:10.1007/s00592-014-0698-4
30. Owens DR, Traylor L, Mullins P, Landgraf W. Patient-level meta-analysis of efficacy and hypoglycaemia in people with type 2 diabetes initiating insulin glargine 100U/mL or neutral protamine Hagedorn insulin analysed according to concomitant oral antidiabetes therapy. *Diabetes Res Clin Pract*. 2017;124:57-65. doi:10.1016/j.diabres.2016.10.022
31. Haukka J, Hoti F, Erästö P, Saukkonen T, Mäkimattila S, Korhonen P. Evaluation of the incidence and risk of hypoglycemic coma associated with selection of basal insulin in the treatment of diabetes: a Finnish register linkage study. *Pharmacoepidemiol Drug Saf*. 2013;22(12):1326-1335. doi:10.1002/pds.3534
32. Strandberg AY, Khanfir H, Mäkimattila S, Saukkonen T, Strandberg TE, Hoti F. Insulins NPH, glargine, and detemir, and risk of severe hypoglycemia among working-age adults. *Ann Med*. 2017;49(4):357-364. doi:10.1080/07853890.2016.1278302
33. Strandberg AY, Hoti FJ, Strandberg TE, Christopher S, Haukka J, Korhonen P. All-cause and cause-specific mortality among users of basal insulins NPH, detemir, and glargine. *PLoS One*. 2016;11(3):e0151910. doi:10.1371/journal.pone.0151910
34. Zhang T, Ji L, Gao Y, et al. Observational registry of basal insulin treatment in patients with type 2 diabetes in China: safety and hypoglycemia predictors. *Diabetes Technol Ther*. 2017;19(11):675-684. doi:10.1089/dia.2017.0111
35. Prentice JC, Conlin PR, Gellad WF, Edelman D, Lee TA, Pizer SD. Long-term outcomes of analogue insulin compared with NPH for patients with type 2 diabetes mellitus. *Am J Manag Care*. 2015;21(3):e235-e243.
36. Randall L, Begovic J, Hudson M, et al. Recurrent diabetic ketoacidosis in inner-city minority patients: behavioral, socioeconomic, and psychosocial factors. *Diabetes Care*. 2011;34(9):1891-1896. doi:10.2337/dc11-0701
37. Spain CV, Wright JJ, Hahn RM, Wivel A, Martin AA. Self-reported barriers to adherence and persistence to treatment with injectable medications for type 2 diabetes. *Clin Ther*. 2016;38(7):1653-1664.e1. doi:10.1016/j.clinthera.2016.05.009
38. Piette JD, Heisler M, Wagner TH. Problems paying out-of-pocket medication costs among older adults with diabetes. *Diabetes Care*. 2004;27(2):384-391. doi:10.2337/diacare.27.2.384
39. Karter AJ, Parker MM, Solomon MD, et al. Effect of out-of-pocket cost on medication initiation, adherence, and persistence among patients with type 2 diabetes: the Diabetes Study of Northern California (DISTANCE). *Health Serv Res*. 2018;53(2):1227-1247. doi:10.1111/1475-6773.12700
40. Lipska KJ, Hirsch IB, Riddle MC. Human insulin for type 2 diabetes: an effective, less-expensive option. *JAMA*. 2017;318(1):23-24. doi:10.1001/jama.2017.6939



# Clozapine Treatment for Suicidality in Schizophrenia

## International Suicide Prevention Trial (InterSePT)

Herbert Y. Meltzer, MD; Larry Alphas, MD, PhD; Alan I. Green, MD; A. Carlo Altamura, MD; Ravi Anand, MD; Alberto Bertoldi, MD; Marc Bourgeois, MD; Guy Chouinard, MD; M. Zahur Islam, PhD; John Kane, MD; Ranga Krishnan, MD; J.-P. Lindenmayer, MD; Steven Potkin, MD; for the InterSePT Study Group

**Background:** Approximately 50% of patients with schizophrenia or schizoaffective disorder attempt suicide, and approximately 10% die of suicide. Study results suggest that clozapine therapy significantly reduces suicidal behavior in these patients.

**Methods:** A multicenter, randomized, international, 2-year study comparing the risk for suicidal behavior in patients treated with clozapine vs olanzapine was conducted in 980 patients with schizophrenia or schizoaffective disorder, 26.8% of whom were refractory to previous treatment, who were considered at high risk for suicide because of previous suicide attempts or current suicidal ideation. To equalize clinical contact across treatments, all patients were seen weekly for 6 months and then biweekly for 18 months. Subsequent to randomization, unmasked clinicians at each site could make any interventions necessary to prevent the occurrence of suicide attempts. Suicidal behavior was assessed at each visit. Primary end points included suicide attempts (including those that led to death), hospitalizations to prevent suicide, and a rating of "much worsening of suicidality" from baseline. Masked raters, including an independent

suicide monitoring board, determined when end point criteria were achieved.

**Results:** Suicidal behavior was significantly less in patients treated with clozapine vs olanzapine (hazard ratio, 0.76; 95% confidence interval, 0.58-0.97;  $P=.03$ ). Fewer clozapine-treated patients attempted suicide (34 vs 55;  $P=.03$ ), required hospitalizations (82 vs 107;  $P=.05$ ) or rescue interventions (118 vs 155;  $P=.01$ ) to prevent suicide, or required concomitant treatment with antidepressants (221 vs 258;  $P=.01$ ) or anxiolytics or soporifics (301 vs 331;  $P=.03$ ). Overall, few of these high-risk patients died of suicide during the study (5 clozapine- vs 3 olanzapine-treated patients;  $P=.73$ ).

**Conclusions:** Clozapine therapy demonstrated superiority to olanzapine therapy in preventing suicide attempts in patients with schizophrenia and schizoaffective disorder at high risk for suicide. Use of clozapine in this population should lead to a significant reduction in suicidal behavior.

*Arch Gen Psychiatry.* 2003;60:82-91

**S**UICIDE IS the leading cause of premature death among patients with schizophrenia.<sup>1</sup> Overall, patients with schizophrenia have approximately a 50% lifetime risk for suicide attempts and a 9% to 13% lifetime risk for completed suicide.<sup>2,3</sup> In comparison, the lifetime risk for suicide in the general population of the United States is approximately 1%<sup>4</sup> and in persons with mood disorders, 9% to 15%.<sup>5</sup>

Studies<sup>6-11</sup> examining the effects of typical antipsychotic medications on suicide and suicidal ideation have not identified a change in incidence of suicide with their use. However, several studies suggest that using clozapine, an atypical antipsychotic drug, may reduce suicidal behavior in schizophrenia. Meltzer and Okayli<sup>12</sup> found that among 88 neuroleptic-resistant patients treated with clozapine for

0.5 to 7.0 years and evaluated for changes in suicidal behavior, the percentage with no suicidality of any type increased from 53% to 88%. In another study (which used the Clozaril National Registry to identify 57 000 current, recent, and past clozapine-treated patients, followed by linkage of these data with the National Death Index and Social Security Administration Death Master Files to identify cause of death), Walker and colleagues<sup>13</sup> found that mortality from suicide was markedly decreased in current clozapine users compared with past users. Analyses<sup>14,15</sup> of clozapine registry data from the United States and England support the view that the suicide completion rate in treated patients with schizophrenia is substantially reduced. Similarly, a retrospective study<sup>16</sup> of 295 treatment-resistant outpatients (mainly African Americans) found that

Author affiliations are listed at the end of this article. A complete list of the members of the InterSePT Study Group appears in the box on page 90.

clozapine treatment reduced the expected rate of suicidality during continuous drug administration. In contrast to these findings, a study<sup>17</sup> of the effect of clozapine treatment on completed suicides in the Veterans Administration system did not demonstrate a significant effect of clozapine therapy on the suicide rate. However, despite failure to match the comparison group on variables related to risk of suicide and follow-up that in some cases extended for prolonged periods after patients discontinued clozapine treatment, this study demonstrated a trend toward lowering the suicide rate for clozapine compared with nonclozapine treatment.

Taken together, these studies provide evidence for the ability of clozapine therapy to reduce suicidal behaviors<sup>18</sup>; however, they were retrospective and did not control for possible differences in the risk for suicide between the clozapine and comparison groups, relative differences in the dosage of clozapine vs the comparison antipsychotic drug, differences in the use of concomitant medications, or differences in the frequency of clinical contact (usually increased in patients treated with clozapine because blood monitoring to detect potential emergence of agranulocytosis is required). Furthermore, comparisons of the effects of clozapine therapy on suicidal behavior and the effects produced by using other atypical antipsychotic drugs, for example, quetiapine fumarate, olanzapine, risperidone, or ziprasidone hydrochloride, have not been performed. Because of their potential advantages regarding safety and efficacy relative to typical neuroleptic drugs, these comparators are the most relevant to the issue of drug management of suicidal patients with schizophrenia.

To address these limitations, a randomized, parallel-group study (International Suicide Prevention Trial [InterSePT]) was designed to compare changes in suicidality during treatment with clozapine vs olanzapine in patients with schizophrenia or schizoaffective disorder who are at high risk for suicide. Typical antipsychotic drugs were not included as comparator agents because, as mentioned previously, they have not been demonstrated to reduce the overall rate of suicidal behavior.<sup>6-11</sup> If anything, several studies have suggested that use of typical neuroleptic drugs increases the risk, possibly because of a combination of akathisia and secondary depression, leading to shorter hospitalizations (see Caldwell and Gottesman<sup>9</sup> for a review). Of the newer atypical antipsychotic drugs, olanzapine was selected because results of posthoc analyses suggest that its use may reduce the annual suicide attempt rate compared with haloperidol therapy and, in particular, may produce significantly greater improvement in the "suicidal thoughts" item of the Montgomery Asberg Depression Rating Scale.<sup>19</sup>

## METHODS

### PATIENTS

Men and women aged 18 to 65 years were included. They had a *DSM-IV* diagnosis of schizophrenia or schizoaffective disorder and were considered to be at high risk for committing suicide. High risk for suicide was defined as any one of the following: (1) a history of previous attempts or hospitalizations

to prevent a suicide attempt in the 3 years before enrollment, (2) moderate to severe current suicidal ideation with depressive symptoms, or (3) command hallucinations for self-harm within 1 week of enrollment. Treatment resistance was not a requirement for inclusion in this study. This study was conducted following Good Clinical Practice guidelines to ensure the protection of the participants. Before implementation of this study, institutional review board and ethics committee approvals of the protocol and the consent form were obtained from each of the participating sites.

### STUDY SITES AND DESIGN

Patients were recruited from 67 medical centers in 11 countries (the United States, Canada, France, Italy, the United Kingdom, the Czech Republic, Hungary, Croatia, South Africa, Argentina, and Chile). Selection of sites was based on demonstrated expertise in managing large numbers of schizophrenic patients at risk for suicide, expertise in conducting clinical trials, and the presence of adequate staff to manage these patients for 2 years.

The methods of the InterSePT study were developed collaboratively with academic experts. The design and data analysis were approved by the US Food and Drug Administration (FDA).

A 2-year study duration was chosen to provide time to obtain sufficient end points to differentiate between the 2 treatments. Because managing patients at high risk for suicide in a 2-year, double-blind study was not considered clinically feasible or acceptable and because treatment with clozapine is often associated with a variety of unique, clinically obvious adverse effects, the InterSePT was conducted as a randomized, open-label trial with masked ratings. After discussing the protocol and other treatment options in detail with the patient, written informed consent was obtained. Screened patients meeting inclusion criteria were assigned to treatment with either clozapine or olanzapine. For randomization, patients were blocked by country and medical center. The 2 treatment groups were allocated randomly in a 1:1 ratio within blocks of 4 patients in each medical center.

Recognizing the inherent problems associated with an open-label trial, extensive efforts were made to ensure masking of the raters. All aspects of the primary end points were limited to ratings from 1 of 2 groups of masked raters. A 3-member suicide monitoring board (SMB) determined whether potential end points met the criteria for a suicide attempt or a hospitalization to prevent suicide. The SMB was nominated by the study sponsor, Novartis Pharmaceuticals Corp, East Hanover, NJ, and approved by an academic steering committee. The SMB was chaired by one of us (R.K.) and included Hannale Heila, MD, from the National Public Health Institute in Helsinki, Finland, and Isaac Sakinofsky, PhD, from the University of Toronto, Toronto, Ontario. Each member had extensive experience working with suicidal patients. This team remained constant throughout the study.

Every type 1 end point (see the "Outcome Definitions" subsection) was reviewed by all of the members of the SMB, and consensus was obtained. At the onset of the study, the SMB developed the specific procedures for which data and how data would be reviewed and guidelines for how case evaluations would occur. Potential end point packages were reviewed and edited if necessary to eliminate any indication of the antipsychotic drug used, for example, adverse effects and blood monitoring. Masking of the SMB was monitored by Kevin Cox from Ingenix Pharmaceutical Services, San Diego, Calif, who ensured that all of the data received by the SMB was masked before delivery to the SMB. The institutions for which members of the SMB worked were not included as participant sites in the study.

**Table 1. Clinical Global Impression for Severity of Suicidality**

|   |   |
|---|---|
| Considering all of the information available to you, what was the most severe level of suicidality experienced by this patient in the past 7 days?  | 1 Not at all suicidal<br>2 Mildly suicidal<br>3 Moderately suicidal<br>4 Severely suicidal<br>5 Attempted suicide                           |
| Considering all of the information available to you, how much has the patient's suicidality changed compared with his or her condition at baseline? | 1 Very much improved<br>2 Much improved<br>3 Minimally improved<br>4 No change<br>5 Minimally worsened<br>6 Much worse<br>7 Very much worse |

For purposes of end point assessment, suicide attempts were defined as actions committed by a patient, either with willful intent or as a response to internal compulsions or disordered thinking, that put him or her at high risk for death. Potential end points were identified by investigators and by study monitors who reviewed medical records and adverse event data for potential end points. The SMB determined whether these potential end points met predefined criteria of intent and seriousness to qualify as a study end point.

In addition to these ratings by the SMB, masked psychiatrists at each participating site rated the global assessment of suicidality for type 2 events (see the "Outcome Definitions" subsection). Because these were ratings of an individual patient's suicidality compared with baseline, they could not be assessed by the independent SMB. These site raters' interactions with patients in this study were limited to these ratings, and they had no other clinical contact with them. Raters were required to verify their masking at the time of each rating, and they were regularly monitored by Ingenix Pharmaceutical Services to ensure that they had remained masked to the patient's treatment.

The weekly or biweekly clinical contact required to monitor for clozapine-associated agranulocytosis was matched with a similar visit schedule for olanzapine-treated patients, during which vital signs were obtained. To ensure the safety of patients during this trial, clinicians were allowed to make any interventions necessary to prevent the occurrence of suicide attempts, including changing dosages, adding other medications, switching medications, and increasing surveillance. In addition, patients were queried at each visit about suicidality, and they were referred to their treating physician if a full evaluation was indicated. Rescue interventions required to prevent suicide-related events were analyzed as secondary end points.

The first patient was enrolled in the InterSePT on March 19, 1998, and the last patient was enrolled on February 14, 1999. The last patient completed his last visit on February 14, 2001.

#### OUTCOME DEFINITIONS

It was recognized at the outset that use of suicide deaths alone as an end point would require approximately 20 000 patients if the study were powered to detect a decreased relative risk with clozapine therapy by 20%. Furthermore, because failure to intervene to prevent suicide during the follow-up care of patients in this study was considered ethically unacceptable, every effort was made to prevent this outcome. Thus, suicide deaths alone were not regarded as the primary end point but were included as a subset of a much larger group of suicidal behaviors that might occur during the study.

After consultation with external clinical experts in psychiatry, statistical experts, and the Neuropharmacology Division of the US FDA, 2 types of primary end points were prespecified for this trial. A type 1 event was defined as the occurrence of a significant suicide attempt (which included completed suicides as a subset) or hospitalization because of imminent suicide risk (which included increased levels of surveillance) that was confirmed by an external masked group (the SMB). A type 2 event was defined as ratings from a masked psychiatrist on the Clinical Global Impression of Suicide Severity (CGI-SS) of "much worse" or "very much worse" from baseline (**Table 1**). Because patients with potential type 2 events were not always observed by a masked psychiatrist, criteria for a significant level of worsening from baseline were also defined to have been met whenever a type 1 event occurred.

Other objective measures of suicidality included the individual components of the main outcome variable—the specific time to an SMB-determined suicide attempt (including death by suicide) or hospitalization to prevent suicide or the number of suicide attempts, the number of hospitalizations to prevent suicide, and the number of interventions to prevent suicide, irrespective of SMB validation.

Patients were enrolled for 2 years of follow-up. Attempts were made to continue to collect basic end point information even if the patient discontinued study participation early. This information was included in the intent-to-treat analysis.

In addition to these measures, the change in suicide risk was assessed clinically using the CGI-SS as an additional element of the combined end point. Information for rating this scale was collected during a semistructured interview designed to elicit the necessary information for ratings. Validation of this standardized clinical assessment demonstrates that this instrument is reliable and valid for assessing current suicidal thinking in patients with schizophrenia and schizoaffective disorder by clinicians and researchers. Additional information about the validation of this instrument is available elsewhere.<sup>20</sup> This 2-item scale measured severity of suicidality and change in suicidality from baseline (Table 1). Individuals were assessed on the CGI-SS by masked raters at each site at baseline and at study weeks 8, 16, 24, 32, 40, 48, 52, 60, 68, 80, 92, and 104.

Staff at the investigational sites were trained on InterSePT procedures and scales and were certified by study monitors as qualified raters before being permitted to participate. All new raters in the study were trained to meet prespecified criteria. Additional training was provided during the study to ensure that reliability was maintained over time.

#### STATISTICAL ANALYSIS

After the primary data were locked and verified, statistical analyses were conducted by Ingenix Pharmaceutical Services and supplied to us for examination, checking, and reporting. There were no interim analyses of data other than a masked review of safety data. Every effort was made to follow all patients to completion of the 2-year evaluation. All data obtained were used in the intent-to-treat analysis.

The null hypothesis for this study stated that the relative hazard for type 1 and type 2 events during treatment with clozapine compared with olanzapine treatment would not differ from 1. After consultation with the FDA, the predefined main analysis to test this null hypothesis used the method of Wei et al<sup>21</sup> for analysis of multiple events. In this analysis, the factor country was used as strata and randomized treatment group was used as the only covariate.

Supplementary supportive analyses of the SMB-determined end points were completed using the Cox proportional hazards regression model.<sup>22,23</sup> Putative explanatory variables, that



**Table 2. Demographic Characteristics of the Treatment Groups at Baseline**

| Characteristic                              | Clozapine Group<br>(n = 490) | Olanzapine Group<br>(n = 490) | Total<br>(N = 980) | P Value*<br>(Clozapine vs Olanzapine) |
|---|------------------------------|-------------------------------|--------------------|---------------------------------------|
| Age, mean ± SD, y                           | 37.1 ± 10.3                  | 37.0 ± 10.3                   | 37.1 ± 10.3        | .74                                   |
| Male sex, No. (%)                           | 301 (61.4)                   | 301 (61.4)                    | 602 (61.4)         | .98                                   |
| Race, No. (%)                               |                              |                               |                    | .49                                   |
| White                                       | 356 (72.7)                   | 337 (68.8)                    | 693 (70.7)         |                                       |
| Black                                       | 65 (13.3)                    | 86 (17.6)                     | 151 (15.4)         |                                       |
| Oriental                                    | 6 (1.2)                      | 7 (1.4)                       | 13 (1.3)           |                                       |
| Other                                       | 63 (12.9)                    | 60 (12.2)                     | 123 (12.6)         |                                       |
| Diagnosis, No. (%)                          |                              |                               |                    | .50                                   |
| Schizophrenic                               | 300 (61.2)                   | 309 (63.1)                    | 609 (62.1)         |                                       |
| Schizoaffective                             | 190 (38.8)                   | 181 (36.9)                    | 371 (37.9)         |                                       |
| Lifetime suicide attempts†                  |                              |                               |                    |                                       |
| No. (%)‡                                    | 413 (84.3)                   | 403 (82.2)                    | 816 (83.3)         | .58                                   |
| No., mean ± SD                              | 3.6 ± 7.5                    | 3.2 ± 4.5                     | 3.4 ± 6.2          | .80                                   |
| Suicide attempts in the past 36 mo, No. (%) | 307 (62.7)                   | 311 (63.5)                    | 618 (63.1)         | .76                                   |

\*Continuous variables were analyzed using an analysis of variance model (eg, model age = treatment + pooled country). Categorical variables were analyzed using the Cochran-Mantel-Haenszel method stratified by pooled country.

†Results were analyzed using the Wilcoxon test.

‡Number of patients who made ≥1 suicide attempt before baseline.

is, factors that may have contributed to the primary end point in this model, included treatment, number of previous suicide attempts, active substance or alcohol abuse, country, medical center, sex, and age group (18-32, 33-44, and ≥45 years) at baseline. The hazard ratio, a measure of the relative risk between groups, and its 95% confidence interval (CI), were computed on the basis of the fitted model. In addition, Kaplan-Meier estimates of survival probabilities were calculated. The average number of patients needed to treat to show a benefit of clozapine use over olanzapine use regarding the primary end point was computed using the method of Altman.<sup>24</sup> Annualized rates of suicide attempts were calculated as the number of total suicide attempts (including suicide) among the randomized patients divided by the total number of patient-years in the study. For this calculation, the total number of patient-years included the time during follow-up after dropout from the study. Compliance was estimated by dividing the number of study drug pills returned by the number of study drug pills dispensed, subtracting from 1 and multiplying by 100.

Because the CGI-SS was assessed relatively infrequently during the study, based on an agreement with the FDA, an inferred rating of “much worsening from baseline” was assumed on this scale if a patient attempted suicide or was hospitalized to prevent suicide. The number of patients with “much worsening from baseline CGI-SS” was analyzed using the Fisher exact test. Time to the “much worsening of CGI-SS” was analyzed with the methods used for the primary efficacy variable.

Event rates were computed for other efficacy and safety assessments, and the tests of significance were performed using the Fisher exact test. All *P* values were based on 2-sided alternative hypotheses.

## RESULTS

### PATIENTS

Of 1065 patients screened for participation in this study, 980 (92%; 490 per group) met the inclusion criteria and gave written informed consent. They were then randomly assigned to treatment with either clozapine or olanzapine. Of the total sample, 609 patients (62%) were diagnosed as having schizophrenia and 371 (38%) as having schizoaffective disorder. At the time of enrollment, 263

patients (27%) were rated as treatment resistant. Of the total population studied, 477 patients received olanzapine, 479 received clozapine, and 24 were never treated for various administrative reasons. Eighty-three percent of patients had attempted suicide at least once during their lifetime, and 84% had been hospitalized to prevent a suicide attempt. Sixty-three percent of patients had attempted suicide in the previous 36 months. Patients treated with olanzapine or clozapine did not significantly differ in age, sex, race, diagnosis, treatment resistance, number of previous suicide attempts, or baseline concomitant medications (**Table 2**).

### TREATMENT AND EVENTS

The mean±SD prescribed dosages of study drugs were 16.6±6.4 mg daily for olanzapine and 274.2±155.0 mg daily for clozapine. Compliance was 94.4% for the clozapine-treated group and 95.8% for the olanzapine-treated group. Overall dropout rates for patients treated with clozapine vs olanzapine were not different. Reasons for dropout are summarized in **Table 3**. A total of 380 patients (39%) discontinued taking the study drug early: 99 (10%) withdrew consent, 74 (8%) for adverse events, and 72 (7%) were lost to follow-up. The rates of discontinuation for other reasons were infrequent and did not differ between treatment groups, except for discontinuations for unsatisfactory antisuicidal effect (6 olanzapine-treated patients [1%] and 0 clozapine-treated patients; *P* = .03). Every effort was made to follow patients for study end points for the full 2 years of evaluation, even after they formally discontinued using the study drug. Such information from “retrieved dropouts” was included in the intent-to-treat analyses. More discontinuations occurred early in the course of clozapine treatment (usually for adverse events), whereas there tended to be more olanzapine dropouts later in the study.

Of the 577 cases sent to the SMB for review (representing 443 unique patients), 483 were determined to

**Table 3. Patient Study Discontinuations\***

| Reason†   | Clozapine Group<br>(n = 490) | Olanzapine Group<br>(n = 490) | Total<br>(N = 980) | P Value‡   |
|---|------------------------------|-------------------------------|--------------------|------------|
| Adverse event(s)  | 41 (8.4)                     | 33 (6.7)                      | 74 (7.6)           | .40        |
| Abnormal laboratory value(s)                                | 2 (0.4)                      | 0                             | 2 (0.2)            | .50        |
| Abnormal test procedure result(s)                           | 1 (0.2)                      | 0                             | 1 (0.1)            | >.99       |
| Unsatisfactory therapeutic effect on psychosis              | 5 (1.0)                      | 9 (1.8)                       | 14 (1.4)           | .42        |
| Unsatisfactory therapeutic effect for lowering suicide risk | 0                            | 6 (1.2)                       | 6 (0.6)            | .03§       |
| Deaths  | 8 (1.6)                      | 5 (1.0)                       | 13 (1.3)           | .42        |
| Protocol violation  | 29 (5.9)                     | 20 (4.1)                      | 49 (5.0)           | .24        |
| Consent withdrawal  | 50 (10.2)                    | 49 (10.0)                     | 99 (10.1)          | >.99       |
| Lost to follow-up   | 33 (6.7)                     | 39 (8.0)                      | 72 (7.3)           | .54        |
| Administrative problems                                     | 23 (4.7)                     | 26 (5.3)                      | 49 (5.0)           | .77        |
| <b>Total</b>  | <b>192 (39.2)</b>            | <b>187 (38.2)</b>             | <b>379 (38.7)</b>  | <b>.79</b> |

\*Values are given as number (percentage) of patients.

†Rated by treating physician.

‡Fisher exact test.

§Statistically significant difference.

**Table 4. SMB-Certified and Other Measures of Suicidality\***

|   | Clozapine Group<br>(n = 490) | Olanzapine Group<br>(n = 490) | P Value†<br>(95% CI of the Difference)<br>(Olanzapine – Clozapine) |
|---|------------------------------|-------------------------------|--|
| SMB-determined end points                                       |                              |                               |  |
| Patients with end points, total‡                                | 102 (20.8)                   | 141 (28.8)                    | .005 (.03 to .13)  |
| Patients with significant suicide attempts‡                     | 34 (6.9)                     | 55 (11.2)                     | .03 (.01 to .08)   |
| Patients with hospitalizations to prevent suicide‡              | 82 (16.7)                    | 107 (21.8)                    | .05 (.00 to .10)   |
| All SMB-determined end points, total No.                        | 217                          | 266                           | ...  |
| Patients showing "much worsening from baseline" on the CGI-SS‡§ | 120 (24.5)                   | 161 (32.9)                    | .005 (.03 to .14)  |
| Discontinuations for unsatisfactory antisuicidal effect‡        | 0                            | 6 (1.2)                       | .03  |
| Concomitant medications   |                              |                               |  |
| Antidepressants   | 235 (49.1)                   | 263 (55.1)                    | .01 (.02 to .14)   |
| Anxiolytics/soporifics  | 297 (60.2)                   | 331 (69.4)                    | .05 (.01 to .13)   |
| Adverse events  |                              |                               |  |
| Patients experiencing "suicide attempt"                         | 37 (7.7)                     | 66 (13.8)                     | .002 (.02 to .10)  |
| Patients experiencing "suicidal ideation"                       | 125 (26.1)                   | 153 (32.1)                    | .05 (.00 to .12)   |
| Suicide deaths‡¶  | 5 (1.0)                      | 3 (0.6)                       | .73 (-.02 to .01)  |
| All rescue interventions to prevent suicide‡                    | 118 (24.1)                   | 155 (31.6)                    | .01 (.02 to .13)   |

Abbreviations: CGI-SS, Clinical Global Impression of Suicide Severity; CI, confidence interval; SMB, Suicide Monitoring Board.

\*Values are given as number (percentage) of patients, unless otherwise indicated. Ellipses indicate not applicable.

†Fisher exact test.

‡Intention-to-treat population (all randomized patients): clozapine, n = 490; olanzapine, n = 490.

§Ratings by a masked psychiatrist. Includes implied worsening if a patient was hospitalized to prevent a suicide attempt or attempted suicide.

¶Safety population (all randomized patients who took the study drug): clozapine, n = 479; olanzapine, n = 477.

¶¶Includes all suicide deaths and all events leading to death by suicide.

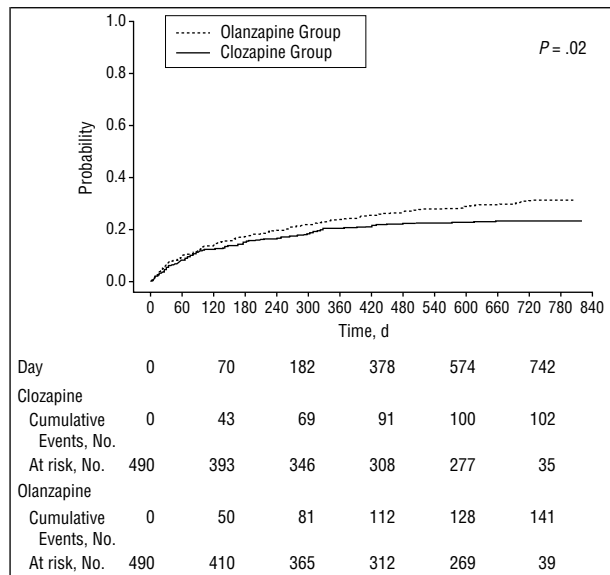
meet end point criteria; of these, 111 (clozapine, 43; olanzapine, 68) were considered suicide attempts and 372 (clozapine, 174; olanzapine, 198) were hospitalizations to prevent suicide.

## MAIN OUTCOME VARIABLES

### Primary Analysis

The results of the main analysis for primary efficacy, based on the method of Wei et al,<sup>21</sup> demonstrate a significant difference ( $P = .03$ ) between the clozapine group and the olanzapine group in reducing suicidality as measured by SMB-determined suicide attempts or hospitalizations to prevent suicide and much or very much worsening in the CGI-SS from baseline as assessed by the masked psychia-

trist. Results from the individual components of this end point indicate that compared with olanzapine therapy, clozapine use had hazard ratios of 0.76 (95% CI, 0.58-0.97) for type 1 events (suicide attempts or hospitalizations to prevent suicide) ( $P = .03$ ) and 0.78 (95% CI, 0.61-0.99) for type 2 events (worsening on the CGI-SS or implicit worsening of suicidality severity as demonstrated by occurrence of a type 1 event) ( $P = .04$ ), indicating a 24% and a 22% advantage, respectively, for clozapine therapy. Other efficacy results are summarized in Table 3 and **Table 4**. Because patients may have experienced more than 1 end point that met SMB criteria during the 2 years of observation, the absolute number of end points in each treatment group was also determined. There were more SMB-determined end points in the olanzapine-treated group than in the clozapine-



Kaplan-Meier estimates of the probability of a suicide attempt or hospitalization to prevent suicide.

treated group (266 vs 217 events). Of the prognostic factors examined using the Cox proportional hazards regression model as potential predictors of suicide events, only the number of lifetime suicide attempts at baseline (hazard ratio, 1.03; 95% CI, 1.01-1.04;  $P < .001$ ) and current or previous history of alcohol or other drug abuse at baseline (hazard ratio, 1.48; 95% CI, 1.11-1.99;  $P = .008$ ) predicted the SMB-determined suicide events. The results were largely consistent across countries, regions, and diagnoses. Treatment resistance did not predict differential response to clozapine or olanzapine regarding suicidality (data not shown).

### Secondary Analysis for Primary End Points

During the InterSePT, the percentage of patients experiencing at least 1 significant suicide attempt or hospitalization to prevent suicide that met criteria established by the masked SMB was greater in the olanzapine-treated group than in the clozapine-treated group (28.8% vs 20.8%;  $P = .005$ ) (Table 4). The olanzapine-treated group experienced significantly more SMB-determined suicide attempts (11.2% vs 6.9%;  $P = .03$ ) and SMB-determined hospitalizations to prevent suicide (21.8% vs 16.7%;  $P = .05$ ). In the analyses of the changes from baseline in CGI-SS scores, there was a significant difference in the percentage of patients who experienced “much worsening from baseline” (olanzapine, 32.9% vs clozapine, 24.5%;  $P = .005$ ).

Overall, there were fewer deaths from suicide than expected, particularly considering that only individuals at high risk for suicide were included in this study. Three suicide attempts in the olanzapine-treated group (0.6%) and 5 in the clozapine-treated group (1.0%) resulted in death (95% CI, 0.40%-7.04%;  $P = .73$ ). On the other hand, when suicide attempts were rated for probability of success by the SMB, a “high probability for success” of a completed suicide was found for 8 events in the clozapine group and 14 events in the olanzapine group. The latter group included

**Table 5. Summary Statistics for Individuals Who Received Rescue Interventions\***

| Rescue Intervention                             | Clozapine Group (n = 490) | Olanzapine Group (n = 490) | P Value† |
|---|---------------------------|----------------------------|----------|
| Hospitalization for imminent risk of suicide    | 100 (20.4)                | 128 (26.1)                 | .04      |
| Increase in level of surveillance               | 41 (8.4)                  | 57 (11.6)                  | .11      |
| Addition of an antidepressant agent             | 15 (3.1)                  | 33 (6.7)                   | .01      |
| Addition of or change in an antipsychotic agent | 15 (3.1)                  | 29 (5.9)                   | .04      |
| Increase in dose of study medication            | 37 (7.6)                  | 34 (6.9)                   | .81      |
| Addition of a mood stabilizer                   | 6 (1.2)                   | 14 (2.9)                   | .11      |
| Addition of other medications                   | 29 (5.9)                  | 42 (8.6)                   | .14      |
| Admission to a partial hospital program         | 2 (0.4)                   | 3 (0.6)                    | >.99     |
| Increase in medication monitoring visits        | 5 (1.0)                   | 9 (1.8)                    | .42      |
| Institution of psychotherapy                    | 6 (1.2)                   | 4 (0.8)                    | .75      |
| Increase in frequency of psychotherapy visits   | 1 (0.2)                   | 7 (1.4)                    | .07      |
| Emergency department visits                     | 34 (6.9)                  | 43 (8.8)                   | .34      |
| Crisis team visit                               | 7 (1.4)                   | 4 (0.8)                    | .55      |
| Electroconvulsive therapy                       | 5 (1.0)                   | 0                          | .06      |
| Increase in frequency of clinic visits          | 4 (0.8)                   | 10 (2.0)                   | .18      |
| Other   | 19 (3.9)                  | 26 (5.3)                   | .36      |
| Total   | 118 (24.1)                | 155 (31.6)                 | .01      |

\*Values are given as number (percentage) of patients.

†Fisher exact test.

the 6 patients discontinued from olanzapine therapy for unsatisfactory control of suicidality. None of the clozapine-treated patients were discontinued for this reason.

The supportive analysis using the Cox proportional hazards regression model with treatment, number of previous suicide attempts, active substance or alcohol abuse, country, sex, and age group at baseline in the model demonstrated a 26% reduced risk for suicide attempt or hospitalization to prevent suicide (type 1 event) for patients randomized to clozapine treatment compared with olanzapine treatment ( $P = .02$ ; hazards ratio, 0.74; 95% CI, 0.57-0.96). Kaplan-Meier estimates for probability of an event were estimated for the 2 treatment groups (Figure). The time in days to observe the first 70 type 1 events was 185 days for the clozapine-treated patients and 126 days for the olanzapine-treated patients. A significant reduction in the 2-year event rate at the end of the study (olanzapine, 32.2% vs clozapine, 24.0%; 95% CI of the difference, 0.02-0.14; number needed to treat, 12) and a delay in time to event were demonstrated for clozapine-treated patients. The number needed to treat of 13 indicates that under similar treatment conditions, on average, for every 13 high-risk patients treated, suicidal events, as defined herein, would be observed in 1 less patient if they were treated with clozapine rather than olanzapine. The overall annualized rate for attempted suicides (including suicide deaths) was 7.1%, with a rate of 8.5% for olanzapine-treated patients and 5.6% for clozapine-treated patients.

**Table 6. Adverse Events of Interest in the Safety Population\***

| Adverse Event                 | Clozapine Group (n = 479) | Olanzapine Group (n = 477) | P Value† |
|-------------------------------|---------------------------|----------------------------|----------|
| Suicidal ideation             | 125 (26.1)                | 153 (32.1)                 | .05      |
| Suicide attempt               | 37 (7.7)                  | 66 (13.8)                  | .002     |
| Laceration                    | 2 (0.4)                   | 19 (4.0)                   | <.001    |
| Depression NEC                | 137 (28.6)                | 173 (36.3)                 | .01      |
| Mood alteration               | 5 (1.0)                   | 13 (2.7)                   | .06      |
| Mood disorder                 | 0                         | 8 (1.7)                    | .004     |
| Insomnia NEC                  | 96 (20.0)                 | 155 (32.5)                 | <.001    |
| Somnolence                    | 220 (45.9)                | 118 (24.7)                 | <.001    |
| Convulsions                   | 12 (2.5)                  | 2 (0.4)                    | .01      |
| Diabetes mellitus NOS         | 16 (3.3)                  | 21 (4.4)                   | .41      |
| Weight gain                   | 150 (31.3)                | 265 (55.6)                 | <.001    |
| Cardiomyopathy                | 0                         | 1 (0.2)                    | .50      |
| Postural hypotension          | 21 (4.4)                  | 1 (0.2)                    | <.001    |
| Dizziness (excluding vertigo) | 129 (26.9)                | 59 (12.4)                  | <.001    |
| Syncope                       | 15 (3.1)                  | 5 (1.0)                    | .03      |
| Akathisia                     | 21 (4.4)                  | 39 (8.2)                   | .02      |
| Muscle rigidity               | 1 (0.2)                   | 6 (1.3)                    | .02      |
| Dysarthria                    | 23 (4.8)                  | 2 (0.4)                    | <.001    |
| Constipation                  | 120 (25.1)                | 46 (9.6)                   | <.001    |
| Dry mouth                     | 26 (5.4)                  | 43 (9.0)                   | .03      |
| Salivary hypersecretion       | 229 (47.8)                | 28 (5.9)                   | <.001    |
| Dyspepsia                     | 66 (13.8)                 | 40 (8.4)                   | .01      |
| Nausea                        | 79 (16.5)                 | 47 (9.9)                   | .003     |
| Vomiting NOS                  | 79 (16.5)                 | 42 (8.8)                   | <.001    |
| Urinary incontinence          | 31 (6.5)                  | 6 (1.3)                    | <.001    |
| Weakness                      | 36 (7.5)                  | 11 (2.3)                   | <.001    |
| Drug abuse                    | 4 (0.8)                   | 14 (2.9)                   | .02      |
| Alcoholism                    | 13 (2.7)                  | 21 (4.4)                   | .17      |
| WBC count decreased           | 28 (5.8)                  | 4 (0.8)                    | <.001    |

Abbreviations: NEC, not elsewhere classified; NOS, not otherwise specified; WBC, white blood cell.

\*Values are given as number (percentage) of patients.

†Fisher exact test.

### Secondary Outcome Analysis

To maintain the safety of individuals participating in this study, clinicians were permitted to make any interventions they believed necessary to prevent an impending suicide attempt. Rescue interventions to prevent suicide were significantly greater in the olanzapine-treated group vs the clozapine-treated group (31.6% vs 24.1%;  $P = .01$ ) (Table 5).

Comparison of the overall use of concomitant medications for any reason, not only suicidality, shows that antidepressants were used more frequently in patients treated with olanzapine (55.1%) than in those treated with clozapine (49.1%) ( $P = .01$ ) (Table 4). Anxiolytics and soporifics were also used more frequently in patients treated with olanzapine (69.4%) than in those treated with clozapine (60.2%) ( $P = .05$ ). Significant differences in the use of other concomitant medications (antipsychotics and mood stabilizers) between treatment groups were not seen. There were more discontinuations because of an unsatisfactory antisuicidal therapeutic effect as determined by the treating physician in the olanzapine-treated group ( $n = 6$ ; 1.2%) vs the clozapine-treated group ( $n = 0$ ) ( $P = .03$ ) (Table 4). More suicide attempts met the criteria for an adverse event (but not necessarily the more rigorous criteria of the SMB) in the olanzapine-treated group vs the

clozapine-treated group (13.8% vs 7.7%;  $P = .002$ ) (Table 4). Similarly, there were more adverse events of "suicidal ideation" in the olanzapine-treated group vs the clozapine-treated group (32.1% vs 26.1%;  $P = .05$ ).

### SAFETY AND TOLERABILITY

The overall number of adverse events and clinically serious adverse events did not differ between treatment groups in this 2-year, prospective comparative study of 2 widely used antipsychotic drugs. However, several differences in the specific adverse event profile for clozapine use and olanzapine use were noted (Table 6). The most frequently observed adverse events attributed to clozapine treatment were salivary hypersecretion, somnolence, weight gain, and dizziness (excluding vertigo). The most frequently observed adverse events attributed to olanzapine treatment were weight gain, somnolence, dry mouth, and dizziness (excluding vertigo). These results will be presented in more detail elsewhere (John Kane, MD, unpublished observations, 2002; Tom Fahy, MD, unpublished observations, 2002). Decreased white blood cell counts were reported as an adverse event in 0.8% of olanzapine-treated patients and 5.8% of clozapine-treated patients ( $P < .001$ ). However, no agranulocytosis or deaths related to granulocytopenia were reported for either treatment group.

There were 8 deaths (1.7%) for any reason in the olanzapine group and 12 (2.5%) in the clozapine group ( $P = .50$ ). Causes of death for olanzapine-treated patients were suicide ( $n = 3$ , 0.6%), cardiorespiratory arrest ( $n = 2$ , 0.4%), and carcinoma, cardiac arrhythmia, and myocardial infarction (after randomization but before treatment) ( $n = 1$  each, 0.2%). Causes of death for clozapine-treated patients were suicide ( $n = 5$ , 1.0%), cardiac arrhythmia ( $n = 2$ , 0.4%), and lymphoma, coronary artery disease, pulmonary embolism, car accident, and stroke ( $n = 1$  each, 0.2%).

### COMMENT

The major finding of this randomized study is that clozapine therapy is superior to olanzapine therapy in reducing key measures of suicidality in patients with schizophrenia or schizoaffective disorder who are at high risk for suicide. In particular, the treatment effect on the most objective measures of suicidality (time to suicide attempts [including deaths by suicide] and time to hospitalizations to prevent suicide) significantly favored clozapine treatment over olanzapine use when analyzed by multiple approaches (including various analyses of time to event, survival analysis methods, and evaluation of the total number of events). The hazard ratios identified in this study suggest that serious suicide attempts and hospitalizations to prevent suicide can be reduced by about one fourth with clozapine treatment vs olanzapine treatment.

Although the total number of suicide-related deaths was greater in the olanzapine-treated group, this was not significant and, as indicated already, the study was not powered to evaluate this as an end point. To have made this the sole primary end point, the observed suicide rate



in this study indicates that approximately 20000 patients would have been needed to find a 20% reduction in relative risk between the 2 drugs. Many factors contribute to whether a serious suicide attempt will lead to death. In addition, when the overall use of various interventions to prevent suicide attempts was compared between treatments, the results consistently supported the superiority of clozapine therapy over olanzapine therapy to reduce the risk of suicide. In particular, antidepressant and anxiolytic drugs were most likely given to alleviate depression, hopelessness, or anxiety or agitation—conditions that are frequently associated with increased risk for suicide.<sup>4</sup> Although use of these agents may have served to diminish the rate of suicidal behavior observed in both treatment groups, the more frequent prescription of these agents in olanzapine-treated patients did not succeed in equalizing the effects of olanzapine therapy and clozapine therapy on suicidality.

Khan et al<sup>11</sup> recently used an FDA database to access data from 10118 patients participating in pivotal clinical trials of treatment with olanzapine, risperidone, and quetiapine fumarate and compared the rates of suicide and suicide attempts with those of patients randomized to receive placebo or established (“typical”) antipsychotic drugs. Annualized rates of attempted suicide (including completed suicides) were 3.3% during placebo treatment, 5.7% during treatment with an established antipsychotic agent, and 5.0% during atypical antipsychotic drug treatment (not including clozapine). Despite the statistical power provided by the large sample size, the rates for suicide attempts (including completed suicides) among these 3 schizophrenic treatment groups (not preselected for suicidality) were not significantly different from each other. In addition, these data affirm the high risk for suicide in patients with schizophrenia.

The annualized rate for attempted suicide (including completed suicide) in the InterSePT, which selected schizophrenic patients at high risk for suicide, was 7.1%, a rate not appreciably different from that for use of atypical antipsychotic drugs in the study by Khan et al.<sup>11</sup> Given that the selection criteria for the InterSePT study population required evidence for a high risk of suicide, a higher rate of treatment-emergent suicidal behaviors was expected than was described in the FDA sample. The fact that the rate is similar to that reported by Khan et al<sup>11</sup> suggests that the InterSePT itself (through the psychosocial interventions, the drugs used, or both) seemed to decrease the expected suicide rate. The advantages of using clozapine over olanzapine for reducing suicidality might be even more evident in clinical practice than reported herein because the increased clinical contact required for clozapine treatment to manage the risk of agranulocytosis might further reduce suicide risk relative to other antipsychotic drug treatments. It is unlikely that the extra contact available to the olanzapine-treated patients would be available in usual clinical practice.

Because risk of suicide is now one of the chief indicators for hospitalization of patients with schizophrenia or schizoaffective disorder, these results suggest that wider use of clozapine in suicidal patients with schizo-

phrenia could reduce costs of their treatment. In addition, the decrease in suicidal behaviors observed with clozapine treatment has important potential quality-of-life benefits for individual patients, their families, and society. Together with the reduced risk of suicide when receiving clozapine therapy, these considerations suggest a more favorable risk-benefit analysis for the use of clozapine, especially in patients at risk for suicide.

Some limitations of this study should be noted. Although treatment assignment was randomized and the key ratings were masked, the study was not completely double-blinded. The decision not to use a double-blind design was based on concerns that a true blinding could not be maintained during a 2-year study given the well-known, recognizable, and common differentiating adverse effects of the 2 treatments. Another concern was that masked drug treatment might have hampered the flexibility of clinical care necessary to reduce the possibility of death by suicide in patients during a 2-year study. With full awareness of potential problems in a masked rater study, care was taken to ensure that the SMB and the masked raters did not have access to any source of data that might unblind them. Second, although clozapine treatment requires additional clinical contact related to white blood cell count monitoring, this study was not designed to determine whether any beneficial effects of clozapine treatment on suicidality are related to this additional contact. However, the equivalent clinical contact in the olanzapine-treated group demonstrates that increased contact alone cannot account for the clozapine effect on suicidality relative to olanzapine observed in this and other studies. Third, this study did not include a typical neuroleptic drug as a comparator. However, given the evidence that these drugs do not reduce the risk of suicide and<sup>6-9</sup> that the adverse effects of these drugs may be associated with increased risk for suicide<sup>2</sup> and independent evidence indicating that clozapine treatment reduces the suicide attempt and completion rate by approximately 80% compared with typical neuroleptic-treated patients,<sup>9,15</sup> there is no reason to expect that use of a typical antipsychotic agent would have had an effect on suicidal behavior comparable to clozapine therapy.

Strengths of this study include (1) the large sample size; (2) the use of a masked SMB, which utilized a uniform set of criteria for classifying potential suicide events; (3) clinician freedom to use adjunctive treatments as needed to minimize suicidality; (4) the inclusion of a broad range of nonrefractory patients at risk for suicide, including schizoaffective patients, whose risk for suicide tends to be greater than that of schizophrenic patients<sup>25</sup>; and (5) a dropout rate well within the range found in clinical trials of schizophrenia (despite the length of the study and the demanding nature of the protocol). Moreover, clinicians were free to use any dose of clozapine or olanzapine that they believed was merited by their patients' clinical conditions. The mean dose of clozapine was much lower than that usually used to treat refractory patients, reflecting the fact that only one quarter of the patients were refractory. Plasma levels of clozapine were not determined during the study. On the other hand, the olanzapine dose was similar to the mean daily dose cur-

In addition to the authors of this article, the InterSePT Study Group included the following principal investigators, Steering Committee members, Suicide Monitoring Board members, and Novartis Pharmaceuticals Corp employees: Saide Altinsan, MD; Siemion Altman, MD; Likiana Avigo, MD; Richard Balon, MD; Vanda Benešová, MD; Luis Bengochea, MD; Istvan Bitter, MD; Elisabeth Bokowska, MD; Bernardo Carpiello, MD; Daniel Casey, MD; Giovanni Cassano, MD; James Chou, MD; Guy Chouinard, MD; Libor Chvila, MD; Jean Dalery, MD; Pedro Delgado, MD; Liliana Dell'Osso, MD; Carl Eisdorfer, MD, PhD; Robin A. Emsley, MD; Dawn Eng, MD; Tom A. Fahy, MD; Vera Folnegovic, MD; Sophie Frangou, MD; Pedro Gargoloff, MD; Alberto Giannelli, MD; Ira Glick, MD; Richard Greenberg, MD; George T. Grossberg, MD; Doris Gundersen, MD; Hannale Heila, MD; George Hsu, MD; Naveed Iqbal, MD; M. Miro Jakovljevic, MD; Richard C. Josiassen, PhD; Akos Kassaifarkas, MD; Rob Kerwin, MD; Frederic Khidichian, MD; Mary Ann Knesevich, MD; Jack Krasuski, MD; Vinod Kumar, MD; Veronica Walters Larach, MD; Michael Lesem, MD; Shon Lewis, MD; Pierre-Michel Llorca, MD; H. Edward Logue, MD; Stephen Martin, MD; Muriel Maurel-Raymondet, MD; Laszlo Mod, MD; Eva Morik, MD; Carlos Morra, MD; Ann Mortimer, MD; Mojtaba Noursalehi, PhD; Gyorgy Ostorharics-Horvath, MD; Ivo Paclt, MD; Jorg J. Pahl, MD; Linda Pestreich, Jeffrey Lee Peters, MD; Rosario Pioli, MD; Michael G. Plopper, MD; Thomas Posever, MD; Mark Rapaport, MD; Delbert Robinson, MD; Carlo Andrea Robotti, MD; Harry Rohme, PhD; Frederic Rouillon, MD; David Sack, MD; Isaac Sakinsolfsky, PhD; Phillip Seibel, MD; George Simpson, MD; Nancy Temkin, PhD; Oladapo Tomori, MD; Santha Vaidain, MD; Zdeňka Vyhnanđová, MD; Frederick Young, PhD; Daniel Zimbhoff, MD; Marie-Agathe Zimmerman, MD.

rently used in the United States to treat patients with schizophrenia.

The results reported herein are consistent with a large body of data from the United States, the United Kingdom, and elsewhere indicating that clozapine treatment can reduce the suicide rate in these patients.<sup>12-16</sup> Some studies suggest that treatment response to clozapine administration is particularly evident in patients who have increased suicidality, and that this response may extend to patients with bipolar disorders.<sup>25-28</sup> Although these data provide compelling evidence for an effect of clozapine use in reducing suicidality, the mechanism for this effect requires further study. Data from this study suggest that the effect of clozapine therapy may not relate to its superior efficacy for treatment-resistant psychotic symptoms. Alternative mechanisms that have been suggested for the effect of clozapine use include an intrinsic antidepressant activity,<sup>12</sup> as also suggested by effects on mood symptoms and the differential antidepressant drug use in this study. Other data<sup>20</sup> suggest that suicidality may represent a separate symptom domain that is related to, but independent of, depression or psychosis.<sup>20</sup> The failure of treatment with typical antipsychotic agents to reduce suicidal behavior indicates that these symptoms are distinct from the major positive symptoms—

delusions and hallucinations—for which these drugs are effective in approximately 70% of patients. Similarly, some studies indicate that classifying patients with schizophrenia as treatment responders and nonresponders to the antipsychotic effects of neuroleptic therapy does not differentiate them with regard to suicidality<sup>12</sup> or, as shown here, with regard to the ability of clozapine therapy to reduce suicidality. Together, these data indicate that the effect of clozapine use on suicidal behavior, although perhaps related to some of its other clinical advantages, could represent a separate outcome of clozapine treatment.

Suicidal behavior in persons with schizophrenia and schizoaffective disorder is recognized as a pressing public health problem.<sup>29,30</sup> To our knowledge, except for clozapine therapy, no pharmacologic treatment has been demonstrated to be useful in reducing suicidal behavior in patients with schizophrenia. The InterSePT indicates that, on average, treatment of only 12 patients with clozapine rather than olanzapine will show benefit for clozapine to reduce suicidal behavior. As discussed herein, these results suggest an advantage of clozapine therapy over olanzapine therapy to reduce the risk of suicide in patients with schizophrenia and schizoaffective disorder. Additional study is needed to determine whether the advantage of clozapine therapy for reducing suicidal behavior also holds for patients with other conditions in which antipsychotic drug use is widespread and suicide occurs at high rates, particularly bipolar disorder, major depression with psychotic features, and borderline personality disorder.

*Submitted for publication February 14, 2002; final revision received June 14, 2002; accepted June 14, 2002.*

*From the Department of Psychiatry, Vanderbilt University, Nashville, Tenn (Dr Meltzer); Pfizer Inc, Ann Arbor, Mich (Dr Alphs); the Departments of Psychiatry, Dartmouth Medical School, Hanover, Mass (Dr Green), and University of Milan, Milan, Italy (Dr Altamura); Organon, Oss, the Netherlands (Dr Anand); the Departments of Psychiatry, La Plata University, La Plata, Argentina (Dr Bertoldi), University of Bordeaux, Bordeaux, France (Dr Bourgeois), McGill University, Montreal, Quebec (Dr Chouinard), and University of Montreal, Montreal (Dr Chouinard); Novartis Pharmaceuticals Corp, East Hanover, NJ (Dr Islam); and the Departments of Psychiatry, Hillside Hospital, Glen Oaks, NY (Dr Kane), Albert Einstein University, Bronx, NY (Dr Kane), Duke University, Durham, NC (Dr Krishnan), Manhattan Psychiatric Center—Nathan Kline Institute for Psychiatric Research, Wards Island, NY (Dr Lindenmayer), and University of California, Irvine (Dr Potkin). Drs Meltzer, Green, Chouinard, Kane, Lindenmayer, and Potkin have received grants from or are consultants to Novartis Pharmaceuticals Corp. Dr Chouinard has also received research support from Merck Frosst Canada, Inc, Dorval, Quebec; Janssen-Ortho Inc, Toronto, Ontario; and Pfizer Canada Inc, Dorval; and is a consultant to Janssen-Ortho Inc and Organon.*

*This study was supported by Novartis Pharmaceuticals Corp and by additional grants from the William K. Warren Research Foundation, Tulsa, Okla, and the Donald Test Foundation Trust and Lydia Bryant Test, Dallas, Tex (Dr Meltzer).*

We thank Ingenix Pharmaceutical Services for monitoring this study, ensuring that it was conducted according to protocol, and maintaining the study masking by coordinating communication and data transfers among the sponsor, the sites, and the SMB. We also thank the InterSePT Study Group for work on this project.

Corresponding author and reprints: Herbert Y. Meltzer, MD, Psychiatric Hospital at Vanderbilt, 160123rd Ave S, Suite 306, Nashville, TN 37212 (e-mail: herbert.meltzer@mcmail.vanderbilt.edu).

## REFERENCES

1. Cohen LJ, Test MA, Brown RL. Suicide and schizophrenia: data from a prospective community treatment study. *Am J Psychiatry*. 1990;147:602-607.
2. Meltzer HY, Fatemi H. Suicide in schizophrenia: the effect of clozapine. *Clin Neuropharmacol*. 1995;18(suppl):S18-S24.
3. Nyman A, Jonsson H. Patterns of self-destructive behaviour in schizophrenia. *Acta Psychiatr Scand*. 1986;73:252-262.
4. Caldwell CB, Gottesman II. Schizophrenia: a high-risk factor for suicide: clues to risk reduction. *Suicide Life Threat Behav*. 1992;22:479-493.
5. Harris EC, Barraclough B. Suicide as an outcome for mental disorders: a meta-analysis. *Br J Psychiatry*. 1997;170:205-228.
6. Axelsson R, Lagerkvist-Briggs M. Factors predicting suicide in psychotic patients. *Eur Arch Psychiatr Clin Neurosci*. 1992;241:259-266.
7. Winokur G, Tsuang M. The Iowa 500: suicide in mania, depression and schizophrenia. *Am J Psychiatry*. 1975;132:650-651.
8. Johns CA, Stanley M, Stanley R. Suicide in schizophrenia. *Ann N Y Acad Sci*. 1986;487:294-300.
9. Caldwell CB, Gottesman II. Schizophrenics kill themselves too: a review of risk factors for suicide. *Schizophr Bull*. 1990;16:571-589.
10. Siris SG. Suicide and schizophrenia. *J Psychopharmacol*. 2001;15:127-135.
11. Khan A, Khan SR, Leventhal RM, Brown WA. Symptom reduction and suicide risk among patients treated with placebo in antipsychotic clinical trials: an analysis of the Food and Drug Administration database. *Am J Psychiatry*. 2001;158:1449-1454.
12. Meltzer HY, Okayli G. Reduction of suicidality during clozapine treatment of neuroleptic-resistant schizophrenia: impact on risk-benefit assessment. *Am J Psychiatry*. 1995;152:183-190.
13. Walker AM, Lanza LL, Arellano F, Rothman KJ. Mortality in current and former users of clozapine. *Epidemiology*. 1997;8:671-677.
14. Reid WH, Mason M, Hogan T. Suicide prevention effects associated with clozapine therapy in schizophrenia and schizoaffective disorder. *Psychiatr Serv*. 1998;40:1029-1033.
15. Munro J, O'Sullivan D, Andrews C, Arana A, Mortimer A, Kerwin R. Active monitoring of 12,760 clozapine recipients in the UK and Ireland: beyond pharmacovigilance. *Br J Psychiatry*. 1999;175:576-580.
16. Reinstein MJ, Chasonov MA, Colombo KD, Jones LE, Sonnenberg JG. Reduction in suicide inpatients with schizophrenia receiving clozapine. *Clin Drug Invest*. 2002;22:341-346.
17. Sernyak MJ, Desai R, Stolar M, Rosenheck R. Impact of clozapine on completed suicide. *Am J Psychiatry*. 2001;158:931-937.
18. Meltzer HY. Treatment of suicidality in schizophrenia. *Ann N Y Acad Sci*. 2001;932:44-60.
19. Tran PV, Hamilton SH, Kuntz AJ, Potvin JH, Andersen SW, Beasley C Jr, Tollefson GD. Double-blind comparison of olanzapine versus risperidone in the treatment of schizophrenia and other psychotic disorders. *J Clin Psychopharmacol*. 1997;17:407-418.
20. Lindenmayer JP, Czobor P, Alphas L, et al, for the InterSePT Study Group. The InterSePT Scale for Suicidal Thinking (ISST) reliability and validity. *Schizophr Res*. In press.
21. Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *JAMA*. 1989;84:1065-1073.
22. Allison PD. *Survival Analysis Using the SAS® System: A Practical Guide*. Cary, NC: SAS Institute Inc; 1995.
23. Collett D. *Modeling Survival Data in Medical Research*. London, England: Chapman & Hall; 1994.
24. Altman DG. Confidence intervals for the number needed to treat. *BMJ*. 1998;317:1309-1312.
25. Radomsky ED, Haas GL, Mann JJ, Sweeney JA. Suicidal behavior in patients with schizophrenia and other psychotic disorders. *Am J Psychiatry*. 1999;156:1590-1595.
26. Suppes T, Webb A, Paul B, Carmody T, Kraemer H, Rush AJ. Clinical outcome in a randomized 1-year trial of clozapine versus treatment as usual for patients with treatment-resistant illness and a history of mania. *Am J Psychiatry*. 1999;156:1164-1169.
27. Ciapparelli A, Dell'Osso L, Pini S, Chiavacci MC, Fenzi M, Cassano GB. Clozapine for treatment-refractory schizophrenia, schizoaffective disorder, and psychotic bipolar disorder: a 24-month naturalistic study. *J Clin Psychiatry*. 2000;61:329-334.
28. Sajatovic M, Bingham CR, Garver D, Ramirez LF, Ripper G, Blow F, Lehmann LS. An assessment of clinical practice of clozapine therapy for veterans. *Psychiatr Serv*. 2000;51:669-671.
29. Levin A. Conference focuses on mental health stigma. *Psychiatric News*. 2001;36(9):8.
30. Singh BS. Suicide: the public health crisis of our time. *Aust N Z J Med*. 1998;28:295-300.

# Design and Rationale of the Paliperidone Palmitate Research in Demonstrating Effectiveness (PRIDE) Study: A Novel Comparative Trial of Once-Monthly Paliperidone Palmitate Versus Daily Oral Antipsychotic Treatment for Delaying Time to Treatment Failure in Persons With Schizophrenia

Larry Alphas, MD, PhD; Lian Mao, PhD; Stephen C. Rodriguez, MS; Joe Hulihan, MD; and H. Lynn Starr, MD

## ABSTRACT

**Background:** Public health considerations require that clinical trials address the complex “real-world” needs of patients with chronic illnesses. This is particularly true for persons with schizophrenia, whose management is frequently complicated by factors such as comorbid substance abuse, homelessness, and contact with the criminal justice system. In addition, barriers to obtaining health care in the United States often prevent successful community reentry and optimal patient management. Further, nonadherence to treatment is common, and this reinforces cycles of relapse and recidivism.

Long-acting injectable antipsychotic therapy may facilitate continuity of treatment and support better outcomes, particularly in patients who face these challenges. Clinical trials with classical explanatory designs may not be the best approaches for evaluating these considerations. We describe the design and rationale of a novel trial that combines both explanatory and pragmatic design features and studies persons with schizophrenia who face these challenges.

**Design and Rationale:** The Paliperidone Palmitate Research in Demonstrating Effectiveness (PRIDE) study is a prospective, open-label, randomized, 15-month study conducted between May 5, 2010, and December 9, 2013, comparing long-acting injectable paliperidone palmitate and oral antipsychotic medications in subjects with schizophrenia (according to *DSM-IV* criteria). Investigators and subjects had broad flexibility for treatment decision-making, thus making it a model that better reflects real-world practice. The primary end point was time to treatment failure, defined as arrest/incarceration psychiatric hospitalization; suicide; treatment discontinuation or supplementation due to inadequate efficacy, safety, or tolerability; or increased psychiatric services to prevent hospitalization. This end point was adjudicated by a blinded event monitoring board. Patients were followed to the 15-month end point, regardless of whether they were maintained on their initial randomized treatment. This article provides some of the reasoning behind the authors’ choices when combining features from both explanatory and pragmatic approaches to this trial’s design.

**Conclusions:** The PRIDE study incorporates real-world design features in a novel, prospective, comparative study of long-acting injectable and oral antipsychotics in persons with schizophrenia who have had recent contact with the criminal justice system. Insights provided should help the reader to better understand the need for more real-world approaches for clinical studies and how a broader approach can better aid clinical treatment and public health decision-making.

**Trial Registration:** ClinicalTrials.gov identifier: NCT01157351

*J Clin Psychiatry* 2014;75(12):1388–1393

© Copyright 2014 Physicians Postgraduate Press, Inc.

Submitted: December 20, 2013; accepted May 28, 2014.

Online ahead of print: October 28, 2014 (doi:10.4088/JCP.13m08965).

Corresponding author: Larry Alphas, MD, PhD, Therapeutic Team Leader Psychiatry, Janssen Scientific Affairs, LLC, 1125 Trenton-Harbourton Rd–A32404, Titusville, NJ 08560 (lalphs@its.jnj.com).

To help patients with schizophrenia lead more productive lives and better fulfill their potential, their complex “real-world” needs must be better understood. Clinical trials that study these needs must inform an ever-widening group of stakeholders, including patients, clinicians, regulatory authorities, and health care payers. To address these diverse expectations, trials with real-world (pragmatic) designs are necessary that build upon foundational studies establishing treatment safety and efficacy (explanatory trials). Pragmatic trials require attention to distinctive methodological considerations. This article focuses on design considerations as they relate to a comparative study of antipsychotics in schizophrenia.

Trials designed for regulatory approval typically focus on explanatory considerations; that is, they concentrate on whether the particular drug under study is safe and effective. Such trials require careful selection of a study population that is otherwise medically healthy and is likely to be adherent to treatment.<sup>1,2</sup> Teams conducting these studies must be highly trained to provide reliable safety and efficacy information. To adequately capture the required information, assessment measures are often highly specialized and are applied more frequently than related assessments used in standard clinical practice.<sup>2</sup> Considerable attention is also given to reinforcing treatment adherence so as to enhance the likelihood that results can be attributed to the study drug.<sup>2</sup>

Such explanatory trials often leave many questions related to clinical practice unanswered. Pragmatic or real-world trials seek to address these questions. To do this, they require a different approach. Exclusionary criteria should be limited, so that study subjects better represent the diverse population expected to be exposed to the treatment.<sup>1</sup> Treatment providers should fully reflect the range of clinicians and staff with skill sets customarily available when the treatment is delivered. Comparative interventions



should include the range of options that are available to the treating clinician.<sup>2</sup> Outcome assessments should be unambiguous measures of response that are meaningful to both the patient and clinician.<sup>2</sup> In practice, few trials are purely explanatory or pragmatic, and many prospective clinical trials include a range of pragmatic and explanatory features.<sup>2</sup> These considerations were used to design a study comparing daily oral and long-acting injectable treatments for schizophrenia.

### RATIONALE

Symptoms of schizophrenia can be treated effectively with antipsychotic medication; however, poor adherence to prescribed treatment is one of the biggest challenges of managing the symptoms of schizophrenia and delaying time to relapse.<sup>3,4</sup> Long-acting injectable antipsychotics deliver therapeutic concentrations over several weeks, eliminating the need for daily dosing<sup>5</sup> and providing clinicians with certain knowledge of adherence or nonadherence. As a result, these agents increase the likelihood of continuous and effective treatment and may reduce patients' risk for relapse. This, in turn, could decrease the likelihood of institutionalization in hospitals and incarceration. Studies comparing long-acting injectable versus oral antipsychotic treatment have provided inconsistent results,<sup>6-15</sup> with some indication that demonstrating a differential effectiveness among these formulations is better established with a more pragmatic clinical trial design than with one that is more explanatory.<sup>16-19</sup> With this in mind, we designed the Paliperidone Palmitate Research in Demonstrating Effectiveness (PRIDE) study (ClinicalTrials.gov identifier: NCT01157351) to compare once-monthly paliperidone palmitate and daily oral antipsychotics in real-world schizophrenia, as defined by subject inclusion criteria, treatment, and outcomes. It was hypothesized that paliperidone palmitate would be more effective than oral antipsychotics. We provide a description of that trial's design and the reasoning behind it, addressing the selection of patients, outcome measures, and study end points.

### STUDY OBJECTIVES

The primary objective of the PRIDE study was to compare the effectiveness of paliperidone palmitate treatment with daily oral antipsychotic treatment in delaying time to treatment failure (as defined by several real-world outcomes) over 15 months in subjects with schizophrenia.

Key secondary objectives were to compare paliperidone palmitate with oral antipsychotic treatment in (1) time to first psychiatric hospitalization or arrest/incarceration; (2) overall patient functioning, as measured by the Personal and Social Performance Scale<sup>20</sup>; (3) time to first psychiatric hospitalization; and (4) overall symptom improvement, as measured by the Clinical Global Impressions-Severity of Illness scale.<sup>21</sup> An additional objective was to examine the safety and tolerability of paliperidone palmitate treatment compared with oral antipsychotic treatment.

- Paliperidone palmitate may provide benefits over oral antipsychotics, but explanatory-designed trials have not characterized the real-world outcomes resulting from use of these agents.
- Pragmatic trials, with their limited exclusionary criteria, wide range of treatment providers, broad treatment options, and naturalistic outcome assessments, are designed to better reflect real-world situations and may lead to more practical understanding of treatment decision choices.
- The Paliperidone Palmitate Research in Demonstrating Effectiveness (PRIDE) study combines both explanatory and pragmatic design elements to examine the relative effectiveness of paliperidone palmitate, a once-monthly antipsychotic, and daily oral antipsychotics in patients with schizophrenia and a history of recent criminal justice system contact.

### STUDY DESIGN

The PRIDE study was a randomized, prospective, open-label, active-controlled, parallel-group comparative efficacy and effectiveness study of paliperidone palmitate versus oral antipsychotic treatment in adults with schizophrenia. The study was conducted between May 5, 2010, and December 9, 2013. It consisted of a screening phase of up to 2 weeks, followed by a 15-month randomized, open-label treatment phase (Figure 1). An independent event monitoring board, blinded to individual subject treatment assignment, certified the occurrence and time of the first treatment failure (the primary end point) for each randomly assigned subject.

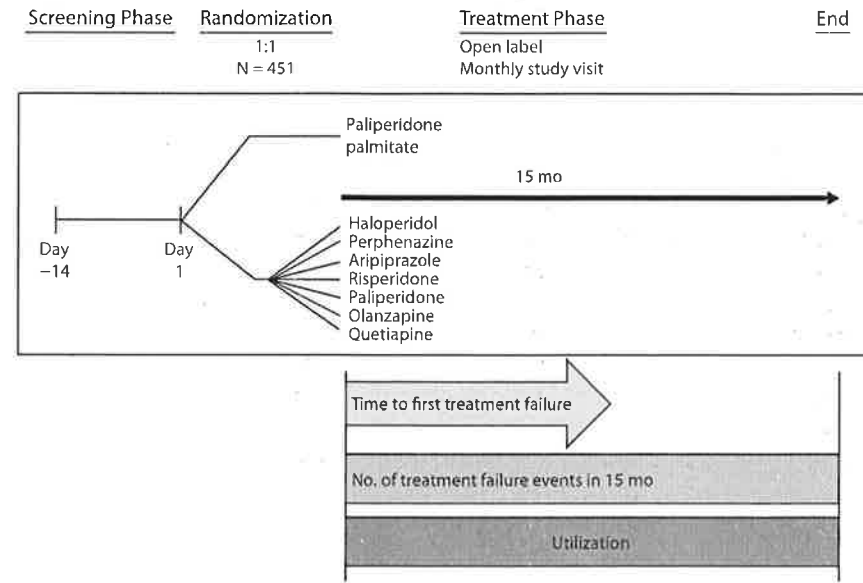
Incorporating both explanatory (efficacy) and pragmatic (effectiveness) design elements, the study was randomized and controlled but was conducted in the context of a naturalistic treatment setting rather than in a highly controlled clinical trial environment. This required balancing often competing considerations regarding pragmatic or explanatory choices for study design elements. The final design allowed information to be gathered on both efficacy and effectiveness (see Table 1 for detail) outcomes. Design domains that characterize a study along the pragmatic-explanatory continuum include patient selection criteria, site and investigator selection criteria, flexibility in dosing and use of concomitant medications, outcome selection, intensity of follow-up, and practices related to treatment adherence and patient retention.<sup>2,12</sup> Study sites and investigators were selected on the basis of access to subjects who fit the entry criteria and ability to follow subjects in a clinical trial setting.

Reasoning behind the choices surrounding some of the more important aspects of the trial design is discussed in more detail below.

### Study End Points

The PRIDE study included some end points that were more explanatory in nature and others that were more pragmatic. Because the trial was designed to meet standards

**Figure 1. Overview of the Paliperidone Palmitate Research in Demonstrating Effectiveness (PRIDE) Study**



**Table 1. Study Design Features**

| Design Feature                                 | Rationale   |
|--|---|
| Randomized                                     | Reduces potential sources of bias by balancing confounding sources of known and unrecognized bias.  |
| Study population with history of incarceration | Includes persons with a diagnosis of schizophrenia and a recent history of at least 1 incarceration. This is a substantial proportion of the schizophrenia population and previously has not been well studied.   |
| No stabilization period before randomization   | In real-world practice, prior stabilization of patients on medication is not possible; therefore, this study is designed to compare treatments without prior stabilization.   |
| Medication adherence                           | Real-world practice was mimicked. Patients randomized to oral treatment were given prescriptions to be filled (at no cost) at a local pharmacy. Patients randomized to paliperidone palmitate were given injections by an injection nurse at the site. No pill counts or other measures of adherence were made. |
| Duration of follow-up                          | The study has been powered such that an adequate number of treatment failure events would be expected to be observed after 15 months of follow-up. This consistent, long-term follow-up permits better understanding of real-world outcomes in a defined time period.   |
| Visit schedule                                 | Clinic visit frequency reflects naturalistic treatment, with similar visit structure for frequency, duration, intensity, and content relative to those seen in standard clinical practice.  |
| No placebo comparison                          | Use of placebo is not ethical or practical for long-term studies. An active comparator is the most clinically relevant comparison.  |
| Open label                                     | Eliminates the need for complex, double-dummy designs that are not naturalistic. Allows for decision-making that reflects real-world practice.  |
| Treatment failure end point                    | Captures a comprehensive, contemporary set of clinical outcomes reflecting failure of treatment under real-world circumstances.   |

for regulatory submission, it could not be purely pragmatic. Patients enrolled in the PRIDE study were encouraged to continue in the study to their predefined, 15-month completion date, even after a change from their initial, randomized treatment assignment because, as in real-life, events that occur after early discontinuation may be extremely relevant to the patient's predefined outcome.

The primary end point and the key secondary end points examined treatment response only while subjects were

taking their randomly assigned medication. This condition represented an explanatory approach and permits specific understanding of the relative safety and efficacy of the assigned treatments. The exploratory pragmatic end points examined treatment to the 15-month end point or the final recorded observation, regardless of whether subjects were maintained on the initial randomized treatment. This permitted understanding of the longer term consequences of the choice among treatments assigned at randomization.

The primary study end point, treatment failure, was developed as a pragmatic construct that is more relevant to the experience of patients with schizophrenia than other assessments. In particular, it incorporated arrest or incarceration as a component of treatment failure—an important pragmatic aspect of contemporary mental illness outcomes in the United States. Treatment failure was defined as 1 of the following: arrest/incarceration, psychiatric hospitalization, or suicide; or, as determined by the study physician, discontinuation of antipsychotic treatment due to inadequate efficacy, treatment supplementation with another antipsychotic due to inadequate efficacy, discontinuation of antipsychotic treatment due to safety or tolerability concerns, or an increase in the level of psychiatric services to prevent imminent psychiatric hospitalization.

Precise definitions were developed for all major elements of the treatment failure end point. An arrest was defined as the taking of a subject into custody by a legal authority for any reason. The definition did not include times when a subject was stopped, questioned, or temporarily detained by a law enforcement officer or by a preplanned, probation-associated, or court-ordered contact with the criminal justice system.

### Study Subjects

Because a more pragmatic approach was used to better reflect the broad range of patients found in regular clinical practice, few inclusionary or exclusionary entry criteria were applied beyond the primary intention to focus on persons with schizophrenia who had had recent contact with the criminal justice system. The few additional inclusion/exclusion criteria included were those required by the ethics of conducting a clinical trial and those required to capture specific efficacy and safety data (some of which are required by regulatory authorities) in an effort to understand the relative response to comparator treatments.

Selection criteria for investigators included their knowledge and connections with their local criminal justice system. Field-based medical staff from the sponsor worked with investigators to further develop an understanding of their local criminal justice system landscape.

Upon study start, it was found that traditional clinical trial recruitment efforts were ineffective. As a result, a recruitment outreach strategy was developed focusing on streets, homeless shelters, and single-residence units. Additional alternative stakeholders, such as law enforcement, case managers, and behavioral health departments, were included in the outreach process.

It was planned that approximately 442 male and female subjects between 18 and 65 years of age with schizophrenia (diagnosed according to *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition criteria,<sup>22</sup> confirmed by using the Mini-International Neuropsychiatric Interview, version 6.0<sup>23</sup>) who had been taken into custody at least twice in the previous 2 years would be enrolled in this study. At least 1 instance of custody must have led to an incarceration, and the most recent release from custody must have occurred within 90 days of the screening visit.

Subjects were excluded if they had been actively abusing intravenous drugs within the past 3 months or had an opiate dependence disorder. Otherwise, substance abuse was not an exclusionary factor. This approach permitted a compromise between being highly inclusive and ensuring the subjects would actually be available later for data collection. It also allowed many subjects to be enrolled in the PRIDE study who would have been excluded from more traditional explanatory studies.

### Treatment

For subjects taking more than 1 oral antipsychotic at screening, 1 was chosen by the investigator as the primary oral antipsychotic. Any additional antipsychotics were tapered and discontinued during the first 3 days of screening so that all subjects received only 1 oral antipsychotic on the day before randomization.

Before randomization, the treating clinician and the subject reviewed the 7 oral antipsychotics available for use in this study (aripiprazole, haloperidol, olanzapine, paliperidone, perphenazine, quetiapine, and risperidone) to determine their acceptability on the basis of prior experience with these medications. One or more (up to 6) could be deselected. Thereafter, subjects with the same set of prespecified oral antipsychotics were placed within similar randomization strata. At visit 1, subjects were randomly assigned within their stratum in a 1:1 ratio to paliperidone palmitate or oral antipsychotic treatment. If the subject was randomly assigned to receive oral antipsychotic treatment, the specific oral medication was randomly selected from the prespecified oral antipsychotics. This randomization approach was chosen over a simpler clinician choice design to reduce treatment selection bias among the oral antipsychotic treatments available.<sup>24</sup>

In general, dosing followed label instructions for the assigned treatments, but monotherapy was required between day 8 and day 15 so as to ensure that this approach was tried. Although supplemental oral antipsychotic treatment was allowed on or after day 15, the investigator was asked to first consider increasing the randomly assigned study drug dosage or adding adjunctive nonantipsychotic psychotropic therapy (ie, anxiolytics, antidepressants, or mood stabilizers) to manage worsening symptoms. This strategy encouraged use of optimal monotherapy for the primary assigned treatment.

Subjects in the paliperidone palmitate arm who were not taking 234 mg of paliperidone palmitate could receive supplemental oral paliperidone (if the investigator deemed a higher dose to be necessary) until the dose of paliperidone palmitate could be increased at the next injection day without declaration of a treatment failure. The questions addressed by this study related to relative comparisons of paliperidone palmitate and oral antipsychotic medications and hypothesized superiority for paliperidone palmitate. Therefore, subjects were allowed to switch from once-monthly paliperidone palmitate to oral antipsychotics, as it would have allowed the study hypotheses to be evaluated for both

the explanatory and pragmatic outcomes. However, subjects were not allowed to switch from a daily oral antipsychotic to once-monthly paliperidone palmitate and still remain in the study because, given the study hypothesis that paliperidone palmitate was superior to oral antipsychotic, this would not have allowed the pragmatic question regarding relative superiority of paliperidone palmitate to oral antipsychotic to be fully addressed.

### Clinical Assessments and Evaluations

During the screening period, psychiatric/medical histories were obtained, diagnostic criteria for schizophrenia ascertained, safety screening procedures performed, and other eligibility criteria evaluated. Throughout the 15-month treatment period, visits occurred on a similar schedule for both treatment arms on a monthly basis. Subjects were assessed at each study visit for occurrence of treatment failure.

To the extent possible, all subjects were followed for 15 months if they consented to ongoing participation. Subjects who experienced treatment failure had the option of continuing to take their randomly assigned treatment or changing to a new oral treatment if they did not find the randomized medication tolerable or adequately effective. Such subjects returned for regularly scheduled visits and underwent all assessments, including assessments for treatment failure. If a subject discontinued drug treatment, end-of-treatment procedures were completed as soon as possible thereafter. Subjects who left the study before reaching their 15-month end point were allowed to reenter the study any number of times until 15 months from randomization had elapsed.

### Statistical Analysis

Sample size determination was based on testing the primary null explanatory hypothesis that there is no difference between paliperidone palmitate and oral antipsychotic treatment in distribution of time to first treatment failure. Assuming time to first treatment failure follows an exponential distribution, the primary null and alternative hypothesis could be expressed in terms of the hazard ratio of the 2 treatment groups. For detecting treatment differences measured by a hazard ratio of 0.516 with 80% power, at a 2-sided .05 significance level, using an exponential maximum likelihood test of equality of survival curves, at least 72 first-treatment failures would have been needed. If it is assumed that the maximum follow-up time is 15 months and 30% of the randomly assigned subjects would drop out by 15 months before experiencing a treatment failure event (ie, common exponential dropout rate of 0.0238), a total of 442 subjects were required (221 per group). The hazard ratio of 0.516 corresponds to event rate differences ranging from 10% to 20%. This difference was judged to be clinically relevant.

For this study, the intent-to-treat (ITT) population was defined as all randomly assigned subjects who received at least 1 dose of their assigned study treatment. The

explanatory ITT (eITT) analysis set for the primary efficacy end point was defined as time to first treatment failure observed before the eITT end point (last injection date + 28 days, or last prescription date of the randomly assigned oral medication + number of days' supply + 1 day) for all ITT subjects. First treatment failure times for subjects who did not experience any treatment failure before the eITT end point were censored at the eITT end point, and treatment differences were compared using a log rank test based on the eITT analysis set.

### CONCLUSIONS

The PRIDE study is a 15-month, prospective, randomized, active-controlled, open-label comparative trial, conducted in a naturalistic, real-world setting. It is designed to compare once-monthly paliperidone palmitate with daily oral antipsychotic treatment in delaying time to treatment failure in adults with schizophrenia who have had recent contact with the criminal justice system. The primary end point for the PRIDE study is novel, as it encompasses criminal justice contact as a component of treatment failure. The design allows both explanatory and pragmatic questions to be addressed in the context of a single study. The explanatory end points should support prior findings with paliperidone palmitate in an important segment of the population of persons with schizophrenia relative to alternative oral treatments. The choice of time to treatment failure and the more pragmatic aspects of the trial permit a broad understanding of the effectiveness of paliperidone palmitate when used under standard practice conditions compared with a range of treatment alternatives. Indeed, the inclusion of time to arrest as part of the primary end point incorporates a measure that is relevant to the real-world experience of many persons with schizophrenia. This end point is usually neglected but has important implications for individuals, families, and broader public health considerations.

The findings from this study will permit a deeper understanding of differences between the 2 primary treatments studied (paliperidone palmitate and oral antipsychotics) and the longer term consequences of making treatment decisions, regardless of the resultant treatment conditions. This knowledge should help inform better treatment decisions and generate better public health policy.

**Drug names:** aripiprazole (Abilify), haloperidol (Haldol and others), olanzapine (Zyprexa and others), paliperidone (Invega), paliperidone palmitate (Invega Sustenna), quetiapine (Seroquel and others), risperidone (Risperdal and others).

**Author affiliations:** Scientific Affairs, Janssen Scientific Affairs, LLC (Drs Alphs and Starr and Mr Rodriguez), Titusville; Biostatistics and Reporting Department, Janssen R & D, LLC (Dr Mao), Titusville; and Global Medical Affairs, Janssen Global Services, LLC (Dr Hulihan), Raritan, New Jersey.

**Potential conflicts of interest:** Drs Alphs and Starr and Mr Rodriguez are employees of Janssen Scientific Affairs and Johnson & Johnson stockholders. Dr Mao is an employee of Janssen Research & Development and a Johnson & Johnson stockholder. Dr Hulihan is an employee of Janssen Global Services and is a Johnson & Johnson stockholder.

**Funding/support:** This research was funded by Janssen Scientific Affairs, LLC, Titusville, New Jersey (primary study protocol number R092670-SCH3006).

**Previous presentations:** 13th International Congress on Schizophrenia Research; April 2–6, 2011; Colorado Springs, Colorado • American Society for Experimental NeuroTherapeutics 13th Annual Meeting; February 24–26, 2011; Bethesda, Maryland • American College of Neuropsychopharmacology 49th Annual Meeting; December 5–9, 2010; Miami Beach, Florida • International Society for CNS Clinical Trials and Methodology Autumn Conference; October 13–14, 2010; Baltimore, Maryland • and National Conference on Correctional Health Care Annual Meeting; October 9–13, 2010; Las Vegas, Nevada.

**Acknowledgments:** The authors wish to acknowledge Reuven Ferziger, MD (employee of Janssen Scientific Affairs, LLC, at the time of the study); Ed Crumbley, BA (employee of Janssen Scientific Affairs, LLC, at the time of the study); and Norris Turner, PharmD, PhD (employee of Johnson & Johnson Health Care Systems), for their participation during the development of the study design. The authors also thank Cynthia Bossie, PhD (employee of Janssen Scientific Affairs, LLC), and Matthew Grzywacz, PhD, of ApotheCom, LLC, Yardley, Pennsylvania (funded by Janssen Scientific Affairs, LLC), for providing writing and editorial assistance. Drs Ferziger, Turner, and Bossie and Mr Crumbley are Johnson & Johnson stockholders.

### REFERENCES

- Roland M, Torgerson DJ. What are pragmatic trials? *BMJ*. 1998;316(7127):285.
- Thorpe KE, Zwarenstein M, Oxman AD, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *CMAJ*. 2009;180(10):E47–E57.
- Morken G, Widen JH, Grawe RW. Non-adherence to antipsychotic medication, relapse and rehospitalisation in recent-onset schizophrenia. *BMC Psychiatry*. 2008;8(1):32.
- Kane JM. Improving treatment adherence in patients with schizophrenia. *J Clin Psychiatry*. 2011;72(9):e28.
- Gopal S, Hough DW, Xu H, et al. Efficacy and safety of paliperidone palmitate in adult patients with acutely symptomatic schizophrenia: a randomized, double-blind, placebo-controlled, dose-response study. *Int Clin Psychopharmacol*. 2010;25(5):247–256.
- Chue P, Eerdeken M, Augustyns I, et al. Comparative efficacy and safety of long-acting risperidone and risperidone oral tablets. *Eur Neuropsychopharmacol*. 2005;15(1):111–117.
- Bai YM, Chen TT, Wu B, et al. A comparative efficacy and safety study of long-acting risperidone injection and risperidone oral tablets among hospitalized patients: 12-week randomized, single-blind study. *Pharmacopsychiatry*. 2006;39(4):135–141.
- Emsley R, Oosthuizen P, Koen L, et al. Oral versus injectable antipsychotic treatment in early psychosis: post hoc comparison of two studies. *Clin Ther*. 2008;30(12):2378–2386.
- Weiden PJ, Schooler NR, Weedon JC, et al. A randomized controlled trial of long-acting injectable risperidone vs continuation on oral atypical antipsychotics for first-episode schizophrenia patients: initial adherence outcome. *J Clin Psychiatry*. 2009;70(10):1397–1406.
- Macfadden W, Ma YW, Thomas Haskins J, et al. A prospective study comparing the long-term effectiveness of injectable risperidone long-acting therapy and oral aripiprazole in patients with schizophrenia. *Psychiatry (Edmont)*. 2010;7(11):23–31.
- Gaebel W, Schreiner A, Bergmans P, et al. Relapse prevention in schizophrenia and schizoaffective disorder with risperidone long-acting injectable vs quetiapine: results of a long-term, open-label, randomized clinical trial. *Neuropsychopharmacology*. 2010;35(12):2367–2377.
- Rosenheck RA, Krystal JH, Lew R, et al; CSP 555 Research Group. Challenges in the design and conduct of controlled clinical effectiveness trials in schizophrenia. *Clin Trials*. 2011;8(2):196–204.
- Schooler NR, Buckley PF, Mintz J, et al. PROACTIVE: initial results of an RCT comparing long-acting injectable risperidone to 2nd generation oral antipsychotics. Presented at the 50th Annual Meeting of the American College of Neuropsychopharmacology; December 4–8, 2011; Kona, Hawaii: S104–S105.
- Zhornitsky S, Stip E. Oral versus long-acting injectable antipsychotics in the treatment of schizophrenia and special populations at risk for treatment nonadherence: a systematic review. *Schizophr Res Treatment*. 2012;2012:407171.
- Barrio P, Batalla A, Castellví P, et al. Effectiveness of long-acting injectable risperidone versus oral antipsychotics in the treatment of recent-onset schizophrenia: a case-control study. *Int Clin Psychopharmacol*. 2013;28(4):164–170.
- Kirson NY, Weiden PJ, Yermakov S, et al. Efficacy and effectiveness of depot versus oral antipsychotics in schizophrenia: synthesizing results across different research designs. *J Clin Psychiatry*. 2013;74(6):568–575.
- Kishimoto T, Nitta M, Borenstein M, et al. Long-acting injectable versus oral antipsychotics in schizophrenia: a systematic review and meta-analysis of mirror-image studies. *J Clin Psychiatry*. 2013;74(10):957–965.
- Kishimoto T, Robenzadeh A, Leucht C, et al. Long-acting injectable vs oral antipsychotics for relapse prevention in schizophrenia: a meta-analysis of randomized trials. *Schizophr Bull*. 2014;40(1):192–213.
- Fusar-Poli P, Kempton MJ, Rosenheck RA. Efficacy and safety of second-generation long-acting injections in schizophrenia: a meta-analysis of randomized-controlled trials. *Int Clin Psychopharmacol*. 2013;28(2):57–66.
- Morosini PL, Magliano L, Brambilla L, et al. Development, reliability and acceptability of a new version of the DSM-IV Social and Occupational Functioning Assessment Scale (SOFAS) to assess routine social functioning. *Acta Psychiatr Scand*. 2000;101(4):323–329.
- Guy W. *ECDEU Assessment Manual for Psychopharmacology*. US Dept Health, Education, and Welfare publication (ADM) 76-338. Rockville, MA: National Institute of Mental Health;1976:218–222.
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition. Washington, DC: American Psychiatric Association; 1994.
- Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry*. 1998;59(suppl 20):22–33, quiz 34–57.
- Lavori PW, Rush AJ, Wisniewski SR, et al. Strengthening clinical effectiveness trials: equipoise-stratified randomization. *Biol Psychiatry*. 2001;50(10):792–801.

# Real-World Outcomes of Paliperidone Palmitate Compared to Daily Oral Antipsychotic Therapy in Schizophrenia: A Randomized, Open-Label, Review Board–Blinded 15-Month Study

Larry Alphas, MD, PhD; Carmela Benson, MS, MSHP; Kimberly Cheshire-Kinney, BA; Jean-Pierre Lindenmayer, MD; Lian Mao, PhD; Stephen C. Rodriguez, MS; and H. Lynn Starr, MD

## ABSTRACT

**Objective:** The Paliperidone Palmitate Research in Demonstrating Effectiveness (PRIDE) study compared the effects of once-monthly paliperidone palmitate with daily oral antipsychotics on treatment failure in adults with schizophrenia.

**Method:** The PRIDE study is a 15-month, randomized, multicenter study (May 5, 2010, to December 9, 2013) of adult subjects with a *DSM-IV* diagnosis of schizophrenia and a history of incarceration. Subjects were randomly assigned to once-monthly paliperidone palmitate injections or daily oral antipsychotics (randomly assigned from 7 acceptable, prespecified oral antipsychotics) for 15 months. The primary end point was time to first treatment failure, defined as arrest/incarceration; psychiatric hospitalization; suicide; treatment discontinuation or supplementation due to inadequate efficacy, safety, or tolerability; or increased psychiatric services to prevent hospitalization. Time to first treatment failure was determined by a blinded event-monitoring board and analyzed with the Kaplan-Meier method.

**Results:** In this study, 450 patients were randomly assigned, and 444 were included in the intent-to-treat population. Paliperidone palmitate was associated with significant delay in time to first treatment failure versus oral antipsychotics (hazard ratio, 1.43; 95% CI, 1.09–1.88; log rank  $P = .011$ ). Observed treatment failure rates over 15 months were 39.8% and 53.7%, respectively. Arrest/incarceration and psychiatric hospitalization were the most common reasons for treatment failure in the paliperidone palmitate and oral antipsychotic groups (21.2% vs 29.4% and 8.0% vs 11.9%, respectively). The 5 most common treatment-emergent adverse events for the paliperidone palmitate treatment group were injection site pain (18.6% of subjects), insomnia (16.8%), weight increased (11.9%), akathisia (11.1%), and anxiety (10.6%).

**Conclusions:** In a trial designed to reflect real-world management of schizophrenia, once-monthly paliperidone palmitate demonstrated superiority compared to oral antipsychotics in delaying time to treatment failure.

**Trial Registration:** Clinicaltrials.gov Identifier: NCT01157351

*J Clin Psychiatry* 2015;76(5):554–561

© Copyright 2015 Physicians Postgraduate Press, Inc.

Submitted: October 13, 2014; accepted February 27, 2015.

Online ahead of print: April 14, 2015 (doi:10.4088/JCP.14m09584).

Corresponding author: Larry Alphas, MD, PhD, Therapeutic Team Leader, Psychiatry, Janssen Scientific Affairs, LLC, 1125 Trenton-Harbourton Rd–A32404, Titusville, NJ 08560 (lalphs@its.jnj.com).

Schizophrenia is a serious chronic mental illness characterized by hallucinations, delusions, and significant functional disabilities that affects approximately 1.1% of adults in the United States.<sup>1,2</sup> Schizophrenia places a large economic burden on the health care system, resulting in estimated direct annual medical costs of \$43 to \$58 billion<sup>3</sup> and a significant societal impact in terms of overall health care burden.<sup>4</sup>

For individuals with schizophrenia to fulfill their potential and lead more meaningful lives their real-world treatment needs must be better understood and addressed. At present, effective management of schizophrenia is complicated by a variety of factors, including contacts with the criminal justice system, multiple hospitalizations, comorbid substance abuse, challenges to treatment adherence, unemployment, and unstable living conditions.<sup>5–7</sup> Most clinical trials select for individuals who are not broadly representative of patients from these real-world settings,<sup>8–10</sup> limiting the generalizability of their results. Their broad applicability is further complicated by the frequent choice of scale-based rather than clinically defined end points.<sup>11,12</sup> A consequence of such designs is that they fail to evaluate many of the complex issues associated with the daily management of schizophrenia.

Poor treatment adherence is common among individuals with schizophrenia, particularly in patients with prior involvement with the criminal justice system or comorbid substance abuse.<sup>6,13,14</sup> Such problems with adherence is frequently a precursor to cycles of relapse and recidivism. Long-acting injectable (LAI) antipsychotic therapies can deliver therapeutic concentrations continuously over several weeks and eliminate the need for daily medication administration.<sup>15</sup> Further, their mode of administration provides physicians with certain knowledge of adherence. As a result, use of LAI antipsychotic therapy may facilitate continuity of treatment and support better outcomes. Despite these apparent advantages, clinical trials comparing LAI and oral antipsychotics have produced inconsistent results.<sup>16–21</sup> We hypothesize that the inconsistencies in these reports might be a consequence of the study designs chosen for these comparisons and, possibly, a failure to follow a broad spectrum of patients using measures that reflect an adequate breadth of real-world outcomes.<sup>22</sup> We describe a study, Paliperidone Palmitate Research in Demonstrating Effectiveness (PRIDE), that compares once-monthly paliperidone palmitate with daily oral antipsychotics in patients with schizophrenia who are at risk for relapse. The study was designed to reflect real-world management of schizophrenia, as defined by the patients included, and clinically meaningful outcome measures.

## METHOD

### Study Design

The PRIDE study is a randomized, prospective, open-label, event-monitoring board-blinded, parallel-group study that compared paliperidone palmitate and oral antipsychotics on treatment failure in subjects with schizophrenia. The study incorporated both explanatory (efficacy) and pragmatic (effectiveness) design elements, allowing documentation of efficacy and effectiveness outcomes. Conducted between May 5, 2010, and December 9, 2013, the study included a screening phase of up to 2 weeks, followed by a 15-month randomized, open-label treatment phase. It was registered on ClinicalTrials.gov (identifier: NCT01157351). More complete details of the study design have been previously published.<sup>23</sup>

### Participants

Participants were enrolled from 50 sites across 25 US states and Puerto Rico. To enhance enrollment of subjects often excluded from trials, efforts were made to recruit subjects from nontraditional locations, such as homeless shelters, soup kitchens, and jail-release or diversion programs. The study's major inclusion criteria enlisted adults aged 18 to 65 years with a current diagnosis of schizophrenia (*Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition [DSM-IV]* criteria<sup>24</sup> that was confirmed by the Mini-International Neuropsychiatric Interview [MINI], version 6.0<sup>25</sup>). Subjects must have been taken into custody by the criminal justice system  $\geq 2$  times in the previous 2 years, with  $\geq 1$  of these events leading to incarceration; released from most recent custody within 90 days of the screening visit; and accepting of a once-monthly, LAI antipsychotic. To maximize study retention, subjects designated a reliable external contact (eg, family member, case manager). Major exclusion criteria included use of either clozapine within 3 months of screening or an injectable antipsychotic within 2 injection cycles of screening. Substance abuse was not exclusionary, but subjects who had abused intravenous drugs within 3 months of screening or had an opiate dependence disorder (*DSM-IV*) were excluded. The study was approved by each site's institutional review board and was conducted in accordance with the Declaration of Helsinki. All subjects provided written informed consent.

### Interventions

Before random treatment assignment, clinicians, together with each subject, reviewed the 7 oral antipsychotics available in this study (aripiprazole, haloperidol, olanzapine, paliperidone, perphenazine, quetiapine, and risperidone) to determine their acceptability based on prior experience. Up to 6 medications could be deselected by the participant or physician.

### Randomization

To reduce treatment selection bias, an equipoise-stratified randomization scheme<sup>26</sup> was used for treatment assignment and implemented via an interactive voice response system.

- Most randomized clinical trials do not enroll patients with complex comorbidities/histories; thus, these trials may not broadly reflect important treatment considerations for persons with schizophrenia.
- The Paliperidone Palmitate Research in Demonstrating Effectiveness (PRIDE) study was designed to reflect real-world treatment of schizophrenia, as defined by patient selection, treatment approaches, and clinically meaningful outcomes.
- The PRIDE study results demonstrate that, by using a design reflecting real-world considerations, once-monthly treatment with paliperidone palmitate was more effective in delaying treatment failure than treatment with daily oral antipsychotics.

Subjects with the same oral antipsychotic selections were placed in the same randomization strata and assigned in a 1:1 ratio to treatment with flexibly dosed paliperidone palmitate (78–234 mg) or a flexibly dosed oral antipsychotic. For subjects assigned to the oral antipsychotics arm, the specific agent was randomly selected from the group of prespecified, acceptable oral antipsychotics. Any oral antipsychotic prescribed before randomization was tapered and discontinued over the first 8 days after randomization. Paliperidone palmitate injection was administered by staff at the clinical site. Patients randomized to oral antipsychotics received a prescription and a voucher to cover its cost. Prescriptions were filled at a local pharmacy.

### Study Medications

Subjects assigned to the paliperidone palmitate group were initiated with 2 injections in the deltoid muscle that were given approximately 1 week apart: 234 mg on day 1 and 156 mg on day 8 ( $\pm 4$  days). Flexible monthly maintenance doses of paliperidone palmitate within a range of 78–234 mg (50–150 mg equivalents; recommended target maintenance dose was 156 mg) were started on day 38. Doses of oral antipsychotic monotherapy were selected and adjusted within the dose range of the package insert (occasional dosing outside of package insert range was allowed). Nonantipsychotic psychotropic medications (ie, mood stabilizers, antidepressants, anxiolytics, or hypnotics) were allowed.

### Clinical Assessments and Outcome Measure

Throughout the 15-month treatment period, study visits occurred on days 8 ( $\pm 4$ ), 15 ( $\pm 3$ ), and 38 ( $\pm 7$ ), and monthly thereafter (every 30 [ $\pm 7$ ] days). Subjects were assessed at each visit for treatment failure. Subjects were encouraged to continue in the study to their predefined, 15-month end-of-observation date, regardless of early discontinuation of their randomized treatment assignment or achievement of the primary end point. A subject was considered to be a completer for the efficacy analysis if he or she either experienced a treatment failure event or completed the 15-month study follow-up.



The primary study end point was time to first treatment failure, as determined by an independent event-monitoring board that was blinded to individual subject treatment assignment. Treatment failure was defined as 1 of the following: arrest/incarceration, psychiatric hospitalization, suicide, discontinuation of antipsychotic treatment due to inadequate efficacy, treatment supplementation with another antipsychotic due to inadequate efficacy, discontinuation of antipsychotic treatment due to safety or tolerability concerns, or an increase in the level of psychiatric services to prevent imminent psychiatric hospitalization.

Prespecified key secondary efficacy end points listed in order of priority were time to first psychiatric hospitalization or arrest/incarceration, change in Personal and Social Performance Scale (PSP) scores, time to first psychiatric hospitalization, and change in Clinical Global Impressions-Severity of Illness scale (CGI-S) score. Safety assessments included monitoring of adverse events (AEs), vital signs, physical examinations, and clinical laboratory tests.

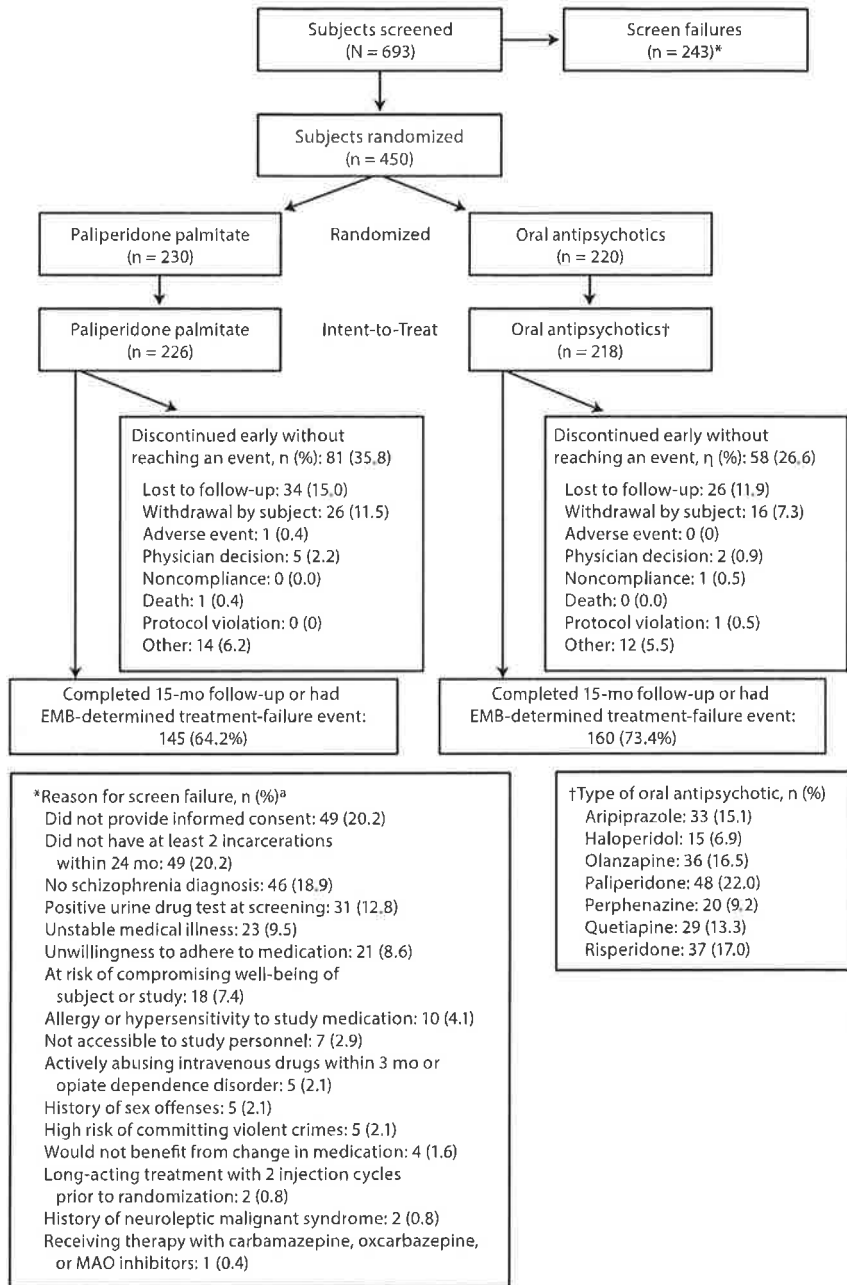
**Sample Size**

Sample size estimation was based on testing treatment group differences measured by hazard ratio (HR), using a 2-sided exponential maximum likelihood test at a .05 significance level. Detecting an HR (treatment/control) of 0.516 with 80% power requires a total of at least 72 treatment failure events. Assuming 30% of randomized subjects would drop out of the study before experiencing a treatment failure event, a total of 442 subjects (221 per group) would need to be randomized to achieve the required number of treatment failure events.

**Statistical Analysis**

The intent-to-treat (ITT) population, defined as all randomly assigned subjects who received ≥ 1 dose of their study treatment, was used for efficacy and safety analysis. To determine the relative effects of assigned treatments, an explanatory approach was used to assess the primary, key secondary, and safety end points. Primary and

**Figure 1. Study Flow**



<sup>a</sup>Percentages based on total number of screen failures (n=243). Screen failures could be due to more than 1 reason. Abbreviations: EMB = Event Monitoring Board, MAO = monoamine oxidase.

secondary analyses included all data from randomization until the end of randomly assigned treatment (28 days after the last injection of paliperidone palmitate or 1 day after the last dose of oral antipsychotic).

Demographic and baseline characteristics and AEs were summarized using descriptive statistics. Event-free probabilities of treatment failure and components of treatment failure were estimated by the Kaplan-Meier method. Treatment differences were compared using a log rank test. The HR and 95% confidence interval (CI) were estimated using a Cox proportional hazards model, with treatment as a fixed factor. Statistical significance was based on a 2-sided  $\alpha$  of .05. A mixed-model repeated-measures analysis of covariance, using an unstructured covariance matrix



with terms for treatment, time, treatment-by-time interaction, and baseline score, compared PSP total scores and CGI-S scores. To preserve the overall type I error rate at the 2-sided .05 significance level, the primary and key secondary hypotheses were tested using a fixed sequence gatekeeper approach. The hierarchy of the procedure was to test the primary hypothesis first at the .05 significance level, then to test the key secondary hypothesis at the .05 significance level. At each step, if the null hypothesis failed to be rejected ( $P \geq .05$ ), formal testing would be terminated and current and all subsequent null hypotheses would not be rejected. Testing of subsequent hypotheses continued, but the results were considered exploratory. All analyses were performed using SAS version 9.2 (SAS Institute Inc; Cary, North Carolina).

## RESULTS

### Patient Disposition and Baseline Characteristics

In this study, 693 subjects were screened, and 450 were randomly assigned (230 to paliperidone palmitate and 220 to oral antipsychotics; Figure 1); 444 subjects were included in the ITT population (paliperidone palmitate,  $n = 226$ ; oral antipsychotics,  $n = 218$ ). Overall, 305 subjects (68.7%) either had an event-monitoring board-identified treatment failure event or completed the 15-month study; 60 subjects (13.5%) were lost to follow-up, and 42 (9.5%) withdrew consent (Figure 1).

Baseline demographics and clinical characteristics did not differ significantly between arms (Table 1). Most subjects were male (86.3%) and black/African American (62.1%). The mean (SD) age was 38.1 (10.5) years, and the mean (SD) time since release from last incarceration was 42.2 (51.7) days. The majority of arrests prior to enrollment in the study were for non-violent or drug offenses; 54.2%, 37.8%, and 29.1% of subjects were previously arrested due to felonies, misdemeanors, or infractions, respectively. It should be noted that definitions for arrest-type classifications vary by jurisdiction. A total of 59.5% of subjects had comorbid substance abuse. Additionally, 16.1% of the oral antipsychotic group had been taking their randomly selected medication within 7 days of randomization. For the paliperidone palmitate group, 22.6% had been taking either paliperidone or risperidone within 7 days of randomization. Overall, 74.3% in the paliperidone palmitate treatment arm and 80.3% in the oral antipsychotic treatment arm used  $\geq 1$  concomitant psychotropic medication (including antipsychotics) during the study (see eTable 1 at [PSYCHIATRIST.COM](http://PSYCHIATRIST.COM)). Frequently used concomitant nonantipsychotic psychotropic medications for paliperidone palmitate versus oral antipsychotic groups included antidepressants (37.2% vs 41.3%), benzodiazepines (19.9% vs 22.5%), mood stabilizers/antiepileptics (17.3% vs 17.0%), and nonbenzodiazepine hypnotics/anxiolytics (14.6% vs 10.6%) (eTable 1).

**Table 1. Demographic and Baseline Characteristics (intent-to-treat population)**

| Characteristic   | Paliperidone Palmitate (n = 226) | Oral Antipsychotics (n = 218) |
|--|----------------------------------|-------------------------------|
| Age, mean (SD), y  | 37.7 (10.6)                      | 38.6 (10.4)                   |
| Male, n (%)  | 193 (85.4)                       | 190 (87.2)                    |
| Race, n (%)  | (n = 226)                        | (n = 217)                     |
| White  | 73 (32.3)                        | 74 (34.1)                     |
| Black/African American   | 145 (64.2)                       | 130 (60.0)                    |
| Other  | 8 (3.5)                          | 13 (6.0)                      |
| Ethnicity, n (%)   | (n = 216)                        | (n = 212)                     |
| Hispanic or Latino   | 31 (14.4)                        | 36 (17.0)                     |
| Not Hispanic or Latino   | 185 (85.6)                       | 176 (83.0)                    |
| BMI (kg/m <sup>2</sup> ), mean (SD)                              | (n = 225)<br>27.9 (5.6)          | (n = 218)<br>27.8 (5.0)       |
| Time since release from the last incarceration, mean (SD), d     | (n = 226)<br>38.9 (50.3)         | (n = 217)<br>45.7 (53.0)      |
| Duration of illness, n (%)                                       | (n = 226)                        | (n = 216)                     |
| $\leq 5$ y   | 42 (18.6)                        | 35 (16.2)                     |
| $> 5$ y  | 184 (81.4)                       | 181 (83.8)                    |
| No. of psychiatric hospitalizations in lifetime, mean (SD)       | (n = 176)<br>7.3 (16.4)          | (n = 170)<br>5.7 (5.6)        |
| No. of psychiatric hospitalizations in the past 12 mo, mean (SD) | (n = 176)<br>1.3 (7.6)           | (n = 173)<br>1.0 (1.5)        |
| Concurrent substance abuse (including alcohol), n (%)            | 130 (57.5)                       | 134 (61.5)                    |
| Type of arrest, n (%)  |                                  |                               |
| Infraction <sup>a</sup>  | 68 (30.1)                        | 61 (28.0)                     |
| Most common ( $\geq 10\%$ , either group)                        |                                  |                               |
| Violation of probation/parole                                    | 44 (19.5)                        | 39 (17.9)                     |
| Misdemeanor <sup>b</sup>   | 90 (39.8)                        | 78 (35.8)                     |
| Most common ( $\geq 10\%$ , either group)                        |                                  |                               |
| Vagrancy, public intoxication                                    | 26 (11.5)                        | 26 (11.9)                     |
| Felony <sup>c</sup>  | 108 (47.8)                       | 128 (58.7)                    |
| Most common ( $\geq 10\%$ , either group)                        |                                  |                               |
| Drug charges   | 34 (15.0)                        | 40 (18.3)                     |
| Assault (private citizen)  | 24 (10.6)                        | 28 (12.8)                     |
| Burglary/larceny/breaking and entering                           | 16 (7.1)                         | 22 (10.1)                     |
| Homelessness, <sup>d</sup> n (%)                                 | (n = 221)<br>28 (12.7)           | (n = 210)<br>34 (16.2)        |
| PSP total score, mean (SD)                                       | (n = 226)<br>54.8 (12.7)         | (n = 215)<br>54.9 (12.8)      |
| CGI-S score, mean (SD)   | (n = 226)<br>3.8 (0.8)           | (n = 217)<br>3.9 (0.7)        |

<sup>a</sup>Infractions are defined as a violation of a rule, ordinance, or regulation. They are considered minor crimes and are sometimes called petty crimes or summary offenses. They are punishable usually by a fine, rather than jail time; typically, these are local crimes related to traffic, parking, or noise violations.

<sup>b</sup>Misdemeanors are defined as lesser crimes (ie, do not rise to severity of a felony). Misdemeanors are considered crimes of low seriousness.

<sup>c</sup>Felonies are defined as the most serious classification of crimes. Both property crimes and person crimes are considered felonies. Persons committing the crime, as well as anyone who aided and abetted the felon before the crime, during the crime, or as an accessory to the crime after it was committed, can be charged with a felony.

<sup>d</sup>Homelessness is defined as living on the streets or in an emergency shelter for the homeless since the time of release from jail.

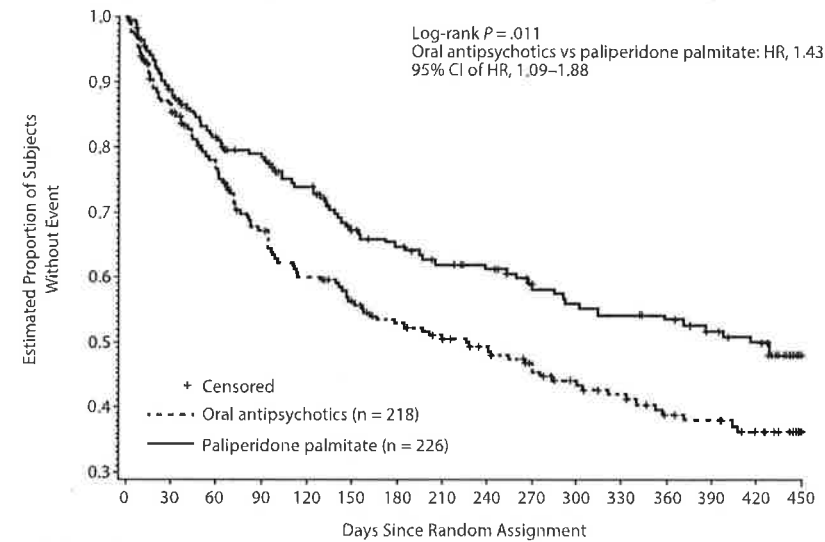
Abbreviations: BMI = body mass index, CGI-S = Clinical Global Impressions-Severity of Illness, PSP = Personal Social Performance Scale.

### Primary Outcome: First Treatment Failure

Ninety subjects (39.8%) in the paliperidone palmitate group and 117 subjects (53.7%) in the oral antipsychotic group had a treatment failure event. Paliperidone palmitate was superior to oral antipsychotics in delaying time to first treatment failure (HR, 1.43; 95% CI, 1.09–1.88;  $P = .011$ ) (Figure 2). Median times to first treatment failure were 416 and 226 days in the paliperidone palmitate and oral antipsychotic groups, respectively. The most common

**Figure 2. Kaplan-Meier Estimate of Time to First Treatment Failure (A) and Reasons for Treatment Failure (B)<sup>a</sup>**

**A. Time to First Treatment Failure**



| No. of subjects at risk |  |
|-------------------------|--|
| Oral antipsychotics     | 218 183 152 126 112 102 92 86 79 71 61 58 49 47 41 29  |
| Paliperidone palmitate  | 226 190 162 148 128 107 100 92 88 76 70 68 65 60 56 31 |

**B. Reason for First Treatment Failure**

| Reason  | Paliperidone Palmitate (n = 226), n (%) | Oral Antipsychotics (n = 218), n (%) |
|---|---|--------------------------------------|
| Any   | 90 (39.8)                               | 117 (53.7)                           |
| Arrest/incarceration  | 48 (21.2)                               | 64 (29.4)                            |
| Psychiatric hospitalization   | 18 (8.0)                                | 26 (11.9)                            |
| Discontinuation of antipsychotic treatment due to safety or tolerability                  | 15 (6.6)                                | 8 (3.7)                              |
| Treatment supplementation with another antipsychotic due to inadequate efficacy           | 5 (2.2)                                 | 6 (2.8)                              |
| Discontinuation of antipsychotic treatment due to inadequate efficacy                     | 1 (0.4)                                 | 9 (4.1)                              |
| Increase in level of psychiatric services to prevent imminent psychiatric hospitalization | 3 (1.3)                                 | 4 (1.8)                              |
| Suicide   | 0                                       | 0                                    |

<sup>a</sup>Data from randomization until end of randomly assigned treatment (28 days after last injection of paliperidone palmitate or 1 day after last dose of oral antipsychotic). Abbreviation: HR = hazard ratio.

reasons for first treatment failure were arrest/incarceration (21.2% vs 29.4%) and psychiatric hospitalization (8.0% vs 11.9%) (Figure 2). No suicides were reported.

**Secondary Outcomes**

**Time to first psychiatric hospitalization or arrest/incarceration.** Seventy-six subjects (33.6%) in the paliperidone palmitate group and 98 subjects (45.0%) in the oral antipsychotic group had a psychiatric hospitalization or arrest/incarceration as a first treatment failure event. Paliperidone palmitate was superior to oral antipsychotics in delaying time to first psychiatric hospitalization or arrest/incarceration (HR, 1.43; 95% CI, 1.06–1.93; *P* = .019) (Figure 3). Median time to first psychiatric hospitalization or arrest/incarceration was not reached in the paliperidone palmitate group (>450 days) and was 274 days in the oral antipsychotic group.

**Personal and Social Performance Scale.** No significant between-group differences were observed in mean change in PSP total scores (least squares mean [standard error (SE)] difference = 0.39 [0.98]; *P* = .689). Subsequent analyses of other secondary efficacy variables were considered exploratory.

**CGI-S.** No significant between-group differences were observed in mean change in CGI-S scores (least squares mean [SE] difference = -0.06 [0.05]; nominal *P* = .296).

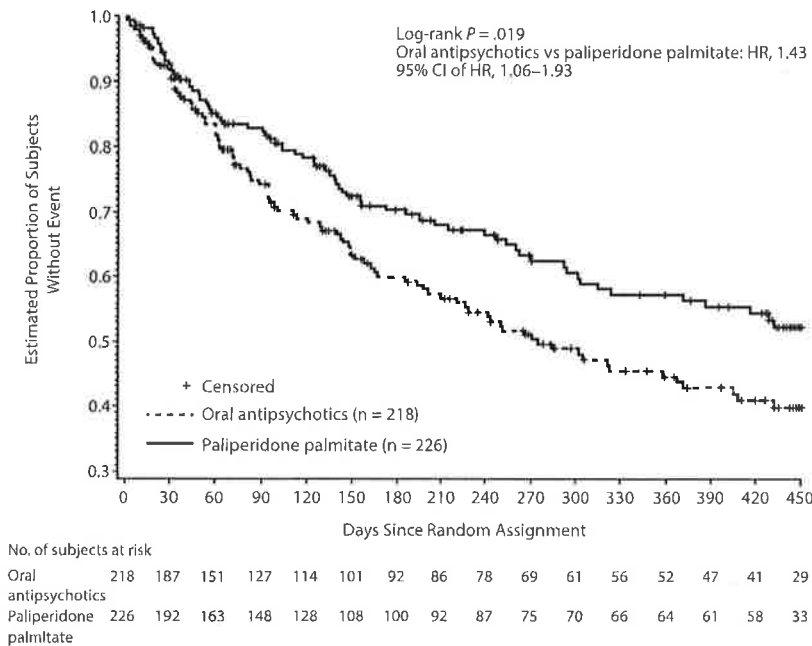
**First psychiatric hospitalization and first arrest/incarceration.** No significant difference in time to first psychiatric hospitalization was seen between the paliperidone palmitate and oral antipsychotic groups (HR, 1.19; 95% CI, 0.67–2.11; nominal *P* = .552). In contrast, paliperidone palmitate was superior to oral antipsychotics in delaying time to first arrest/incarceration (HR, 1.49; 95% CI, 1.08–2.06; nominal *P* = .016).

**Adherence.** When injection records were used to assess adherence, 95.2% in the paliperidone palmitate group had a medication possession ratio (MPR) > 80%. When clinician-based “prescription” records were used to assess oral medication adherence, 77.2% had an MPR > 80%. When pharmacy-based “refill” prescription records were used to assess oral medication adherence, 24.3% in the oral group had an MPR > 80%.

**Exposure to study medication.** Mean (SD) exposure to paliperidone palmitate was 266.2 (174.5) days, and mean exposure to oral antipsychotics was 271.5 (178.5) days. Doses are summarized in eTable 2. Additionally, the mean and median number of injections received by paliperidone palmitate patients was generally comparable to the mean and median number of prescriptions received by patients in the oral antipsychotic group (eTable 2).

**Safety.** Treatment-emergent AEs (TEAEs) were reported in 85.8% and 79.8% of subjects in the paliperidone palmitate and oral antipsychotic treatment arms, respectively. Most common were injection site pain (18.6% of subjects), insomnia (16.8%), weight increase (11.9%), akathisia (11.1%), and anxiety (10.6%) in the paliperidone palmitate group and insomnia (11.5%), headache (8.3%), dry mouth (8.3%), anxiety (7.3%), and sedation (7.3%) in the oral antipsychotic group (Table 2). Incidence of serious TEAEs was 17.3% and 21.6% in the paliperidone palmitate and oral antipsychotic groups, respectively. Treatment-emergent AEs leading to study drug discontinuation occurred in 11.9% of paliperidone palmitate and 7.8% of oral antipsychotic subjects (eTable 3). Rates of extrapyramidal symptom-related TEAEs

**Figure 3. Kaplan-Meier Estimate of Time to First Psychiatric Hospitalization or Arrest/Incarceration<sup>a</sup>**



<sup>a</sup>Data from randomization until end of randomly assigned treatment (28 days after last injection of paliperidone palmitate or 1 day after last dose of oral antipsychotic). Abbreviation: HR = hazard ratio.

for paliperidone palmitate versus oral antipsychotics were akathisia (11.1% vs 6.9%), dyskinesia (2.7% vs 1.4%), dystonia (2.2% vs 2.8%), and Parkinsonism (1.8% vs 1.8%). Incidence of prolactin-related TEAEs was 23.5% and 4.1% in the paliperidone palmitate and oral antipsychotic groups, respectively. Incidences of prolactin-related AEs by gender are presented in eTable 4. In all, 32.4% of subjects in the paliperidone palmitate group and 14.4% in the oral antipsychotic group had a  $\geq 7\%$  increase in weight. One death occurred in the paliperidone palmitate group and was considered by the investigator as unlikely related to the study drug. There were no unexpected safety concerns related to vital signs, physical examination findings, or clinical laboratory test results.

**DISCUSSION**

The PRIDE study demonstrated the superiority of once-monthly paliperidone palmitate over daily oral antipsychotics in delaying time to treatment failure in an innovative randomized study that reflects real-world management of schizophrenia. The study's premise that clinical trials with more pragmatic designs are more likely to demonstrate advantages of LAIs over oral antipsychotics was supported.<sup>19-21</sup> To increase the pragmatic focus of the study, persons at high risk for treatment nonadherence (ie, those with recent involvement with the criminal justice system, comorbid substance abuse, or unstable living conditions) were enrolled. In addition, considerable flexibility in treatment/management decisions by physicians and patients was allowed. Finally, objective and clinically relevant outcome measures were chosen as primary end points. The robust results favoring paliperidone palmitate over oral formulations of commonly used antipsychotics in delaying time to treatment failure suggest that these design differences may be relevant for demonstrating these treatment differences.

In the United States, the criminal justice system has become a frequent setting for management of patients with severe mental illness. It has overtaken psychiatric hospitals as a site for their institutionalization.<sup>27</sup> The results of the

**Table 2. Summary of Treatment-Emergent Adverse Events (TEAEs) in  $\geq 5\%$  of Subjects by Preferred Term<sup>a,b</sup>**

| TEAE, n (%)                       | Paliperidone Palmitate (n = 226) | Oral Antipsychotics (n = 218) |
|-----------------------------------|----------------------------------|-------------------------------|
| Any                               | 194 (85.8)                       | 174 (79.8)                    |
| Injection site pain               | 42 (18.6)                        | 0                             |
| Insomnia                          | 38 (16.8)                        | 25 (11.5)                     |
| Weight increased                  | 27 (11.9)                        | 13 (6.0)                      |
| Akathisia                         | 25 (11.1)                        | 15 (6.9)                      |
| Anxiety                           | 24 (10.6)                        | 16 (7.3)                      |
| Depression                        | 17 (7.5)                         | 14 (6.4)                      |
| Fatigue                           | 17 (7.5)                         | 6 (2.8)                       |
| Erectile dysfunction              | 17 (7.5)                         | 0                             |
| Sedation                          | 15 (6.6)                         | 16 (7.3)                      |
| Dry mouth                         | 15 (6.6)                         | 18 (8.3)                      |
| Increased appetite                | 15 (6.6)                         | 8 (3.7)                       |
| Nasopharyngitis                   | 15 (6.6)                         | 12 (5.5)                      |
| Headache                          | 14 (6.2)                         | 18 (8.3)                      |
| Libido decreased                  | 13 (5.8)                         | 3 (1.4)                       |
| Upper respiratory tract infection | 13 (5.8)                         | 10 (4.6)                      |
| Back pain                         | 13 (5.8)                         | 8 (3.7)                       |
| Schizophrenia                     | 10 (4.4)                         | 15 (6.9)                      |
| Somnolence                        | 10 (4.4)                         | 15 (6.9)                      |
| Toothache                         | 10 (4.4)                         | 12 (5.5)                      |
| Dizziness                         | 5 (2.2)                          | 11 (5.0)                      |
| Suicidal ideation                 | 8 (3.5)                          | 13 (6.0)                      |

<sup>a</sup>Preferred terms of adverse events were based on version 12.0 of the Medical Dictionary for Regulatory Activities (MedDRA) (<http://www.meddra.org>).

<sup>b</sup>This table comprises data from randomization until the end of randomly assigned treatment (28 days after the last injection of paliperidone palmitate or 1 day after the last dose of oral antipsychotic).

PRIDE study suggest that outcomes for this vulnerable population can be improved by medication choice. Indeed, outcomes of particular public health and economic importance (ie, arrest, incarceration, and hospitalization) were the most commonly observed primary end points in this study. We speculate that treatment with paliperidone palmitate leads to more consistent treatment exposure, resulting in fewer symptoms that lead to treatment failure.

Patient symptoms and functioning as measured by the CGI-S and PSP failed to demonstrate differences between paliperidone palmitate and oral treatments in this study. Although these findings may be correct, this failure may also be a consequence of ascertainment bias. That is, assessment with the CGI-S and PSP was not possible when subjects were institutionalized, points at which symptoms and functioning would likely be most deviant from baseline. On the other hand, when subjects were available for assessment, they were most likely to have improved, as evidenced by their ability to keep their clinic visit.

The risk of reinstitutionalization in individuals with mental illness and contact with the criminal justice system is high.<sup>28-30</sup> Data from studies conducted in subjects with significant exposure to the criminal justice system reported rehospitalization rates up to 48% and reincarceration rate up to 68%.<sup>28-30</sup> In the PRIDE study, reinstitutionalization (ie, arrest/incarceration and hospitalization) rates were 33.6% for the paliperidone palmitate arm versus 45% for the oral antipsychotic arm. The lower reinstitutionalization rate observed in the PRIDE study suggests that paliperidone palmitate may reduce the risk of reinstitutionalization in this high-risk population.

In this study, neither patients nor clinicians were blinded to treatment assignment, but all primary end points were independently identified by a blinded event-monitoring board that had no knowledge of treatment assignment. Furthermore, the majority of the end points defined as treatment failures, such as incarceration and psychiatric hospitalization, were highly objective and were chosen for this study because they may indicate deterioration in the subject's clinical state. Although some outcomes such as discontinuation due to efficacy or safety may have been influenced by the knowledge of treatment assignment, they represented only approximately 10% of the treatment failures identified by the blinded event-monitoring board.

Paliperidone palmitate treatment was associated with numerically greater AEs, such as prolactin elevations, sexual side effects, and weight gain. This study was not powered or designed to detect differences from 7 individual oral antipsychotics. Given the varied safety profiles among these agents, the pooled data may have obscured tolerability issues relative to any individual oral agent.

The study was designed to reproduce 2 clinical management practices commonly experienced by patients with chronic schizophrenia: clinical visits where patients receive an LAI antipsychotic and visits where patients receive a prescription for an oral antipsychotic, which they fill in an outside pharmacy. In actual practice, treatment with oral antipsychotic medications requires multiple points of adherence. These include whether a clinic visit is made, whether a prescription is received, whether the prescription is filled, and, finally, whether the medication is taken as prescribed. For injectable medications, these requirements are limited to whether the patient attends a clinic visit and receives an injection. In this study, there was indirect assessment of adherence to medications. Study visits, prescriptions written, and injections were carefully documented. In addition, vouchers provided to pay for subjects' prescriptions when filled were also meant to serve as a record of this fulfillment. However, under the real-world conditions that existed for this study, other mechanisms for payment were potentially available. This led to a possible underestimation of medication possession. For these reasons, more reliable estimates of adherence were available for injectable than oral medications. In summary, the number of study visits was similar for subjects receiving oral medications and paliperidone palmitate, and the overall duration of exposure was comparable

between the 2 groups; however, it was uncertain how many prescriptions were filled for the oral group and whether these patients took their medications as prescribed.

The study design retained some biases that may affect generalization of results. Subjects who were not willing to receive LAI therapy would not have enrolled. This may have contributed to a nonrandom selection process. Other notable differences in gender and race from the US population of persons with schizophrenia are also apparent.

## CONCLUSIONS

The PRIDE study results demonstrated that once-monthly treatment with paliperidone palmitate was more effective in delaying treatment failure versus daily oral antipsychotics (median difference of 190 days) in a trial designed to reflect the real-world management of schizophrenia subjects at risk for treatment failure. These findings support the value of real-world study designs when attempting to identify treatment differences that may relate to formulation differences.

**Drug names:** aripiprazole (Abilify), clozapine (Clozaril, FazaClo, and others), haloperidol (Haldol and others), olanzapine (Zyprexa and others), paliperidone (Invega), paliperidone palmitate (Invega Sustenna), quetiapine (Seroquel and others), risperidone (Risperdal and others).

**Author affiliations:** Janssen Scientific Affairs, LLC (Drs Alphs and Starr, Ms Benson, and Mr Rodriguez); Janssen Research and Development, LLC (Dr Mao and Ms Cheshire-Kinney), Titusville, New Jersey; and Department of Psychiatry, New York University, New York (Dr Lindenmayer).

**Author contributions:** All authors were responsible for analysis and interpretation of the data, drafting of the manuscript, critical revision of the manuscript for important intellectual content, and approval of the final version of the manuscript. In addition, Dr Alphs, Mss Benson and Cheshire-Kinney, and Mr Rodriguez were responsible for study concept and design. Dr Starr was responsible for overall study conduct, while Dr Mao was responsible for study concept and design and acquisition of data.

**Potential conflicts of interest:** Drs Alphs and Starr, Ms Benson, and Mr Rodriguez are employees of Janssen Scientific Affairs, LLC, and Johnson & Johnson stockholders. Dr Mao and Ms Cheshire-Kinney are employees of Janssen Research and Development, LLC, and Johnson & Johnson stockholders. Dr Lindenmayer has received grant/research support from Janssen, Alkermes, Pfizer, Neurocrine, EnVivo, and Roche, and is a consultant for Janssen.

**Funding/support:** This study, including writing and editorial assistance, was funded by Janssen Scientific Affairs, LLC.

**Role of the sponsor:** The study sponsor was responsible for the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, and approval of the manuscript; and the decision to submit the manuscript for publication.

**Previous presentations:** This work has been presented in part at the following conferences: • American Psychiatric Association 167th Annual Meeting; May 3-7, 2014; New York, New York • Society of Biological Psychiatry 69th Annual Scientific Meeting; May 8-10, 2014; New York, New York • ASCP Annual Meeting; June 16-19, 2014; Hollywood, Florida • XVI World Congress of Psychiatry; September 14-18, 2014; Madrid, Spain • APNA 28th Annual Conference; October 22-25, 2014; Indianapolis, Indiana • 27th ECNP Congress of Applied and Translational Neuroscience; October 18-21, 2014; Berlin, Germany • NCHC 2014 National Conference on Correctional Health Care; October 18-22, 2014; Las Vegas, Nevada.

**Acknowledgments:** The authors thank Carol Price, BS, from Janssen Scientific Affairs, LLC (administration of the event-monitoring board); William Olson, PhD, from Janssen Scientific Affairs, LLC (contribution to the study design and statistical methods); Jim O'Neill, BS, from Janssen Scientific Affairs, LLC (contributions to the conduct of the study); Lucy Mahalchick, BA, from Janssen Scientific Affairs, LLC (contribution to the conduct of the study and analysis of the data); Qin Li, MS, from Regeneron Pharmaceuticals, Inc (contributions to the statistical programming and analysis); Cynthia A. Bossie, PhD, from Janssen Scientific Affairs, LLC (contributions to manuscript development); and Joe Hulihan, MD, from Janssen Global Services, LLC (for contributions to study design). The acknowledged individuals have no additional conflicts of interest to report.

The authors also thank the event-monitoring board members: Leslie Citrome, MD, MPH (New York Medical College Valhalla, New York); John Lauriello, MD (through contract with University of Missouri, Columbia); Richard Kavoussi, MD (Neurite Consulting LLC, Raleigh, North Carolina); and Douglas Feltner, MD (Vice President, Pharmaceutical Development, Neuroscience, AbbVie Inc, Ann Arbor, Michigan).

The authors also thank the Paliperidone Palmitate Research in Demonstrating Effectiveness (PRIDE) study investigators: Neal Adams, MD, MPH (Oakland, California); Mohammed Alam, MD (Oak Brook, Illinois); Riaz Baber, MD (Naperville, Illinois); Michael Balkunas, MD (New Britain, Connecticut); Mohammed Bari, MD (National City, California); Carl Bell, MD (Chicago, Illinois); Jeffrey Bennett, MD (Springfield, Illinois); Jason Bermak, MD (San Francisco, California); Prakash Bhatia, MD (Escondido, California); Robert Brauer, MD (Willoughby, Ohio); Ronald Brenner, MD (Cedarhurst, New York); Wen Cai, MD (Tucson, Arizona); Alvaro Camacho, MD (Imperial, California); Jesse M. Carr, MD (Glendale, California); Rakesh Chandra, MD (Paducah, Kentucky); Marc Andrew Colon, MD (Shreveport, Louisiana); Kathleen Degen, MD (New London, Connecticut); Dale D'Mello, MD (East Lansing, Michigan); Steven L. Dubovsky, MD (Buffalo, New York); Jean Fils, MD; Brian Keefe, MD (Tampa, Florida); Karen Frei, MD; Johnny A., Edrozo, MD (San Bernardino, California); Paul Gross, MD (Allentown, Pennsylvania); Robert Horne, MD (Las Vegas, Nevada); Javed Iqbal, MD (Paramus, New Jersey); Valentin Isacescu, MD (Oceanside, California); Shahid Jamil, MD (Novi, Michigan); Gregory Kaczinski, MD (Little Rock, Arkansas); Pravin Kansagra, MD (Anaheim, California); Mary A. Knesewich, MD, PA (Irving, Texas); David G. Krefetz, DO (Willingboro, New Jersey); Joseph Kwentus, MD (Flowood, Mississippi); Mark N. Lerman, MD (Hoffman Estate, Illinois); Jean-Pierre Lindenmayer, MD (New York, New York); Brett Y. Lu, MD (Honolulu, Hawaii); Robert J. Malcolm, MD (Charleston, South Carolina); Raymond A. Manning, MD (Pico Rivera, California); Emilio Mantero-Atienza, MD, PhD (Miami, Florida); Denis Mee-Lee, MD (Honolulu, Hawaii); Manuel Melendez, MD; Mario S. Cuervo, MD (Miami, Florida); Lawrence Edward Mobley, III, MD (Pensacola, Florida); Paul W. Murphy, MD (Wichita, Kansas); Syed Jamal Mustafa, MD (Bothell, Washington); Mark A. Novitsky, MD (Philadelphia, Pennsylvania); Nilesh J. Patel, MD (Wharton, Texas); Pierre Pean, MD (Tamarac, Florida); Marvin Lane Peyton, MD (Oklahoma City, Oklahoma); Mohamed Ramadan, MD (Bullhead City, Arizona); Sriram Ramaswamy, MD; Daniel Wilson, MD (Omaha, Nebraska); Robert A. Riesenber, MD (Atlanta, Georgia); Martha Sajatovic, MD (Cleveland, Ohio); Stephen R. Saklad, PharmD, BCCP (San Antonio, Texas); Adonis Sfera, MD (Anaheim, California); Rajinder Shiwak, MD; Ernest Brownlee, Jr, MD (Arlington, Texas); Roger Sommi, PharmD, BCCP, FCCP (Kansas City, Missouri); Dwight S. St. Clair, DO (Wichita, Kansas); Mary L. Stedman, MD (Tampa, Florida); Rajagopal Keerthy Sunder, MD; Lily S. Chung, MD; Sadashiv Rajadhyaksha, MD; Samuel E. Dey, Jr, MD (Riverside, California); Jose R. Torres, MD (Rio Piedras, Puerto Rico); Mark H. Townsend, MD (New Orleans, Louisiana); John Tran, MD (Spokane, Washington); Tram K. Tran-Johnson, PharmD, PsyD (San Diego, California); Thomas J. Valente, MD (Leesburg, Florida); Ingrid Vasiliu-Felles, MD; Richard Steinbook, MD (Miami, Florida); Stephen J. Volk, MD (Long Beach, California); Robert Weisman, DO (Rochester, New York); Mark J. Woysville, MD (Middleburg Heights, Ohio); and Jose T. Zaglul, MD (Bradenton, Florida).

The authors also thank Maxwell Chang, BSc Hons, and Matthew Grzywacz, PhD (employees of ApotheCom, LLC, Yardley, Pennsylvania), for providing writing and editorial assistance.

**Supplementary material:** Available at PSYCHIATRIST.COM.

## REFERENCES

- Regier DA, Narrow WE, Rae DS, et al. The de facto US mental and addictive disorders service system: epidemiologic catchment area prospective 1-year prevalence rates of disorders and services. *Arch Gen Psychiatry*. 1993;50(2):85-94.
- National Institute of Mental Health. Schizophrenia. <http://www.nimh.nih.gov/health/topics/schizophrenia/index.shtml>. Accessed February 5, 2015.
- Kennedy JL, Altar CA, Taylor DL, et al. The social and economic burden of treatment-resistant schizophrenia: a systematic literature review. *Int Clin Psychopharmacol*. 2014;29(2):63-76.
- Knapp M, Mangalore R, Simon J. The global costs of schizophrenia. *Schizophr Bull*. 2004;30(2):279-293.
- Hoge SK. Providing transition and outpatient services to the mentally ill released from correctional institutions. In: Greifinger RB, ed. *Public Health Behind Bars: From Prisons to Communities*. Dobbs Ferry, NY: Springer; 2007:461-477.
- Ascher-Svanum H, Nyhuis AW, Faries DE, et al. Involvement in the US criminal justice system and cost implications for persons treated for schizophrenia. *BMC Psychiatry*. 2010;10(1):11.
- Folsom D, Jeste DV. Schizophrenia in homeless persons: a systematic review of the literature. *Acta Psychiatr Scand*. 2002;105(6):404-413.
- Robinson D, Woerner MG, Pollack S, et al. Subject selection biases in clinical trials: data from a multicenter schizophrenia treatment study. *J Clin Psychopharmacol*. 1996;16(2):170-176.
- Hofer A, Hummer M, Huber R, et al. Selection bias in clinical trials with antipsychotics. *J Clin Psychopharmacol*. 2000;20(6):699-702.
- Fleischhacker WW, Hummer M. Do phase III trials have clinical value? *J Clin Psychopharmacol*. 1999;19(5):391-392.
- Fleischhacker WW, Kemmler G. The clinical relevance of percentage improvements on the PANSS score. *Neuropsychopharmacology*. 2007;32(11):2435-2436.
- Stroup TS, McEvoy JP, Swartz MS, et al. The National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) project: schizophrenia trial design and protocol development. *Schizophr Bull*. 2003;29(1):15-31.
- Wilk J, Marcus SC, West J, et al. Substance abuse and the management of medication nonadherence in schizophrenia. *J Nerv Ment Dis*. 2006;194(6):454-457.
- Hunt GE, Bergen J, Bashir M. Medication compliance and comorbid substance abuse in schizophrenia: impact on community survival 4 years after a relapse. *Schizophr Res*. 2002;54(3):253-264.
- Gopal S, Hough DW, Xu H, et al. Efficacy and safety of paliperidone palmitate in adult patients with acutely symptomatic schizophrenia: a randomized, double-blind, placebo-controlled, dose-response study. *Int Clin Psychopharmacol*. 2010;25(5):247-256.
- Chue P, Eerdeken M, Augustyns I, et al. Comparative efficacy and safety of long-acting risperidone and risperidone oral tablets. *Eur Neuropsychopharmacol*. 2005;15(1):111-117.
- Bai YM, Chen TT, Wu B, et al. A comparative efficacy and safety study of long-acting risperidone injection and risperidone oral tablets among hospitalized patients: 12-week randomized, single-blind study. *Pharmacopsychiatry*. 2006;39(4):135-141.
- Macfadden W, Ma YW, Thomas Haskins J, et al. A prospective study comparing the long-term effectiveness of injectable risperidone long-acting therapy and oral aripiprazole in patients with schizophrenia. *Psychiatry (Edmont)*. 2010;73(1):23-31.
- Kishimoto T, Kobenzadeh A, Leucht C, et al. Long-acting injectable vs oral antipsychotics for relapse prevention in schizophrenia: a meta-analysis of randomized trials. *Schizophr Bull*. 2014;40(1):192-213.
- Kishimoto T, Nitta M, Borenstein M, et al. Long-acting injectable versus oral antipsychotics in schizophrenia: a systematic review and meta-analysis of mirror-image studies. *J Clin Psychiatry*. 2013;74(10):957-965.
- Kane JM, Kishimoto T, Correll CU. Assessing the comparative effectiveness of long-acting injectable vs oral antipsychotic medications in the prevention of relapse provides a case study in comparative effectiveness research in psychiatry. *J Clin Epidemiol*. 2013;66(suppl):S37-S41.
- Fleischhacker WW, Goodwin GM. Effectiveness as an outcome measure for treatment trials in psychiatry. *World Psychiatry*. 2009;8(1):23-27.
- Alphs L, Mao L, Rodriguez SC, et al. Design and rationale of the Paliperidone Palmitate Research in Demonstrating Effectiveness (PRIDE) study: a novel comparative trial of once-monthly paliperidone palmitate versus daily oral antipsychotic treatment for delaying time to treatment failure in persons with schizophrenia. *J Clin Psychiatry*. 2014;75(12):1388-1393.
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition. Washington, DC: American Psychiatric Association; 1994.
- Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry*. 1998;59(suppl 20):22-33, quiz 34-57.
- Lavori PW, Rush AJ, Wisniewski SR, et al. Strengthening clinical effectiveness trials: equipoise-stratified randomization. *Biol Psychiatry*. 2001;50(10):792-801.
- Kuehn BM. Criminal justice becomes front line for mental health care. *JAMA*. 2014;311(19):1953-1954.
- Kunz M, Yates KF, Czobor P, et al. Course of patients with histories of aggression and crime after discharge from a cognitive-behavioral program. *Psychiatr Serv*. 2004;55(6):654-659.
- Harris V, Koepsell TD. Criminal recidivism in mentally ill offenders: a pilot study. *Bull Am Acad Psychiatry Law*. 1996;24(2):177-186.
- Feder L. A comparison of the community adjustment of the mentally ill offenders with those from the general prison population. *Law Hum Behav*. 1991;15(5):477-493.

See supplementary material for this article at PSYCHIATRIST.COM.



THE JOURNAL OF  
**CLINICAL PSYCHIATRY**  
THE OFFICIAL JOURNAL OF THE AMERICAN SOCIETY OF CLINICAL PSYCHOPHARMACOLOGY

### **Supplementary Material**

**Article Title:** Real-World Outcomes of Paliperidone Palmitate Compared to Daily Oral Antipsychotic Therapy in Schizophrenia: A Randomized, Open-Label, Review Board–Blinded 15-Month Study

**Authors:** Larry Alphas, MD, PhD; Carmela Benson, MS, MSHP; Kimberly Cheshire-Kinney, BA; Jean Pierre Lindenmayer, MD; Lian Mao, PhD; Stephen C. Rodriguez, MS; and H. Lynn Starr, MD

**DOI Number:** 10.4088/JCP.00m09584

#### **List of Supplementary Material for the article**

1. **eTable 1** Concomitant Use of Psychotropic Medications (ITT population)
2. **eTable 2** Dose and Exposure of Paliperidone Palmitate and Oral Antipsychotics
3. **eTable 3** Treatment-Emergent Adverse Events (TEAEs) Leading to Study Drug Discontinuation
4. **eTable 4** Treatment-Emergent Prolactin-Related Adverse Events by Preferred Term for Male and Female Subjects

#### **Disclaimer**

This Supplementary Material has been provided by the author(s) as an enhancement to the published article. It has been approved by peer review; however, it has undergone neither editing nor formatting by in-house editorial staff. The material is presented in the manner supplied by the author.

© Copyright 2015 Physicians Postgraduate Press, Inc.

**Online-only material**

**eTable 1. Concomitant Use of Psychotropic Medications (ITT population)**

| <b>Psychotropic medication, No. (%)</b>     | <b>Paliperidone Palmitate (n = 226)</b> | <b>Oral Antipsychotics (n = 218)</b> |
|---|---|--------------------------------------|
| Any   | 168 (74.3)                              | 175 (80.3)                           |
| Atypical antipsychotics                     | 100 (44.2)                              | 111 (50.9)                           |
| Antidepressants                             | 84 (37.2)                               | 90 (41.3)                            |
| Benzodiazepines                             | 45 (19.9)                               | 49 (22.5)                            |
| Antihistamines                              | 45 (19.9)                               | 55 (25.2)                            |
| Antiextrapyramidal symptoms                 | 52 (23.0)                               | 42 (19.3)                            |
| Mood stabilizers and antiepileptics         | 39 (17.3)                               | 37 (17.0)                            |
| Nonbenzodiazepine hypnotics and anxiolytics | 33 (14.6)                               | 23 (10.6)                            |
| Beta blockers                               | 33 (14.6)                               | 27 (12.4)                            |
| Typical antipsychotics                      | 25 (11.1)                               | 28 (12.8)                            |
| Depot antipsychotics                        | 4 (1.8)                                 | 9 (4.1)                              |
| Stimulants                                  | 4 (1.8)                                 | 0                                    |

Abbreviation: ITT = intent-to-treat.



**eTable 2. Dose and Exposure of Paliperidone Palmitate and Oral Antipsychotics**

|  | Oral Antipsychotics     |                      |                     |                      |                     |                      |                         |                     |
|--|-------------------------|----------------------|---------------------|----------------------|---------------------|----------------------|-------------------------|---------------------|
|  | Paliperidone Palmitate  | Aripiprazole         | Haloperidol         | Olanzapine           | Paliperidone        | Perphenazine         | Quetiapine              | Risperidone         |
| Dose per injection records, mg, mean (SD)    | n = 226<br>181.3 (34.2) | NA                   | NA                  | NA                   | NA                  | NA                   | NA                      | NA                  |
| Dose per prescription records, mg, mean (SD) | NA                      | n = 33<br>15.3 (5.9) | n = 15<br>8.2 (5.3) | n = 36<br>13.3 (6.4) | n = 48<br>6.6 (2.4) | n = 20<br>16.5 (8.8) | n = 29<br>339.9 (180.4) | n = 37<br>3.6 (1.6) |
| Dose per refill records, mg, mean (SD)       | NA                      | n = 25<br>16.6 (6.6) | n = 11<br>7.7 (5.4) | n = 31<br>13.2 (6.5) | n = 43<br>6.4 (2.5) | n = 18<br>14.4 (7.5) | n = 24<br>335.9 (177.4) | n = 30<br>3.4 (1.5) |
| Duration of exposure (days) <sup>a</sup>     | n = 226                 | n = 33               | n = 15              | n = 36               | n = 48              | n = 20               | n = 29                  | n = 37              |
| Mean (SD)                                    | 266.2 (174.5)           | 239.1 (176.7)        | 173.5 (160.8)       | 281.3 (183.6)        | 272.8 (185.3)       | 266.8 (168.8)        | 280.8 (167.8)           | 324.2 (178.1)       |
| Median (range)                               | 251.5 (30-479)          | 242.0 (9-492)        | 137.0 (11-474)      | 325.0 (22-496)       | 315.5 (11-506)      | 270.0 (31-496)       | 304.0 (8-511)           | 427.0 (12-491)      |
| Duration category, n (%)                     |                         |                      |                     |                      |                     |                      |                         |                     |
| ≤30  | 18 (8.0)                | 2 (6.1)              | 4 (26.7)            | 4 (11.1)             | 4 (8.3)             | 0 (0)                | 1 (3.4)                 | 1 (2.7)             |
| 31-90  | 37 (16.4)               | 9 (27.3)             | 2 (13.3)            | 5 (13.9)             | 9 (18.8)            | 5 (25.0)             | 3 (10.3)                | 8 (21.6)            |
| 91-180                                       | 36 (15.9)               | 4 (12.1)             | 2 (13.3)            | 4 (11.1)             | 6 (12.5)            | 1 (5.0)              | 7 (24.1)                | 1 (2.7)             |
| 181-270                                      | 26 (11.5)               | 2 (6.1)              | 2 (13.3)            | 4 (11.1)             | 4 (8.3)             | 5 (25.0)             | 2 (6.9)                 | 3 (8.1)             |
| 271-360                                      | 9 (4.0)                 | 5 (15.2)             | 3 (20.0)            | 1 (2.8)              | 3 (6.3)             | 2 (10.0)             | 4 (13.8)                | 2 (5.4)             |
| 361-450                                      | 41 (18.1)               | 4 (12.1)             | 1 (6.7)             | 5 (13.9)             | 7 (14.6)            | 3 (15.0)             | 4 (13.8)                | 6 (16.2)            |
| >450   | 59 (26.1)               | 7 (21.2)             | 1 (6.7)             | 13 (36.1)            | 15 (31.3)           | 4 (20.0)             | 8 (27.6)                | 16 (43.2)           |

<sup>a</sup>Based on number of injections for the paliperidone palmitate group and number of prescriptions for the oral antipsychotic groups.

**eTable 3. Treatment-Emergent Adverse Events (TEAEs) Leading to Study Drug Discontinuation<sup>a,b</sup>**

| TEAE-Related Study Discontinuation, No. (%)          | Paliperidone Palmitate (n = 226) | Oral Antipsychotics (n = 218) |
|--|----------------------------------|-------------------------------|
| Any  | 27 (11.9)                        | 17 (7.8)                      |
| Body system/Preferred term <sup>c</sup>              |                                  |                               |
| Psychiatric disorders                                | 9 (4.0)                          | 10 (4.6)                      |
| Depressive symptom                                   | 2 (0.9)                          | 0                             |
| Psychotic disorder                                   | 2 (0.9)                          | 0                             |
| Schizophrenia  | 1 (0.4)                          | 5 (2.3)                       |
| Anxiety  | 1 (0.4)                          | 2 (0.9)                       |
| Insomnia   | 1 (0.4)                          | 0                             |
| Agitation  | 1 (0.4)                          | 0                             |
| Abnormal dreams                                      | 1 (0.4)                          | 0                             |
| Libido decreased                                     | 1 (0.4)                          | 0                             |
| Substance abuse                                      | 1 (0.4)                          | 0                             |
| Suicidal ideation                                    | 1 (0.4)                          | 0                             |
| Paranoid schizophrenia                               | 1 (0.4)                          | 0                             |
| Depression   | 0                                | 2 (0.9)                       |
| Auditory hallucination                               | 0                                | 2 (0.9)                       |
| Disorganized schizophrenia                           | 0                                | 1 (0.5)                       |
| Suicide attempt                                      | 0                                | 1 (0.5)                       |
| Nervous system disorders                             | 8 (3.5)                          | 8 (3.7)                       |
| Akathisia  | 2 (0.9)                          | 2 (0.9)                       |
| Headache   | 2 (0.9)                          | 0                             |
| Dystonia   | 1 (0.4)                          | 3 (1.4)                       |
| Dizziness  | 1 (0.4)                          | 1 (0.5)                       |
| Tardive dyskinesia                                   | 1 (0.4)                          | 1 (0.5)                       |
| Dyskinesia   | 1 (0.4)                          | 0                             |
| Oromandibular dystonia                               | 1 (0.4)                          | 1 (0.5)                       |
| Extrapyramidal disorder                              | 1 (0.4)                          | 0                             |
| Parkinsonism   | 0                                | 2 (0.9)                       |
| Reproductive system and breast disorders             | 6 (2.7)                          | 0                             |
| Erectile dysfunction                                 | 2 (0.9)                          | 0                             |
| Amenorrhea   | 1 (0.4)                          | 0                             |
| Ejaculation failure                                  | 1 (0.4)                          | 0                             |
| Galactorrhea   | 1 (0.4)                          | 0                             |
| Gynecomastia   | 1 (0.4)                          | 0                             |
| General disorders and administration site conditions | 4 (1.8)                          | 0                             |
| Injection site pain                                  | 2 (0.9)                          | 0                             |
| Chest discomfort                                     | 1 (0.4)                          | 0                             |
| Sudden death   | 1 (0.4)                          | 0                             |
| Investigations                                       | 3 (1.3)                          | 0                             |
| Weight increased                                     | 2 (0.9)                          | 0                             |
| Semen volume decreased                               | 1 (0.4)                          | 0                             |

<sup>a</sup>This table comprised data from randomization until the end of randomly assigned treatment (28 days after the last injection of paliperidone palmitate or 1 day after the last dose of oral antipsychotic).

<sup>b</sup>Defined as adverse events with an incidence of  $\geq 1$  by body group in either treatment group.

<sup>c</sup>Preferred terms of adverse events were based on version 12.0 of the Medical Dictionary for Regulatory Activities (MedDRA; Medical Dictionary for Regulatory Activities (MedDRA). Version 12.0. <http://www.meddra.org>. Accessed March 13, 2015).

**eTable 4. Treatment-Emergent Prolactin-Related Adverse Events by Preferred Term<sup>a</sup> for Male and Female Subjects<sup>b</sup>**

|                           | <b>Paliperidone Palmitate</b> | <b>Oral Antipsychotics</b> |
|---------------------------|-------------------------------|----------------------------|
| Male subjects, no. (%)    | (n = 193)                     | (n = 190)                  |
| Any                       | 42 (21.8)                     | 7 (3.7)                    |
| Erectile dysfunction      | 17 (8.8)                      | 0                          |
| Decreased libido          | 13 (6.7)                      | 3 (1.6)                    |
| Increased blood prolactin | 7 (3.6)                       | 1 (0.5)                    |
| Breast tenderness         | 2 (1.0)                       | 1 (0.5)                    |
| Gynecomastia              | 3 (1.6)                       | 0                          |
| Breast pain               | 1 (0.5)                       | 0                          |
| Sexual dysfunction        | 1 (0.5)                       | 0                          |
| Hyperprolactinemia        | 4 (2.1)                       | 2 (1.1)                    |
| Anorgasmia                | 3 (1.6)                       | 0                          |
| Loss of libido            | 1 (0.5)                       | 0                          |
| Females, no. (%)          | (n = 33)                      | (n = 28)                   |
| Any                       | 11 (33.3)                     | 2 (7.1)                    |
| Amenorrhea                | 5 (15.2)                      | 1 (3.6)                    |
| Galactorrhea              | 5 (15.2)                      | 0                          |
| Irregular menstruation    | 2 (6.1)                       | 1 (3.6)                    |
| Hyperprolactinemia        | 2 (6.1)                       | 1 (3.6)                    |
| Increased blood prolactin | 2 (6.1)                       | 0                          |

<sup>a</sup>Preferred terms of adverse events were based on version 12.0 of the Medical Dictionary for Regulatory Activities (MedDRA; Medical Dictionary for Regulatory Activities [MedDRA]. Version 12.0. <http://www.meddra.org>. Accessed March 13, 2015).

<sup>b</sup>This table comprised data from randomization until the end of randomly assigned treatment (28 days after the last injection of paliperidone palmitate or 1 day after the last dose of oral antipsychotic).

# Effectiveness and Duration of Protection Provided by the Live-attenuated Herpes Zoster Vaccine in the Medicare Population Ages 65 Years and Older

Hector S. Izurieta,<sup>1,5</sup> Michael Wernecke,<sup>2</sup> Jeffrey Kelman,<sup>3</sup> Sarah Wong,<sup>2</sup> Richard Forshee,<sup>1</sup> Douglas Pratt,<sup>1</sup> Yun Lu,<sup>1</sup> Qin Sun,<sup>2</sup> Christopher Jankosky,<sup>1</sup> Philip Krause,<sup>1</sup> Chris Worrall,<sup>3</sup> Tom MaCurdy,<sup>2</sup> and Rafael Harpaz<sup>4</sup>

<sup>1</sup>Center for Biologics Evaluation and Research, Food and Drug Administration, Silver Spring, Maryland; <sup>2</sup>Acumen LLC, Burlingame, California; <sup>3</sup>Centers for Medicare and Medicaid Services, Washington, DC; <sup>4</sup>National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia; <sup>5</sup>Universidad Rey Juan Carlos, Madrid, Spain

(See the Editorial Commentary by Black on pages 794-5.)

**Background.** Tens of millions of seniors are at risk of herpes zoster (HZ) and its complications. Live attenuated herpes zoster vaccine (HZV) reduces that risk, although questions regarding effectiveness and durability of protection in routine clinical practice remain. We used Medicare data to investigate HZV effectiveness (VE) and its durability.

**Methods.** This retrospective cohort study included beneficiaries ages  $\geq 65$  years during January 2007 through July 2014. Multiple adjustments to account for potential bias were made. HZV-vaccinated beneficiaries were matched to unvaccinated beneficiaries (primary analysis) and to HZV-unvaccinated beneficiaries who had received pneumococcal vaccination (secondary analysis). HZ outcomes in community and hospital settings were analyzed, including ophthalmic zoster (OZ) and postherpetic neuralgia (PHN).

**Results.** Among eligible beneficiaries (average age 77 years), the primary analysis found VE for community HZ of 33% (95% CI: 32%–35%) and 19% (95% CI: 17%–22%), for the first 3, and subsequent 4+ years postvaccination, respectively. In the secondary analysis, VE was, respectively, 37% (95% CI: 36%–39%) and 22% (95% CI: 20%–25%). In the primary analysis, VE for PHN was 57% (95% CI: 52%–61%) and 45% (95% CI: 36%–53%) in the first 3 and subsequent 4+ years, respectively; VE for hospitalized HZ was, respectively, 74% (95% CI: 67%–79%) and 55% (95% CI: 39%–67%). Differences in VE by age group were not significant.

**Conclusions.** In both the primary and secondary analyses, HZV provided protection against HZ across all ages, but effectiveness declined over time. VE was higher and better preserved over time for PHN and HZ-associated hospitalizations than for community HZ.

**Keywords.** Herpes Zoster vaccine; vaccine effectiveness; post-herpetic neuralgia; ophthalmic zoster; elderly.

Herpes zoster (HZ) is a painful condition resulting from the reactivation of the varicella-zoster virus (VZV) from a latent state in sensory nerve ganglia. The disease manifests as a vesicular rash, characteristically unilateral, and restricted to a single dermatome, accompanied by pain along the dermatome. Older populations are particularly affected as the incidence and severity increases with age. There is a marked increase in the incidence of HZ after age 50, which correlates with aging-related decline in cell-mediated immunity [1]. Studies in Canada, Israel, Japan, Taiwan, and the United States reported age-adjusted HZ incidence rates of 8 to 11 per 1000 person-years in populations ages  $\geq 65$  years, whereas incidence in the general population was lower, ranging from 3.4 to 5.0 per 1000 person-years [2, 3]. Complications of HZ include ophthalmic zoster (OZ), bacterial

superinfections of skin lesions and disseminated infections, particularly among immunocompromised patients [4]. The most common serious complication of HZ is postherpetic neuralgia (PHN) [5]. Adults older than 70 years have a 4-fold increased risk of PHN compared with those younger than 60 years [6, 7].

The live-attenuated herpes zoster vaccine, ZOSTAVAX® (Zoster Vaccine Live; Merck & Co., Inc., Whitehouse Station, NJ) was initially licensed in the United States in 2006 to prevent HZ, following a clinical trial in over 38 000 participants ages  $\geq 60$  years with no history of HZ, immunosuppression, or other conditions that could interfere with study participation [8]. In this randomized study, the vaccine (HZV) reduced HZ incidence by 51% (95% CI: 44%, 58%) [1]. Post-marketing studies have generally not been powered to fully explore the roles and interactions of factors such as time since vaccination, age, and disease outcome (e.g., PHN, OZ, and hospitalized HZ) on HZV effectiveness [1, 9–15].

Our objective in this study was to use Medicare databases to robustly evaluate the effectiveness and duration of protection provided by the live-attenuated HZV among beneficiaries ages  $\geq 65$  years by age and risk group.

Received 25 August 2016; editorial decision 18 November 2016; accepted 15 December 2016.  
Correspondence: H. S. Izurieta, MD, MPH, CBER/FDA, 10903 New Hampshire Avenue, Silver Spring, MD 20993 (Hector.izurieta@fda.hhs.gov).

Clinical Infectious Diseases® 2017;64(6):785–93

Published by Oxford University Press for the Infectious Diseases Society of America 2017.  
This work is written by (a) US Government employee(s) and is in the public domain in the US.  
DOI: 10.1093/cid/ciw854

## METHODS

### Study Population

Medicare beneficiaries ages  $\geq 65$  years enrolled in fee-for-service Medicare Parts A (inpatient services), B (outpatient services), and D (prescription drugs) were analyzed retrospectively, comparing those who received a claim for HZV to those who were unvaccinated. Exposure to the HZV was identified using National Drug Codes (NDCs) for ZOSTAVAX, with the vaccination date used as the index date. Vaccinations between January 1, 2007, and July 31, 2014 were eligible for use as index dates. For the unvaccinated cohort, index dates were assigned to match the distribution of vaccination dates, subject to eligibility requirements.

Beneficiaries were included in the analyses if they had 12 months of continuous enrollment in Medicare A/B/D, ensuring that data exist to define baseline characteristics (Figure 1). Additionally, to ensure that unvaccinated beneficiaries were truly unvaccinated, beneficiaries were included in analyses only if they were enrolled in Medicare Part D since at least May 2006, when HZV was first licensed. Beneficiaries were excluded if they had an overlapping nursing home, skilled nursing facility, or hospice stay because claims from these settings are less complete, or if they received an HZ diagnosis or claim for an immunocompromising medical condition in the 12 months prior, or immunosuppressive treatment in the 6 months prior to their index date (Appendix I–II), because these conditions and drugs are known to reduce the effectiveness of the vaccine.

### Cohort Follow-up

Start of follow-up began 30 days after cohort entry [1]. Beneficiaries were followed continuously until: a subsequent HZV claim, death, disenrollment from fee-for-service Medicare, end of study (July 31, 2014), admission to a nursing home, skilled nursing facility, or hospice, occurrence of the outcome of interest, or an immunosuppressive drug claim in the first 2 months after index date.

### Outcomes

Incident HZ cases were defined using *International Classification of Diseases, Ninth Revision* (ICD-9-CM) codes (053.XX for HZ and 053.2X for OZ). Cases were captured in institutional and noninstitutional care settings and categorized into 5 separate outcomes: community HZ, hospitalized HZ, community OZ, hospitalized OZ, and PHN. We do not present our hospitalized OZ results because of low sample size and overall similarity to findings on hospitalized HZ. To identify hospitalized HZ, we used inpatient claims with a primary diagnosis of HZ, which has been shown to have a positive predictive value (PPV) of 86% [15]. For community cases, outcomes in any diagnosis position were used with reported PPVs ranging from 85.2% to 98% [7, 16].

PHN is an important consequence of HZ, especially in older adults. However, detection of PHN may be biased when ascertainment is restricted to medically attended care and is particularly challenging when using administrative data [10, 16]. PHN was defined using a modified version of the algorithm in Klompas et al. [16]. Cases of PHN were identified based on the presence of ICD-9 053.xx in the 90–180 days after an incident HZ diagnosis in combination with at least one of the following events: (1) a new prescription of a PHN treatment (Appendix III) in the 0–60 days from incident HZ diagnosis, (2) the presence of HZ with nervous system complications (ICD-9 053.1x) in the 90–180 days from incident HZ diagnosis, or (3) the presence of a new diagnosis for neuralgia (ICD-9 729.2x) in the 0–180 days from incident HZ diagnosis.

### Covariates

This study accounted for well-studied risk factors of HZ in addition to other potential risk factors believed to be associated with risk of HZ or propensity to seek care once HZ is contracted (Appendix IV) [11, 14, 17, 18]. To achieve balance between cohorts, 1:1 propensity score matching was used, with propensity scores estimated from a logistic regression using all covariates. Beneficiaries were matched using this propensity score and the minimum Mahalanobis distance for key covariates: age, sex, race, and low-income subsidy status [19].

### Falsification Outcomes

HZV recipients might differ from nonrecipients in their ability or desire to access care for HZ or on other unmeasured confounders, introducing ascertainment or selection bias. We adopted the approach in Tseng et al. [11, 17, 20–23] to check for such bias: hazard ratios were calculated for 13 acute symptomatic conditions (Appendix V) in the vaccinated and matched unvaccinated cohorts. Because these conditions were unrelated to HZ, as a group, the hazard ratios were expected to cluster around 1.0; any deviations from 1.0 would alert us to potential biases.

### Statistical Analysis

Incident outcome rates were calculated by dividing the number of cases by the total person-years of observation. Doubly robust Cox regression models were used to estimate the hazard ratios for incident HZ and PHN in the vaccinated compared with the unvaccinated population, in which time intervals were interacted with vaccine status. Doubly robust models adjust for all baseline characteristics (Appendix IV) to account for any residual confounding in post-matching analyses [24, 25]. Additionally, for those beneficiaries in the cohorts who were started on immunosuppressive therapies after the index date, drug use was included in the model as a time-varying covariate (Appendix II). Vaccine effectiveness (VE) was calculated as  $(1 - \text{HR}) \times 100\%$ , where HR

is the estimated hazard ratio between the 2 cohorts for a particular time interval, unless noted otherwise.

The secondary analysis used analogous methods but compared HZ vaccinees to beneficiaries who received a

pneumococcal vaccine (PV; Pneumovax<sup>®</sup>23, Merck & Co Inc., Whitehouse Station, New Jersey) but did not receive HZV. PV beneficiaries were followed from vaccination as their index date. Additional analyses were performed on both primary and

**Table 1. Demographic and Health-related Characteristics at Baseline in Propensity Score and Mahalanobis Matched Medicare Beneficiaries in the Primary Analysis Comparing Vaccinated and Unvaccinated Cohorts From January 2007 to July 2014**

| Demographic Variables         | Vaccinated |     | Unvaccinated <sup>a</sup> |     | Austin Std. Diff <sup>b</sup> |
|-------------------------------|------------|-----|---------------------------|-----|-------------------------------|
| Base population               | 945992     |     | 945992                    |     |                               |
| Age (continuous) <sup>c</sup> |            |     |                           |     |                               |
| Mean                          | 77         |     | 77                        |     | 0.00                          |
| SD                            | 6.15       |     | 6.15                      |     |                               |
| Age (categories) <sup>c</sup> |            |     |                           |     |                               |
| 65–69                         | 106167     | 11% | 106077                    | 11% | 0.00                          |
| 70–74                         | 292343     | 31% | 292606                    | 31% | 0.00                          |
| 75–79                         | 262853     | 28% | 262935                    | 28% | 0.00                          |
| 80–84                         | 172306     | 18% | 172203                    | 18% | 0.00                          |
| 85–89                         | 84236      | 9%  | 84174                     | 9%  | 0.00                          |
| 90+                           | 28087      | 3%  | 27997                     | 3%  | 0.00                          |
| Sex                           |            |     |                           |     |                               |
| Male                          | 316743     | 33% | 316545                    | 33% | 0.00                          |
| Female                        | 629249     | 67% | 629447                    | 67% | 0.00                          |
| Race                          |            |     |                           |     |                               |
| White                         | 850032     | 90% | 850118                    | 90% | 0.00                          |
| Black                         | 22086      | 2%  | 22068                     | 2%  | 0.00                          |
| Asian                         | 44621      | 5%  | 44581                     | 5%  | 0.00                          |
| Hispanic                      | 10905      | 1%  | 10895                     | 1%  | 0.00                          |
| Other                         | 18348      | 2%  | 18330                     | 2%  | 0.00                          |
| Metropolitan statistical area |            |     |                           |     |                               |
| Urban                         | 672857     | 71% | 672710                    | 71% | 0.00                          |
| Rural                         | 271872     | 29% | 272045                    | 29% | 0.00                          |
| Low-income subsidy status     |            |     |                           |     |                               |
| No LIS                        | 740582     | 78% | 740643                    | 78% | 0.00                          |
| Hospital visits               |            |     |                           |     |                               |
| 1                             | 96593      | 10% | 97966                     | 10% | 0.00                          |
| 2+                            | 34458      | 4%  | 34559                     | 4%  | 0.00                          |
| ER visits                     |            |     |                           |     |                               |
| 0                             | 765417     | 81% | 764715                    | 81% | 0.00                          |
| 1                             | 131875     | 14% | 132528                    | 14% | 0.00                          |
| 2+                            | 48700      | 5%  | 48749                     | 5%  | 0.00                          |
| Physician office visits       |            |     |                           |     |                               |
| 0–4                           | 153886     | 16% | 152646                    | 16% | 0.00                          |
| 5–10                          | 285905     | 30% | 287462                    | 30% | 0.00                          |
| 11–20                         | 291555     | 31% | 291723                    | 31% | 0.00                          |
| 21–30                         | 120457     | 13% | 120359                    | 13% | 0.00                          |
| 31+                           | 94189      | 10% | 93802                     | 10% | 0.00                          |
| Medical conditions            |            |     |                           |     |                               |
| Diabetes                      | 274517     | 29% | 280717                    | 30% | 0.01                          |
| Kidney disease                | 86690      | 9%  | 88333                     | 9%  | 0.01                          |
| Heart disease                 | 263650     | 28% | 266710                    | 28% | 0.01                          |
| Lung disease                  | 195309     | 21% | 197762                    | 21% | 0.01                          |
| Liver disease                 | 14889      | 2%  | 14828                     | 2%  | 0.00                          |

Abbreviation: ER, emergency room.

<sup>a</sup> Data were measured 1 year prior to index date. For the vaccinated, the index date was the date of first vaccination. For the unvaccinated, the index date was assigned to match the distribution of vaccination dates. <sup>b</sup>The balance of covariates between the matched cohorts was assessed using the standardized mean difference [20]. A standardized mean difference of 0.1 or less indicated a negligible difference in means or proportions between cohorts. <sup>c</sup>The age distribution corresponds to the eligibility restriction that required beneficiaries be enrolled in Medicare since vaccine licensure. Frailty characteristics consisted of dementia, home oxygen use, urinary catheter use, walker use, and wheelchair use. These characteristics were identified in 1%–4% of the base population, where each was balanced between vaccinated and unvaccinated cohorts.

The full list of covariates that were included in the propensity score model are provided in Appendix IV.

**Table 2. Community Herpes Zoster (HZ) Incidence by Vaccination Status in the Primary Matched Population**

| Characteristic     | No. Community HZ Cases |              | Total Person-years |              | Rate/1000 person-years (95% CI) |                   | % Community HZ Cases that led to PHN <sup>a</sup> |              |
|--------------------|------------------------|--------------|--------------------|--------------|---------------------------------|-------------------|---|--------------|
|                    | Vaccinated             | Unvaccinated | Vaccinated         | Unvaccinated | Vaccinated                      | Unvaccinated      | Vaccinated  | Unvaccinated |
| Overall            | 27 556                 | 29 383       | 2 576 815          | 1 952 710    | 10.7 (10.6, 10.8)               | 15.0 (14.9, 15.2) | 2.9%  | 4.0%         |
| Year of follow-up  |                        |              |                    |              |                                 |                   |   |              |
| 1                  | 7 539                  | 11 166       | 8 153 355          | 7 444 417    | 9.2 (9.0, 9.5)                  | 15.0 (14.7, 15.3) | 2.5%  | 4.1%         |
| 2                  | 6 397                  | 7 556        | 6 113 365          | 4 960 000    | 10.5 (10.2, 10.7)               | 15.2 (14.9, 15.6) | 2.8%  | 3.6%         |
| 3                  | 4 655                  | 4 417        | 4 207 333          | 2 991 588    | 11.1 (10.8, 11.4)               | 14.8 (14.3, 15.2) | 3.4%  | 4.1%         |
| 4                  | 3 666                  | 2 935        | 3 114 440          | 1 970 248    | 11.8 (11.4, 12.2)               | 14.9 (14.4, 15.4) | 3.1%  | 4.4%         |
| 5                  | 2 880                  | 1 976        | 2 294 288          | 1 305 598    | 12.6 (12.1, 13.0)               | 15.1 (14.5, 15.8) | 3.3%  | 4.6%         |
| 6                  | 1 716                  | 990          | 1 324 459          | 637 737      | 13.0 (12.4, 13.6)               | 15.5 (14.6, 16.5) | 2.7%  | 5.3%         |
| 7+                 | 703                    | 343          | 560 366            | 217 776      | 12.5 (11.7, 13.5)               | 15.8 (14.2, 17.5) | 1.0%  | 1.5%         |
| Age                |                        |              |                    |              |                                 |                   |   |              |
| 65–69              | 3 940                  | 3 944        | 4 053 346          | 2 805 048    | 9.7 (9.4, 10.0)                 | 14.1 (13.6, 14.5) | 2.5%  | 3.8%         |
| 70–74              | 8 869                  | 9 782        | 8 711 733          | 6 675 388    | 10.2 (10.0, 10.4)               | 14.7 (14.4, 14.9) | 2.6%  | 3.6%         |
| 75–79              | 7 465                  | 7 860        | 6 759 404          | 5 176 660    | 11.0 (10.8, 11.3)               | 15.2 (14.9, 15.5) | 2.7%  | 4.2%         |
| 80–84              | 4 696                  | 4 959        | 4 068 809          | 3 170 050    | 11.5 (11.2, 11.9)               | 15.6 (15.2, 16.1) | 3.6%  | 5.0%         |
| 85–89              | 2 019                  | 2 179        | 1 706 667          | 1 337 734    | 11.8 (11.3, 12.4)               | 16.3 (15.6, 17.0) | 3.5%  | 3.9%         |
| 90+                | 567                    | 659          | 468 880            | 362 224      | 12.1 (11.1, 13.1)               | 18.2 (16.9, 19.6) | 2.8%  | 3.8%         |
| Sex                |                        |              |                    |              |                                 |                   |   |              |
| Male               | 7 892                  | 8 493        | 8 646 077          | 6 498 840    | 9.1 (8.9, 9.3)                  | 13.1 (12.8, 13.4) | 3.0%  | 4.1%         |
| Female             | 19 664                 | 20 890       | 17 122 209         | 13 028 771   | 11.5 (11.3, 11.6)               | 16.0 (15.8, 16.3) | 2.8%  | 4.0%         |
| Race               |                        |              |                    |              |                                 |                   |   |              |
| White              | 24 559                 | 26 914       | 2 339 618          | 1 767 930    | 10.5 (10.4, 10.6)               | 15.2 (15.0, 15.4) | 2.8%  | 4.0%         |
| Black              | 266                    | 355          | 399 488            | 373 528      | 6.7 (5.9, 7.5)                  | 9.5 (8.6, 10.5)   | 1.9%  | 4.5%         |
| Other <sup>b</sup> | 2 731                  | 2 114        | 197 249            | 147 428      | 13.8 (13.3, 14.4)               | 14.3 (13.7, 15.0) | 3.2%  | 4.1%         |

Abbreviations: CI, confidence interval; HZ, herpes zoster; PHN, postherpetic neuralgia.

<sup>a</sup>The proportion of beneficiaries that were coded for postherpetic neuralgia in the 90–180 days after an incident community herpes zoster event among beneficiaries that received a diagnosis for community herpes zoster.

<sup>b</sup>Asian, Hispanic, Native American, Other, and Unknown Race are included in the Race—Other category.

secondary populations to investigate whether there was effect modification due to age, sex, and race.

This study was performed as part of the SafeRx Project, a joint initiative of the Centers for Medicare and Medicaid Services (CMS) and the US Food and Drug Administration (FDA). It was approved by the Research in Human Subjects Committee of FDA's Center for Biologics Evaluation and Research. Analyses were performed using R 3.0.2 (R Foundation for Statistical Computing, Vienna, Austria) and SAS 9.4 (SAS Institute Inc., Cary, North Carolina).

## RESULTS

### Primary Analysis

Prior to matching, vaccinated and unvaccinated beneficiaries differed considerably. After matching, a total of 945 992 vaccinated and 945 992 unvaccinated beneficiaries were obtained, resulting in well-balanced cohorts (Table 1). The population was 67% female, 90% white, and 77 years old on average at index date. Beneficiaries were followed for up to 7.5 years with the average length of follow-up being 2.5 years. Approximately 4% of beneficiaries were censored due to death and 17% due to admittance to a nursing home, skilled nursing facility, or

hospice. During follow-up, 0.7% of eligible beneficiaries began immunosuppressive treatment.

A total of 56 964 incident zoster outcomes were detected in our matched study population. Among the 29 401 cases in the unvaccinated cohort, 29 383 (>99%) received community care, 429 (1%) were hospitalized, 2,699 (9%) experienced OZ, and 1,229 (4%) experienced PHN. In the unvaccinated, the incidence of HZ was 15.0 per 1000 person-years for community events and 0.21 per 1000 person-years for hospitalized events. About 4% of community and 9% of hospitalized cases were also coded as PHN (Table 2, eTable 1). HZ rates remained steady by calendar year in the unvaccinated cohort (eFigure 2).

Before conducting the main analysis, we assessed whether there were underlying differences between study cohorts. We compared vaccinated to unvaccinated beneficiaries using 13 falsification outcomes, which were expected to have no association with HZ. The adjusted hazard ratio (AHR) of the 13 falsification outcomes in the vaccinated and unvaccinated cohorts ranged from 0.72 (95%CI: 0.70–0.73) for hip fracture outcomes to 1.17 (95%CI: 1.15–1.20) for lipoma (Figure 2). After running the main analysis, none of the 13 outcomes had an AHR as far from the null (1.0) as that of the zoster outcomes.



**Table 3a. Primary Analysis: Adjusted Vaccine Effectiveness and 95% CI Comparing Herpes Zoster Vaccinated to Unvaccinated for Community Herpes Zoster, Hospitalized Herpes Zoster, Community Ophthalmic Zoster and Postherpetic Neuralgia Outcomes by Years of Follow-up<sup>a</sup>**

| Year of Follow-up    | Outpatient Herpes Zoster<br>(No. Outcomes = 56 939) | Hospitalized Herpes Zoster <sup>b</sup> (No.<br>Outcomes = 614) | Outpatient Ophthalmic Zoster (No.<br>Outcomes = 5,282) | Postherpetic Neuralgia (No.<br>Outcomes = 2,033) |
|----------------------|---|---|--|--|
|                      | VE (95% CI)   | VE (95% CI)   | VE (95% CI)  | VE (95% CI)                                      |
| <b>Primary model</b> |   |   |  |  |
| First 3 years        | 33%** (32%, 35%)                                    | 74%** (67%, 79%)  | 31%** (27%, 36%)                                       | 57%** (52%, 61%)                                 |
| 4 or more years      | 19%** (17%, 22%)                                    | 55%** (39%, 67%)  | 21%** (12%, 29%)                                       | 45%** (36%, 53%)                                 |
| <b>Yearly model</b>  |   |   |  |  |
| 1                    | 38%** (37%, 40%)                                    | 77%** (68%, 84%)  | 38%** (32%, 44%)                                       | 70%** (63%, 76%)                                 |
| 2                    | 32%** (29%, 34%)                                    | 68%** (55%, 78%)  | 28%** (20%, 36%)                                       | 49%** (39%, 57%)                                 |
| 3                    | 25%** (22%, 28%)                                    | 75%** (60%, 84%)  | 22%** (11%, 32%)                                       | 50%** (39%, 59%)                                 |
| 4                    | 21%** (17%, 25%)                                    | 45%* (11%, 66%)   | 19%** (6%, 31%)  | 44%** (28%, 56%)                                 |
| 5                    | 17%** (12%, 22%)                                    | 52%* (15%, 73%)   | 21%* (4%, 34%)   | 40%** (21%, 54%)                                 |
| 6                    | 17%** (10%, 23%)                                    | 66%** (30%, 83%)  | 19% (-3%, 37%)   | 52%** (31%, 66%)                                 |
| 7+                   | 21%** (11%, 31%)                                    | 70%* (25%, 88%)   | 33% (-1%, 55%)   | 60%** (28%, 78%)                                 |

Abbreviations: CI, confidence interval; HR, hazard ratio; HZ, herpes zoster; VE, vaccine effectiveness.

<sup>a</sup>Significance: \*\*  $P < .01$ , \*  $P < .05$ .

<sup>b</sup>Approximately 90% of hospitalized HZ cases were identified in hospital stays lasting  $\geq 2$  nights. VE Formula for All:  $1 - [(\text{HR in Cohort}) * (\text{HR interaction of Cohort} \times \text{Years of follow-up}) * 100\%]$  VE Formula for Subgroups:  $1 - [(\text{HR in Cohort}) * (\text{HR interaction of Cohort} \times \text{Years of follow-up}) * (\text{HR interaction of Cohort} \times \text{Age/Sex/Race})] * 100\%$  Demographic factors, socioeconomic conditions, healthcare utilization characteristics, frailty characteristics, functional immunocompromising chronic conditions, and time-varying immunocompromising drugs were adjusted in the model.

**Table 3b. Primary Subgroup Analysis: Adjusted Vaccine Effectiveness and 95% CI Comparing Herpes Zoster Vaccinated to Unvaccinated for Community Herpes Zoster, Hospitalized Herpes Zoster, Community Ophthalmic Zoster, and Postherpetic Neuralgia Outcomes by Years of Follow-up<sup>a</sup>**

| Subgroups <sup>c</sup> | Community Herpes Zoster<br>(No. Outcomes = 56 939) |                  | Hospitalized Herpes Zoster <sup>b</sup><br>(No. Outcomes = 614) |                  | Community Ophthalmic Zoster<br>(No. Outcomes = 5 282) |                  | Postherpetic Neuralgia<br>(No. Outcomes = 2 033) |                  |
|------------------------|--|------------------|---|------------------|---|------------------|--|------------------|
|                        | First 3 Years                                      | 4 or More Years  | First 3 Years   | 4 or More Years  | First 3 Years   | 4 or More Years  | First 3 Years                                    | 4 or More Years  |
|                        | VE (95% CI)  | VE (95% CI)      | VE (95% CI)   | VE (95% CI)      | VE (95% CI)   | VE (95% CI)      | VE (95% CI)                                      | VE (95% CI)      |
| <b>Age</b>             |  |                  |   |                  |   |                  |  |                  |
| 65–69                  | 36%** (33%, 39%)                                   | 22%** (18%, 26%) | 68%** (41%, 83%)  | 47%* (1%, 72%)   | 35%** (24%, 44%)                                      | 24%** (10%, 36%) | 61%** (49%, 70%)                                 | 50%** (34%, 62%) |
| 70–74                  | 35%** (33%, 37%)                                   | 21%** (17%, 24%) | 70%** (57%, 79%)  | 51%** (26%, 67%) | 29%** (22%, 36%)                                      | 18%** (7%, 28%)  | 55%** (46%, 62%)                                 | 42%** (29%, 53%) |
| 75–79                  | 32%** (30%, 34%)                                   | 17%** (13%, 21%) | 73%** (61%, 81%)  | 56%** (32%, 71%) | 33%** (25%, 40%)                                      | 22%** (10%, 32%) | 61%** (54%, 68%)                                 | 50%** (38%, 60%) |
| 80–84                  | 31%** (28%, 34%)                                   | 16%** (11%, 20%) | 74%** (61%, 82%)  | 57%** (31%, 73%) | 32%** (22%, 40%)                                      | 21%** (7%, 32%)  | 55%** (45%, 63%)                                 | 42%** (26%, 54%) |
| 85–89                  | 32%** (27%, 36%)                                   | 17%** (11%, 22%) | 75%** (60%, 85%)  | 59%** (29%, 77%) | 30%** (16%, 42%)                                      | 19% (0%, 35%)    | 47%** (27%, 61%)                                 | 32%* (4%, 52%)   |
| 90+                    | 37%** (29%, 43%)                                   | 23%** (13%, 31%) | 88%** (67%, 95%)  | 79%** (44%, 93%) | 32%* (3%, 52%)  | 21% (-14%, 45%)  | 58%** (22%, 78%)                                 | 46% (-2%, 72%)   |
| <b>Sex</b>             |  |                  |   |                  |   |                  |  |                  |
| Male                   | 35%** (32%, 37%)                                   | 21%** (17%, 24%) | 74%** (63%, 81%)  | 55%** (32%, 70%) | 40%** (33%, 46%)                                      | 30%** (21%, 39%) | 56%** (47%, 63%)                                 | 44%** (31%, 54%) |
| Female                 | 33%** (31%, 34%)                                   | 19%** (16%, 21%) | 74%** (66%, 79%)  | 55%** (37%, 68%) | 27%** (22%, 33%)                                      | 16%** (6%, 25%)  | 57%** (52%, 62%)                                 | 46%** (36%, 54%) |
| <b>Race</b>            |  |                  |   |                  |   |                  |  |                  |
| White                  | 35%** (34%, 36%)                                   | 21%** (18%, 24%) | 74%** (67%, 79%)  | 55%** (39%, 68%) | 31%** (27%, 36%)                                      | 21%** (12%, 29%) | 58%** (53%, 63%)                                 | 47%** (38%, 55%) |
| Black                  | 32%** (21%, 42%)                                   | 18%* (4%, 30%)   | 90%* (21%, 99%)   | 83% (-36%, 98%)  | 44%* (2%, 68%)  | 35% (-14%, 63%)  | 73%* (25%, 90%)                                  | 65%* (5%, 87%)   |
| Asian                  | 12%** (6%, 18%)                                    | -6% (-15%, 1%)   | 70%** (51%, 82%)  | 49%* (13%, 71%)  | 29%** (12%, 43%)                                      | 18% (-3%, 35%)   | 39%** (17%, 55%)                                 | 22% (-7%, 44%)   |
| Other                  | 18%** (9%, 25%)                                    | 0% (-11%, 10%)   |   |                  |   |                  |  |                  |

Abbreviations: CI, confidence interval; HR, hazard ratio; HZ, herpes zoster; VE, vaccine effectiveness.

<sup>a</sup>Significance: \*\*  $P < .01$ , \*  $P < .05$ .

<sup>b</sup>Approximately 90% of hospitalized HZ cases were identified in hospital stays lasting  $\geq 2$  nights.

<sup>c</sup>VE Formula for All:  $1 - [(\text{HR in Cohort}) * (\text{HR interaction of Cohort} \times \text{Years of follow-up}) * 100\%]$ . VE Formula for Subgroups:  $1 - [(\text{HR in Cohort}) * (\text{HR interaction of Cohort} \times \text{Years of follow-up}) * (\text{HR interaction of Cohort} \times \text{Age/Gender/Race})] * 100\%$  Demographic factors, socio-economic conditions, healthcare utilization characteristics, frailty characteristics, functional immunocompromising chronic conditions, and time-varying immunocompromising drugs were adjusted in the model.

**Table 3c. Secondary Analysis: Adjusted Vaccine Effectiveness and 95% CI Comparing Herpes Zoster Vaccinated to Pneumococcal Vaccinated for Community Herpes Zoster, Hospitalized Herpes Zoster, Community Ophthalmic Zoster and Postherpetic Neuralgia Outcomes by Years of Follow-up<sup>a</sup>**

| Subgroups <sup>c</sup> | Community Herpes Zoster<br>(No. Outcomes = 45477) |                  | Hospitalized Herpes Zoster <sup>b</sup><br>(No. Outcomes = 452) |                  | Community Ophthalmic Zoster<br>(No. Outcomes = 4311) |                  | Postherpetic Neuralgia<br>(No. Outcomes = 1519) |                  |
|------------------------|---|------------------|---|------------------|--|------------------|---|------------------|
|                        | First 3 Years                                     | 4 or More Years  | First 3 Years   | 4 or More Years  | First 3 Years  | 4 or More Years  | First 3 Years                                   | 4 or More Years  |
|                        | VE (95% CI)                                       | VE (95% CI)      | VE (95% CI)   | VE (95% CI)      | VE (95% CI)  | VE (95% CI)      | VE (95% CI)                                     | VE (95% CI)      |
| All                    | 37%** (36%, 39%)                                  | 22%** (20%, 25%) | 67%** (58%, 74%)  | 34%* (8%, 53%)   | 39%** (35%, 43%)                                     | 19%** (10%, 28%) | 63%** (57%, 67%)                                | 47%** (36%, 55%) |
| Age                    |   |                  |   |                  |  |                  |   |                  |
| 65–69                  | 40%** (37%, 43%)                                  | 25%** (21%, 29%) | 69%** (38%, 85%)  | 41% (-22%, 71%)  | 43%** (32%, 52%)                                     | 24%** (9%, 36%)  | 63%** (49%, 74%)                                | 47%** (26%, 62%) |
| 70–74                  | 38%** (36%, 40%)                                  | 23%** (19%, 26%) | 51%** (26%, 68%)  | 5% (-52%, 40%)   | 39%** (31%, 46%)                                     | 18%** (6%, 29%)  | 67%** (59%, 73%)                                | 52%** (40%, 62%) |
| 75–79                  | 37%** (35%, 39%)                                  | 21%** (17%, 25%) | 67%** (49%, 78%)  | 35% (-5%, 60%)   | 40%** (33%, 47%)                                     | 20%** (7%, 31%)  | 62%** (53%, 70%)                                | 46%** (30%, 58%) |
| 80–84                  | 36%** (33%, 39%)                                  | 20%** (16%, 24%) | 76%** (63%, 85%)  | 54%** (24%, 72%) | 39%** (29%, 47%)                                     | 18%* (3%, 31%)   | 57%** (46%, 66%)                                | 38%** (19%, 53%) |
| 85–89                  | 36%** (32%, 41%)                                  | 20%** (14%, 26%) | 66%** (43%, 80%)  | 34% (-20%, 63%)  | 35%** (20%, 48%)                                     | 14% (-9%, 32%)   | 61%** (46%, 73%)                                | 44%* (19%, 62%)  |
| 90+                    | 32%** (23%, 40%)                                  | 15%* (3%, 26%)   | 81%** (42%, 94%)  | 62% (-19%, 88%)  | 29% (-7%, 53%)                                       | 5% (-45%, 37%)   | 60%* (13%, 81%)                                 | 42% (-28%, 73%)  |
| Gender                 |   |                  |   |                  |  |                  |   |                  |
| Male                   | 39%** (36%, 41%)                                  | 24%** (20%, 27%) | 63%** (46%, 75%)  | 26% (-16%, 53%)  | 48%** (42%, 54%)                                     | 32%** (21%, 41%) | 63%** (55%, 70%)                                | 48%** (34%, 59%) |
| Female                 | 37%** (35%, 38%)                                  | 22%** (19%, 24%) | 69%** (59%, 76%)  | 37%* (10%, 56%)  | 34%** (29%, 40%)                                     | 13%* (3%, 23%)   | 62%** (56%, 68%)                                | 46%** (35%, 55%) |
| Race                   |   |                  |   |                  |  |                  |   |                  |
| White                  | 39%** (37%, 40%)                                  | 24%** (21%, 26%) | 69%** (60%, 76%)  | 38%** (12%, 56%) | 39%** (34%, 43%)                                     | 19%** (10%, 28%) | 63%** (58%, 68%)                                | 47%** (37%, 56%) |
| Black                  | 36%** (22%, 47%)                                  | 20%* (3%, 34%)   | 86%** (-19%, 98%)   | 71% (-142%, 97%) | 26% (-52%, 64%)                                      | 2% (-103%, 59%)  | 90%** (59%, 98%)                                | 86%** (41%, 97%) |
| Asian                  | 23%** (17%, 29%)                                  | 4% (-4%, 12%)    | 48%** (7%, 71%)   | -3% (-92%, 45%)  | 44%** (29%, 56%)                                     | 26%* (4%, 43%)   | 49%** (25%, 66%)                                | 28% (-8%, 52%)   |
| Other                  | 24%** (15%, 33%)                                  | 6% (-6%, 17%)    |   |                  |  |                  |   |                  |

Abbreviations: CI, confidence interval; HR, hazard ratio; HZ, herpes zoster; VE, vaccine effectiveness.

<sup>a</sup>Significance: \*\*  $P < .01$ , \*  $P < .05$ .

<sup>b</sup>Approximately 90% of hospitalized HZ cases were identified in hospital stays lasting  $\geq 2$  nights.

<sup>c</sup>VE Formula for All:  $1 - [(\text{HR in Cohort}) * (\text{HR interaction of Cohort} \times \text{Years of follow-up}) * 100\%]$ . VE Formula for Subgroups:  $1 - [(\text{HR in Cohort}) * (\text{HR interaction of Cohort} \times \text{Years of follow-up}) * (\text{HR interaction of Cohort} \times \text{Age/Gender/Race}) * 100\%]$ . Demographic factors, socio-economic conditions, healthcare utilization characteristics, frailty characteristics, functional immunocompromising chronic conditions, and time-varying immunocompromising drugs were adjusted in the model.

In the fully adjusted analysis, vaccination was associated with a reduced risk of HZ (Table 3a). Compared to unvaccinated beneficiaries, in the first 3 years of follow-up, the adjusted vaccine effectiveness (AVE) was 33% (95% CI 32%–35%) for community HZ; AVE declined slightly by age ( $P = .029$ ), where it was highest in ages 65–69 (36% [95% CI 33%–39%]) and lowest in ages 80–84 (31% [95% CI 28%–34%]). For hospitalized HZ and PHN, the AVE for the first 3 years of follow-up was higher: 74% (95% CI: 67%–79%) and 57% (95% CI: 52%–61%), respectively. For all outcomes, there was no evidence of effect modification by gender. Key predictors of HZ by setting are displayed in eTable 2.

For community HZ, AVE declined significantly over time (Table 3a). VE was not only higher but also better retained over time for PHN and hospitalized HZ than for community HZ. Trends on the duration of protection for the vaccine are also summarized in Figure 2. The unadjusted hazard rate for unvaccinated beneficiaries was flat in outcomes over time, whereas the

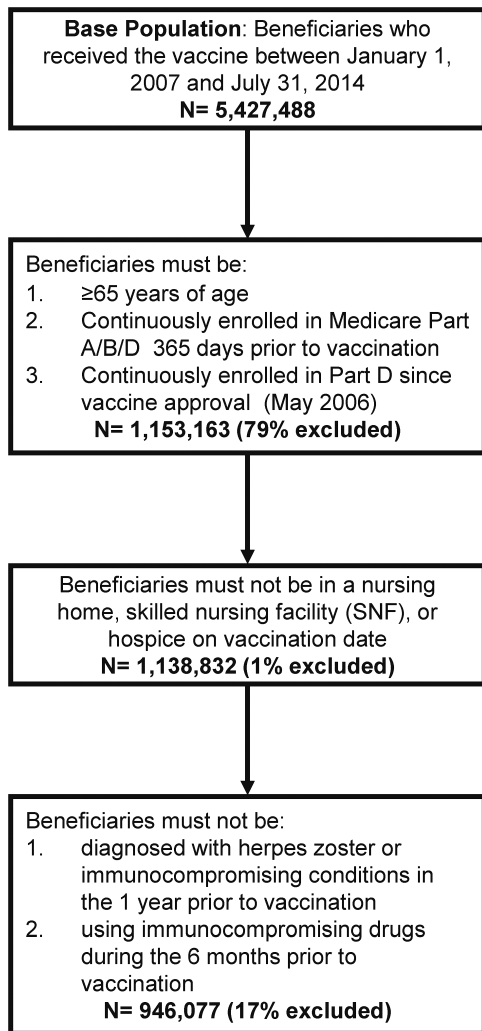
vaccinated started off with a low HZ hazard that increased over time.

### Secondary Analysis

In the secondary analyses, we matched 608 982 HZV recipients to 608 982 recipients of pneumococcal vaccine (Table 3c; Supplemental Material). The AVE in the first 3 years of follow-up for community HZ (37%, 95% CI 36%–39%) was 4% higher than for the primary analysis (Table 3c). Results for all other outcomes, including HZ hospitalizations and PHN, were similar to those in the primary analysis.

### DISCUSSION

We used matched cohort data from approximately 2 million Medicare beneficiaries to investigate HZ vaccine effectiveness and duration of protection, using multiple analytical approaches to identify and address potential bias. This study assesses the



**Figure 1.** Vaccinated cohort flowchart.

effectiveness and durability of the live-attenuated HZV with sufficient power to examine less common outcomes such as PHN and HZ-associated hospitalizations, while controlling for a wide range of HZ risk factors and potential confounders [1, 9, 12, 26, 27]. Furthermore, our Medicare fee-for-service data are sourced from the general US population ages 65 years and older, capturing the entire range of routine clinical practice.

Among eligible study participants averaging 77 years of age, the primary and secondary analyses, respectively, found that HZV was 33% and 37% effective for the prevention of community HZ during the first 3 years postvaccination. For PHN and HZ-associated hospitalization, we found higher HZV effectiveness, 57% and 74%, respectively, for the first 3 years postvaccination in the primary analysis. The differences in these estimates, and their consistency in both the primary and secondary analyses, suggest real differences in VE for each of these endpoints. The higher VE for PHN (57%) than for community HZ (33%) suggests some incremental benefit of the vaccine in prevention of PHN beyond prevention of HZ per

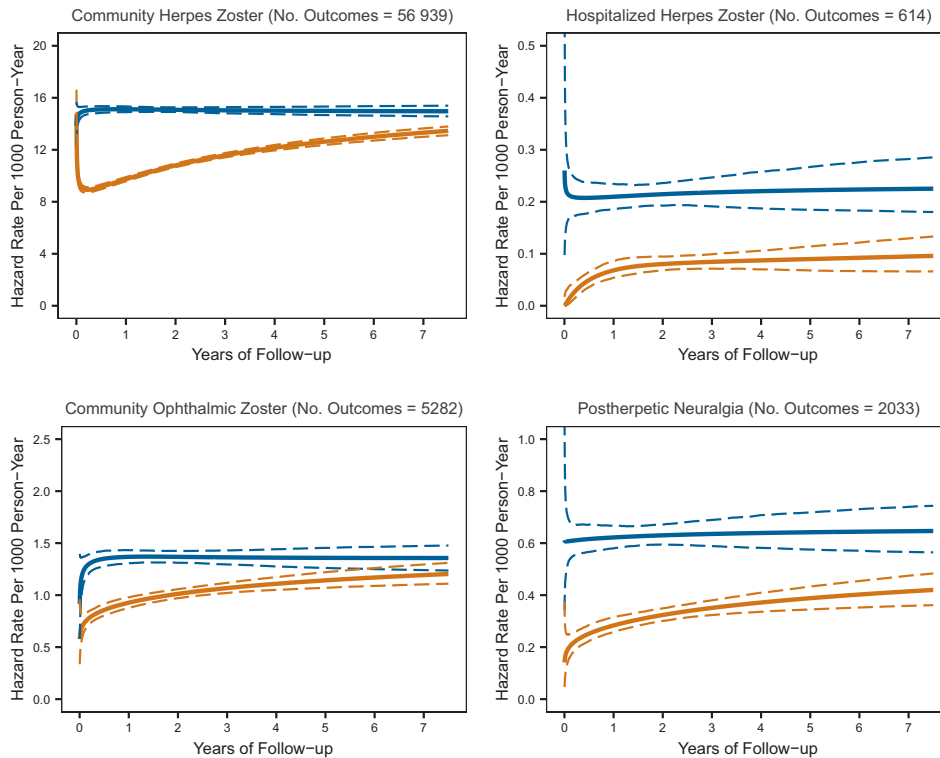
se; this is also suggested by the higher proportion of HZ cases who had PHN, a complication of HZ, among unvaccinated relative to vaccinated cohorts (4.0% vs 2.9%, respectively).

These results were consistent for the primary and secondary analyses; nonetheless, the different estimates could also, in part, reflect the influence of differences in ascertainment, or other potential biases that could have differentially influenced detection of each endpoint in vaccinated versus unvaccinated beneficiaries.

Although the Shingles Prevention Study reported declines in VE with age, our study found only limited evidence of age-related declines in VE, which is a pattern previously reported [1, 11]. The difference with the SPS may relate to case finding: HZ severity increases with age [1]. HZ severity would be less likely to affect case finding in the SPS clinical trial with its active phone-based surveillance, but it would in medical sector-based studies such as ours. Although our results showed a small decline in VE for community HZ between age groups, there was no evidence of decline in other outcomes, which indicates that VE does not change substantially with increasing age. HZV also appears to protect Medicare beneficiaries despite their high burden of chronic illness. Overall, our VE results are consistent with those of other studies, although point estimates vary across studies due to differences in source population characteristics, case definitions and ascertainment, and other aspects of study methodology [11, 12]. In our study, as in other observational studies, the overall vaccine effectiveness estimates may be less accurate than relative VE estimates based on internally controlled comparisons (such as age-specific and time-specific changes in vaccine effectiveness).

In both the primary and secondary analyses, VE for community HZ declined significantly for years 4+ postvaccination to 19% and 22%, respectively, showing that HZV protection declines over time. Similar results have been published based on long-term follow-up of the original study cohort from the SPS, but the long-term follow-up study was uncontrolled and unblinded [12]. A recent prospective cohort study also showed waning of protection, though that study, conducted at a large health maintenance organization [27], had less analytic power and generalizability than our study. For PHN and HZ-associated hospitalization, protection was better preserved than for community HZ with AVEs of 45% and 55%, respectively, for years 4+ in the primary analysis. The differences found in duration of protection would need to be investigated further in other studies.

This study has several limitations. Because Medicare is an administrative database, outcomes and risk factors were not verified by medical record review. However, previous record validation studies have consistently shown relatively high positive predictive value for HZ as an outcome [11, 28–30]. Additionally, published population-based studies have indicated that over 90% of adults self-reporting HZ sought medical attention for the condition [31–34]. If measurement error were



Plots are cubic spline models using 1 knot (Royston and Parmar). Models were ran on matched cohorts without adjusting for covariates.

**Figure 2.** Primary analysis: fitted hazard rates and 95% confidence interval (CI) using a cubic spline model for vaccinated and unvaccinated cohorts.

a problem, we would expect VE for nonspecific outcomes like community HZ to be understated. However, when we analyzed OZ, which is a particularly distressing form of HZ that would almost certainly require medical attention, we found that HZV was similarly effective at preventing community OZ as it was for preventing community HZ. In general, analyzing more specific outcomes in observational studies may sometimes introduce biases by focusing on particular segments of the population. For example, HZ-attributable hospitalization is an imperfect surrogate for HZ disease severity, as even primary-HZ coded hospitalizations are not always attributable to HZ [15]. The rates of HZ-associated hospitalization may be influenced by other risk factors for hospitalization not directly due to HZ. In an effort to address this limitation we explored a more specific outcome with an analysis that restricted community HZ to cases receiving antiviral treatment, and found that estimated VE was only modestly increased (eTable 6).

We believe that this study is unlikely to have much error in measuring vaccination status. Our cohorts are restricted to beneficiaries enrolled in Medicare from the time HZV was licensed so that any HZV would very likely be recorded in Medicare claims. It is also unlikely that many Medicare beneficiaries would choose to pay for this vaccine out-of-pocket, so Medicare claims are likely to be nearly complete for this population. However, because we used a conservative continuous-enrollment criteria,

the average age of our cohorts was older than that for Medicare as a whole; thus, VE estimates in our study populations might have been lower than for the general Medicare population.

Vaccinees may differ from nonvaccinees with respect to both disease risk and healthcare-seeking behavior. Community HZ, in particular, can be missed among beneficiaries who tend to forgo medical care for nonserious conditions. To ensure comparability between cohorts, we matched by multiple potential confounders and conducted a vaccinee-vaccinee analysis as a secondary analysis. Falsification outcomes provide a novel supplementary approach for identifying bias. Our falsification outcomes clustered around the null (1.0), suggesting that there was no systemic bias influencing relative outcomes in vaccinees versus controls. Our aggregate use of this large number of falsification outcomes was intended to provide a more robust comparison group, under the expectation that it is unlikely that all these independently-selected conditions would share causal pathways with HZ. Nonetheless, examination of the results of the individual conditions and their variance can provide clues with which to detect and better understand residual differences between vaccine and control groups.

In the secondary analysis, VE for community HZ in the first 3 years postvaccination was only slightly higher than in the primary analysis (37% vs 33%) and was similar to the primary analysis for all other outcomes. This suggests that

any potential bias related to healthcare-seeking behavior is minimal. Indeed, although falsification outcomes did cluster more tightly around the null in the secondary analysis compared with the primary, the difference was slight. The 67% VE estimate for HZ-associated hospitalizations for the first 3 years postvaccination in the secondary analysis was only 7% lower than in the primary analysis, and differences in VE for HZ-associated hospitalizations were not statistically significant (Table 3b). Further research would be needed to confirm this.

In summary, we observed that the duration of protection against community HZ wanes over time, yet the vaccine protects against HZ regardless of age and chronic illness. We calculated that the number needed to vaccinate to avert an episode of community HZ during our study period was just 30.6 (eTable 8). Our findings provide additional evidence that the live HZV is effective at reducing the incidence of HZ and its complications when used in clinical practice among Medicare beneficiaries. Although effectiveness against community HZ was not high, the protection provided by this vaccine was higher and more durable for HZ-hospitalizations and PHN.

## Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online. Consisting of data provided by the author to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the author, so questions or comments should be addressed to the corresponding author.

## Notes

**Acknowledgements:** We thank Xiyuan Wu from Acumen LLC for assistance with programming and data analysis.

**Funding:** This work was funded by the Food and Drug Administration. Additional funding was provided by the office of the Assistant Secretary of Planning and Evaluation.

**Potential conflicts of interest:** All authors: no reported conflicts. The authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

## References

- Oxman MN, Levin MJ, Johnson GR, et al.; Shingles Prevention Study Group. A vaccine to prevent herpes zoster and postherpetic neuralgia in older adults. *N Engl J Med* **2005**; 352:2271–84.
- Pinchinat S, Cebrián-Cuenca AM, Bricout H, Johnson RW. Similar herpes zoster incidence across Europe: results from a systematic literature review. *BMC Infect Dis* **2013**; 13:170.
- Kawai K, Gebremeskel BG, Acosta CJ. Systematic review of incidence and complications of herpes zoster: towards a global perspective. *BMJ Open* **2014**; 4:e004833.
- Tyring SK. Management of herpes zoster and postherpetic neuralgia. *J Am Acad Dermatol* **2007**; 57(6 Suppl):S136–42.
- Kost RG, Straus SE. Postherpetic neuralgia—pathogenesis, treatment, and prevention. *N Engl J Med* **1996**; 335:32–42.
- Edmunds WJ, Brisson M, Rose JD. The epidemiology of herpes zoster and potential cost-effectiveness of vaccination in England and Wales. *Vaccine* **2001**; 19(23–24):3076–90.
- Yawn BP, Wollan P, St Sauver J. Comparing shingles incidence and complication rates from medical record review and administrative database estimates: how close are they? *Am J Epidemiol* **2011**; 174:1054–61.
- Merck & Co., ZOSTAVAX (Zoster vaccine live): important information regarding storage and handling [package insert]. Available at: [http://www.merck.com/product/usa/pi\\_circulars/z/zostavax/zostavax\\_pi.pdf](http://www.merck.com/product/usa/pi_circulars/z/zostavax/zostavax_pi.pdf). Accessed October 1.
- Marin M, Yawn BP, Hales CM, et al. Herpes zoster vaccine effectiveness and manifestations of herpes zoster and associated pain by vaccination status. *Hum Vaccin Immunother* **2015**; 11:1157–64.
- Tseng HF, Lewin B, Hales CM, et al. Zoster vaccine and the risk of postherpetic neuralgia in patients who developed herpes zoster despite having received the zoster vaccine. *J Infect Dis* **2015**; 212:1222–31.
- Tseng HF, Smith N, Harpaz R, Bialek SR, Sy LS, Jacobsen SJ. Herpes zoster vaccine in older adults and the risk of subsequent herpes zoster disease. *JAMA* **2011**; 305:160–6.
- Morrison VA, Johnson GR, Schmader KE, et al.; Shingles Prevention Study Group. Long-term persistence of zoster vaccine efficacy. *Clin Infect Dis* **2015**; 60:900–9.
- Tseng HF, Tartof S, Harpaz R, et al. Vaccination against zoster remains effective in older adults who later undergo chemotherapy. *Clin Infect Dis* **2014**; 59:913–9.
- Langan SM, Smeeth L, Margolis DJ, Thomas SL. Herpes zoster vaccine effectiveness against incident herpes zoster and post-herpetic neuralgia in an older US population: a cohort study. *PLoS Med* **2013**; 10:e1001240.
- Jackson LA, Reynolds MA, Harpaz R. Hospitalizations to treat herpes zoster in older adults: causes and validated rates. *Clin Infect Dis* **2008**; 47:754–9.
- Klompas M, Kulldorff M, Vilk Y, Bialek SR, Harpaz R. Herpes zoster and postherpetic neuralgia surveillance using structured electronic data. *Mayo Clin Proc* **2011**; 86:1146–53.
- Leung J, Harpaz R, Molinari NA, Jumaan A, Zhou F. Herpes zoster incidence among insured persons in the United States, 1993–2006: evaluation of impact of varicella vaccination. *Clin Infect Dis* **2011**; 52: 332–40.
- Zhang J, Xie F, Delzell E, et al. Association between vaccination for herpes zoster and risk of herpes zoster infection among older patients with selected immune-mediated diseases. *JAMA* **2012**; 308:43–9.
- Rubin DB, Thomas N. Combining propensity score matching with additional adjustments for prognostic covariates. *J Am Stat Assoc* **2000**; 95:573–85.
- Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* **2010**; 21:383–8.
- Prasad V, Jena AB. Prespecified falsification end points: can they validate true observational associations? *JAMA* **2013**; 309:241–2.
- Jackson LA, Jackson ML, Nelson JC, Neuzil KM, Weiss NS. Evidence of bias in estimates of influenza vaccine effectiveness in seniors. *Int J Epidemiol* **2006**; 35:337–44.
- Zaadstra BM, Chorus AM, van Buuren S, Kalsbeek H, van Noort JM. Selective association of multiple sclerosis with infectious mononucleosis. *Mult Scler* **2008**; 14:307–13.m
- Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. *Biometrics* **2005**; 61:962–73.
- Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Stat Med* **2004**; 23:2937–60.
- Schmader KE, Levin MJ, Gnann JW, Jr, et al. Efficacy, safety, and tolerability of herpes zoster vaccine in persons aged 50–59 years. *Clin Infect Dis* **2012**; 54:922–8.
- Tseng HF, Harpaz R, Luo Y, et al. Declining effectiveness of herpes zoster vaccine in adults aged ≥60 years. *J Infect Dis* **2016**; 213:1872–5.
- Mullooly JP, Riedlinger K, Chun C, Weinmann S, Houston H. Incidence of herpes zoster, 1997–2002. *Epidemiol Infect* **2005**; 133: 245–53.
- Jumaan AO, Yu O, Jackson LA, Bohlke K, Galil K, Seward JF. Incidence of herpes zoster, before and after varicella-vaccination-associated decreases in the incidence of varicella, 1992–2002. *J Infect Dis* **2005**; 191: 2002–7.
- Donahue JG, Choo PW, Manson JE, Platt R. The incidence of herpes zoster. *Arch Intern Med* **1995**; 155:1605–9.
- Soni A, SC Hill. Average Annual Health Care Use and Expenses for Shingles among the U.S. Civilian Noninstitutionalized Population, 2003–2005 Washington, DC: Agency for Healthcare Research and Quality, **2007**.
- Bowsher D. The lifetime occurrence of herpes zoster and prevalence of post-herpetic neuralgia: a retrospective survey in an elderly population. *Eur J Pain* **1999**; 3:335–42.
- Lu PJ, Euler GL, Jumaan AO, Harpaz R. Herpes zoster vaccination among adults aged 60 years or older in the United States, 2007: uptake of the first new vaccine to target seniors. *Vaccine* **2009**; 27: 882–7.
- Hales CM, Harpaz R, Bialek SR. Self-reported herpes zoster, pain, and health care seeking in the Health and Retirement Study: implications for interpretation of health care-based studies. *Ann Epidemiol* **2016**; 26:441–446.e3.



# Examining the Impact of Real-World Evidence on Medical Product Development: A Three-Part Workshop Series

## Workshop Three: Application

### Real-World Evidence Study Design Decision Aids

#### *DISCUSSION DRAFT*

The following draft decision aids are intended to help stakeholders evaluate opportunities to use real-world data (RWD) and real-world evidence (RWE) for medical product decision-making and make informed choices about the design of prospective or retrospective studies, primarily for regulatory review, that involve RWD and RWE (akin to clinical decision aids that are designed to help patients make informed decisions about treatment options). There are no right or wrong answers to a given question. Instead, the draft decision aids are intended to lay out key questions for stakeholders to consider early on to help them make thoughtful choices about the development and design of rigorous, but manageable, RWE studies that relevant parties (e.g., patients, clinicians, researchers, sponsors, regulators, payers) agree in advance will generate reliable results.

Some points to keep in mind:

- The draft decision aids are a starting point for workshop discussions and were informed by discussions that took place during the [first](#) and [second](#) workshops in this series.
- Questions ideally would capture relevant information about the potential risks associated with either the treatments themselves or tradeoffs associated with particular decisions; costs in terms of monetary investment, time investment, and/or patient and clinician investment; and reporting and transparency expectations for showing study methods and results.
- The draft decision aids could be further modified and refined to be broadly applicable across study types (e.g., studies on medical products to treat prevalent chronic diseases), account for different types of datasources (e.g., electronic health records, claims data, patient-generated data), and “future-proofed” to accommodate new sources of data going forward (e.g., data from mobile devices, the internet of things).

Each of the draft decision aids highlight a different topic in study design: 1) when a particular real-world data element is fit to assess study eligibility, treatment exposure, or outcomes; 2) considerations for obscuring intervention allocation in trials to generate real-world evidence; 3) considerations for controlling or restricting treatment quality in real-world trials; and 4) assessing and minimizing bias in observational comparisons.

A few preliminary questions to consider before beginning a study:

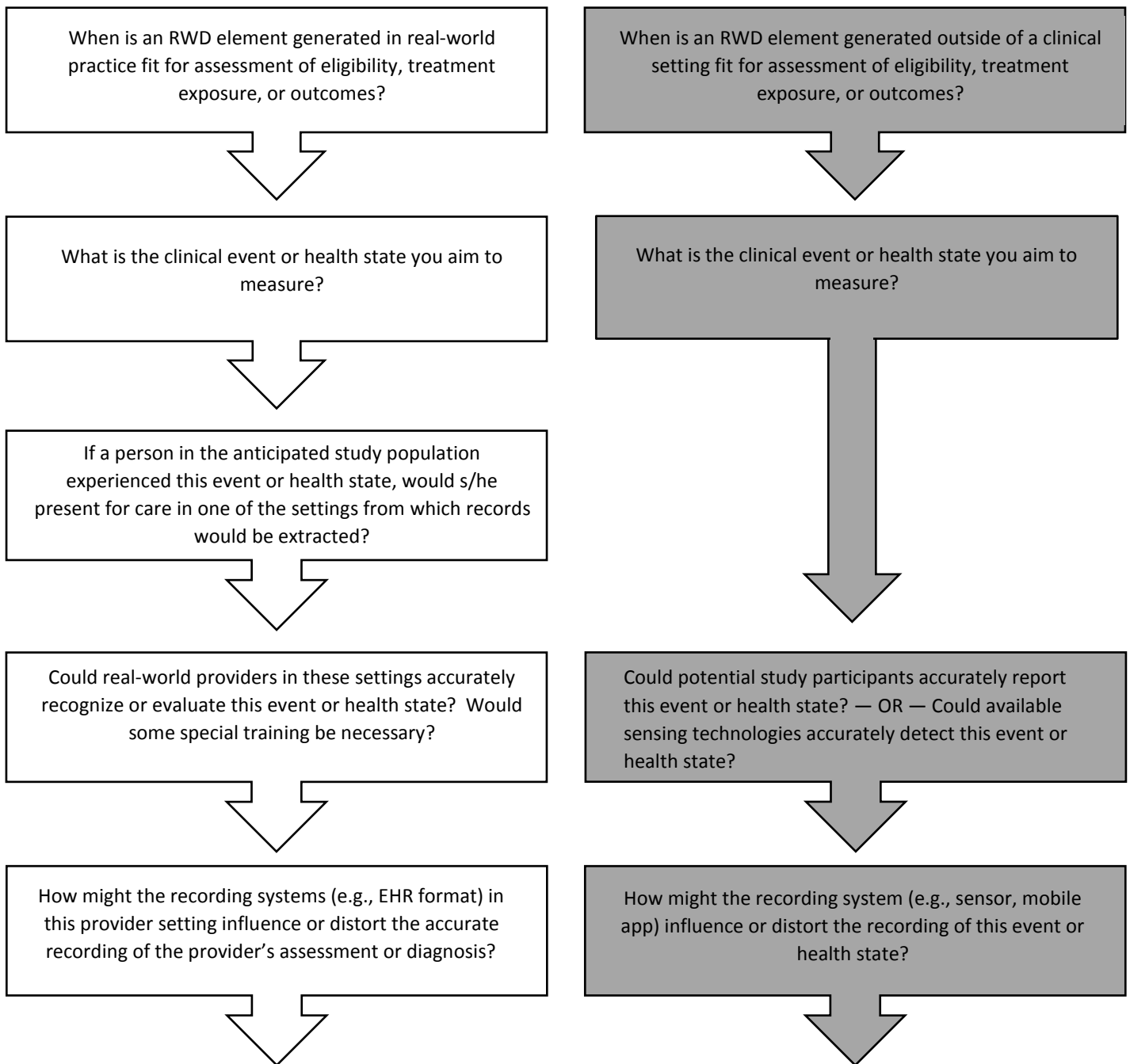
- What is the specific regulatory, clinical, or policy question this real-world evidence should address (effectiveness, comparative effectiveness, or comparative safety)?
- Should the question be framed in terms of superiority or non-inferiority?
- Does there appear to be sufficient power to address the research question? Ideally, this is addressed by querying actual data from likely study settings regarding:
  - Cohort identification (application of inclusion and exclusion criteria); and
  - Primary endpoint in pooled cohort (to obtain overall expected event rate).



---

## WHEN IS A REAL-WORLD DATA ELEMENT FIT FOR ASSESSMENT OF ELIGIBILITY, TREATMENT EXPOSURE, OR OUTCOMES?

---

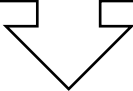


---

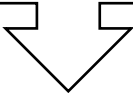
# WHEN IS A REAL-WORLD DATA ELEMENT FIT FOR ASSESSMENT OF ELIGIBILITY, TREATMENT EXPOSURE, OR OUTCOMES?, CONTINUED

---

Could data from different settings or recording systems be harmonized – in both technical and semantic senses (e.g., is there a common data model or standard)?

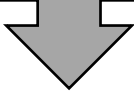


Could the source data elements (e.g., diagnosis codes, procedure codes, clinical text) be consistently reduced to a useful clinical phenotype?

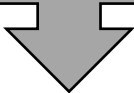


Is the “chain of custody” from source data to research data element transparent and traceable?

Could data from different recording systems be harmonized – in both technical and semantic senses (e.g., is there a common data model or standard)?



Could the source data elements (either sensed or participant-reported) be consistently reduced to a useful clinical phenotype?



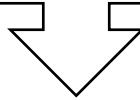
Is the “chain of custody” from source data to research data element transparent and traceable?

---

# OBSCURING INTERVENTION ALLOCATION IN TRIALS TO GENERATE REAL-WORLD EVIDENCE: WHY, WHO, AND HOW?

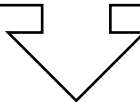
---

What expectations or preferences providers and patients be expected to have regarding benefits and adverse effects of study interventions?

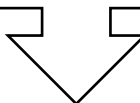


How might those preferences or expectations influence:

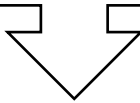
- Intervention uptake or adherence?
- Fidelity or intensity with which an intervention is delivered?
- Likelihood that beneficial or adverse effects would be reported or observed?



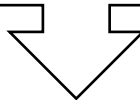
How might those expectations or preferences differ in the settings where trial results will eventually be applied?



How might concealing treatment allocation from patients and/or providers reduce biases due to preferences or expectations?



How might concealing treatment allocation from patients or providers obscure meaningful differences between interventions?



How might procedures necessary to conceal treatment allocation from patients and/or providers impact:

- The acceptability of trial participation to patients and or providers?
- The cost of the trial?
- The risk/benefit ratio of trial participation?

---

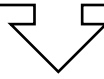
# HOW TIGHTLY SHOULD INVESTIGATORS ATTEMPT TO CONTROL OR RESTRICT TREATMENT QUALITY IN A PRAGMATIC OR “REAL WORLD” TRIAL?

---

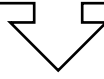
## Starting assumptions for applying this decision aid

- The study question is clearly defined, including the decision to be made and decision-maker the study should inform.
- Data are of adequate quality to assess eligibility, key prognostic factors, treatment exposure, and outcomes.
- Treatments are assigned randomly or by some other method that supports valid inference.

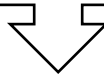
How much would the effectiveness or safety of the study treatment(s) vary among providers or care settings? How is this variability related to different levels of resources, experience, or expertise?



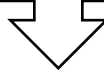
What level(s) of resources/experience/expertise are now present in the care settings in which results of this trial will be applied?



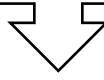
What level(s) of resources/experience/expertise are now present in the care settings in which this trial could be conducted?



What special vulnerabilities or risks are anticipated in the study population?



Is there some minimal or floor level of treatment quality necessary for valid inference regarding the study question?



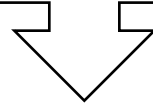
Is there some minimal or floor level of treatment quality necessary to assure participant safety?

---

## HOW CAN BIAS IN OBSERVATIONAL COMPARISONS BE ASSESSED AND MINIMIZED?

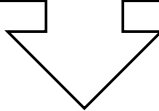
---

What is the clinical and epidemiologic justification for the comparator selected (and the margin, if applicable)?

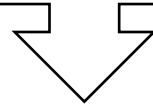


Does there appear to be appropriate balance between the treatment cohorts after matching/weighting? At this stage, there should be no consideration of outcomes by treatment group.

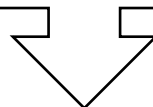
- Display a plot of propensity score distributions for each treatment group (if appropriate).
- Justify weighting methods if used.
- Provide two tables that report covariate balance before and after matching or weighting, respectively.



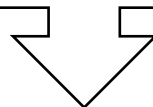
After matching or weighting for balance, do the analytic cohorts appear to represent clinically meaningful groups for study (e.g., has utility or generalizability been sacrificed)?



Are there specific unmeasured confounders thought to be sufficiently influential that they might alter the statistical inference from the study? Is there a supplemental way to measure these confounders? If not, can sensitivity analyses be designed to examine their potential influence?



How can reporting be structured to enable replication by a regulator or another research team?



Where has the study been registered prior to initiation (e.g., clinicaltrials.gov)? If it is a regulatory study, or a study initiated by a regulatory agency, which regulatory agencies have examined the protocol?

February 2018

## Examining the Impact of Real-World Evidence on Medical Product Development: I. Incentives

Proceedings of a Workshop—in Brief

On September 19–20, 2017, the National Academies of Sciences, Engineering, and Medicine held the first workshop of a three-part series titled Examining the Impact of Real-World Evidence on Medical Product Development. The workshops are convened under the auspices of the Forum on Drug Discovery, Development, and Translation and sponsored by the U.S. Food and Drug Administration (FDA). The workshops are intended to advance discussions and common knowledge among key stakeholders about complex issues relating to the generation and use of real-world evidence (RWE). The first workshop focused on how to align incentives to support the collection and use of RWE in health product review, payment, and delivery and how to address the barriers that impede the uptake and application of RWE.

Gregory Simon of the Kaiser Permanente (KP) Washington Health Research Institute told the workshop participants that establishing a common language will be key to understanding and changing the traditional paradigm of evidence generation, and he emphasized that real-world data (RWD) are distinct from RWE. He noted that RWE may come from many sources, including randomized controlled trials (RCTs). Andrew Bindman of the University of California, San Francisco, and Robert Califf of Duke University and Verily Life Sciences said that all stakeholders want access to scientifically derived evidence, but there may be different ways to obtain such evidence and the required degree of confidence in RWE may depend on the needs of the person making a decision. Simon characterized the core qualities of RWE as:

- It is generalizable to correctly predict an outcome for patients; the ability to assess the accuracy of the prediction after a patient was treated imbues an implicit accountability to the evidence-generating system.
- It is relevant to decision makers' specific information needs and stems directly from their priorities. This implies that it is “fit for purpose,” meaning that the evidence is designed to answer the question regardless of its source.
- It is adaptable to embrace the heterogeneity in RWD.
- It is efficient in the sense that the evidence can be produced more quickly and less costly than through traditional methods; this efficiency is necessary because answering fit-for-purpose questions requires the generation of more evidence types.

### ADVANCING PUBLIC HEALTH OPPORTUNITIES WITH REAL-WORLD EVIDENCE

Scott Gottlieb of FDA, the workshop's first keynote speaker, laid out FDA's current thinking about RWE. Gottlieb said that RWE is being more widely used for coverage and reimbursement decisions and its rigor is therefore increasing. “As the breadth and reliability of RWE increases, so do the opportunities for FDA to make use of information,” he said. Clinical care choices are made based on many sources of information that have varying degrees of uncertainty. FDA could therefore support the development of and access to reliable evidence that meets standards for approval. Gottlieb

emphasized that FDA will uphold and promote the “gold standard” for evidence; however, the source of that evidence is not mandated.

Gottlieb said that pre- and postmarket evaluations should be thought of as parts of a continuum rather than as two separate and distinct processes; in particular, he said, the “need for a point of regulatory accountability” should not preclude the possibility of evaluating products over their life cycle of use. RWE offers a way to better inform the benefit–risk profiles for medical products and is already used routinely by FDA to evaluate safety and emerging risk.

To encourage regulated industry to take the risk on RWE, FDA is taking steps to offer more clarity by issuing final guidance for devices and it is developing policies that support the use of RWE in indication expansions, especially in cases of unmet medical need and rare diseases and in meeting postapproval requirements. Rachel Sherman of FDA added that the goal is to achieve “a better evidence base on which to make our medical product approval or clearance decisions.”

## SEEING THE DESTINATION

During the workshop’s first session, speakers and discussants explored the knowledge base that may be necessary to make informed decisions about the use of medical products. Bindman introduced the speakers, noting that if one seeks to redesign the evidence-generation system, it is critical to hear from multiple perspectives.

### A Payer Perspective

Michael Sherman of Harvard Pilgrim Health Care said that Harvard Pilgrim strives to find a balance between treatment access and affordability for patients while also encouraging treatment innovation. M. Sherman said that payers want to pay for new treatments proven to add value for patients, but these therapies must be selected carefully with cost in mind (e.g., payers should evaluate whether a given treatment would lead to premiums increases for employers or individuals). Ultimately, he said, payers should focus on treatments backed by good data and evidence, while also considering the patients’ perspective when determining coverage. M. Sherman acknowledged that some conditions are too rare for treatments to be tested in RCTs, which creates an ethical tightrope for payers.

M. Sherman advocated for manufacturers to enter into value-based agreements for reimbursement that are tied to product success. He suggested that manufacturers could be required to submit data to third-party value analysts, such as the Institute for Clinical and Economic Review,<sup>1</sup> so that pricing could be assessed in a value-based way. Finally, M. Sherman said that payers could encourage FDA to collect postmarketing data for cases in which there are limited data on a given treatment—a practice the agency often already engages in and one that creates RWD that can be used to better study patient outcomes.

### An Integrated Health Care System Perspective

Michael Horberg of KP described the perspective of a health care delivery system on RWE, focusing on KP’s integrated care model. Horberg said that KP makes care-altering decisions after considering the answers to several key questions, which would also apply to RWD assessment and RWE generation:

- What are the efficacy and effectiveness of a new treatment?
- Who conducted the research?
- What is the population at risk?
- What does KP do currently to address a disease/condition?
- What is the added cost of the new treatment?
- If implemented, how would the change in practice be operationalized?

Horberg said that KP recognizes that a gap exists between clinical trial efficacy and real-world effectiveness of new treatments; it also recognizes, he said, the cultural tendencies for patients and providers to rely on anecdotal evidence in making care decisions. Horberg explained that in KP’s integrated care model, decisions to change practice may be based on one of several factors: new knowledge that appeared in the medical literature, pharmacy or physician requests, patient demand, changes to state or federal statutes, and especially KP’s own data or research findings. Changes in practice require the input of multiple stakeholder groups that think and act independently, with the hope that opinions will ultimately converge about treatment recommendations, benefits decisions, and formulary decisions.

---

<sup>1</sup> <https://icer-review.org> (accessed November 20, 2017).



## **An Academic Health System Perspective**

Daniel Ford of the Johns Hopkins University School of Medicine presented an academic health care delivery system perspective. He described the Johns Hopkins Health System (JHHS) Corporation—which includes 2.5 million people in its electronic health records (EHRs) and 300,000 people enrolled in clinical protocol—as an evidence generator and evidence consumer. He said that the system incorporates and relies on both internal and external evidence. In generating evidence, Ford said, JHHS still relies heavily on traditional clinical research, which is often funded by the National Institutes of Health (NIH). However, recent efforts to conduct clinical research at community hospitals opened up new opportunities to gather clinical evidence. Ford discussed JHHS’s data needs, noting that while the system’s data warehouse is improving, it would be difficult to rely solely on its observational data to make judgments about clinical effectiveness. He mentioned several collaborative efforts, including the Chesapeake Regional Information System for our Patients,<sup>2</sup> a health information exchange that serves Maryland and Washington, DC, and said that such efforts provide an important data tool for Johns Hopkins both in research and clinical practice.

Ford discussed the roles and expectations of patients and providers in the generation of RWE. Patients often desire information about their treatment plans in order to inform their personal decision making, but this information is not always accessible in the current environment. Academic researchers, meanwhile, frequently express interest in researching drug effects on off-label indications and applying their findings to usage recommendations, but providers still rely on traditional RCT evidence rather than other potential sources of evidence. Ultimately, Ford said, altering the evidence-generation system will require changes on the part of providers.

## **A Patient Perspective**

Sharon Terry of Genetic Alliance reminded participants that the term “patient” implies a power imbalance between the practitioner and the individual seeking treatment, saying that many health professionals would approach decisions differently if they put their personal health before business or industry preferences. Patients, she said, are concerned about their own health, and groups of them have created disease- or condition-centric communities to represent their interests and generate data to inform health care decisions. However, the data generated by these groups are often discounted when they should be systematically included with other, traditional sources of data, she said.

Terry described the importance of community-based registries to her work and to the evolution of RWD. She noted that community registries, such as those run through PCORnet<sup>3</sup> or Genetic Alliance’s Platform for Engaging Everyone Responsibly (PEER),<sup>4</sup> are distinct from industry- or hospital-led registries in that they are led by community members, they respect and highlight the priorities and concerns of the individuals participating in the registry, and they focus on the education of members through consistent outreach over social media and other media. Terry admitted that one area in which community registries could improve would be to provide more rigorous validation and she suggested that developing a methodology for this could become a priority since it is critical to establish a RWE-generating system that takes into consideration the lived experience of those seeking treatment.

## **A Reaction from Data Generators**

Following the panel, invited discussants Joanne Waldstreicher of Johnson & Johnson and Eleanor Perfetto of the National Health Council (NHC) provided additional observations in reaction to the presentations. Waldstreicher agreed with Terry’s emphasis on creating evidence that maximizes value and minimizes risk for individuals seeking treatment. Waldstreicher encouraged the increased sharing of data between stakeholders as a way of developing an improved understanding of various medical products. The follow-up analysis of product data post-licensure is an important step in building a learning health care system, she said.

Perfetto referred workshop participants to an NHC white paper stemming from a July roundtable on RWE and patient perspectives.<sup>5</sup> She highlighted several takeaways from the meeting and paper. First, patients often do not distinguish between RWE and RCT-derived evidence; they care primarily about answering the questions “Will this work for me, and [will I experience] side effects?” Second, Perfetto said, patients believe that they own their own data and deserve to dictate who can use them and when, in an “opt-in” environment. Third, she said that patients do not view data from EHR and claims as authentic patient data because those data do not reflect patients’ preferences or experiences in a meaningful

<sup>2</sup> <https://www.crisphealth.org> (accessed November 20, 2017).

<sup>3</sup> <http://www.pcornet.org> (accessed November 20, 2017).

<sup>4</sup> <http://www.geneticalliance.org/programs/biotrust/peer> (accessed November 20, 2017).

<sup>5</sup> <http://www.nationalhealthcouncil.org/sites/default/files/Patient%20Perspectives%20on%20Real-World%20Evidence.pdf> (accessed November 20, 2017).

way. Fourth, Perfetto said, patients need clearer definitions of RWD and RWE if they are to be credible stakeholders and make informed decisions. Last, she emphasized Terry's point that patient communities are a great source of RWD and one that should not be discounted in the generating of evidence from a diverse set of sources.

## FIT-FOR-PURPOSE EVIDENCE

Mark McClellan of the Duke–Margolis Center for Health Policy gave a keynote address on fit-for-purpose evidence after the Session I panel and discussion, reiterating many of the points made by the Session I panelists. He noted that fit-for-purpose evidence was a topic that Duke–Margolis addressed at a separate September 2017 workshop<sup>6</sup> and in a September 2017 white paper.<sup>7</sup> Addressing the uncertainty concerning terminology, he offered a crucial distinction between RWD and RWE: RWD are related to patient health status and the delivery of care, are routinely collected through clinical records or similar sources, and should be credible and trustworthy. RWE is derived by culling RWD by applying sound, rigorous analytical methods to RWD. McClellan, referring to a framework developed in the Duke–Margolis white paper,<sup>8</sup> emphasized that successfully generating fit-for-purpose evidence relies on understanding the regulatory and clinical contexts surrounding the data.

Concluding his presentation, McClellan said that steps being taken to develop fit-for-purpose and patient-centered RWE not only may lead to new clinical and regulatory decision-making tools, but also could support the development of new payment models based on the value and quality of care. Ultimately, he said, the use of RWE is limited compared to the amount of RWD currently available, so further stakeholder investment in this area is needed.

## LEARNING FROM SUCCESS

In Session II, Learning from Success, workshop participants discussed examples of the successful use of RWE in decision making. The presentations focused on how the methods and techniques used in the successful examples could apply to future applications.

### **Generalizing and Scaling the Salford Lung Studies**

Martin Gibson and Marie Kane of Northwest eHealth discussed the Salford Lung Studies (NASEM, 2017; Vestbo et al., 2016; Woodcock et al., 2017). Gibson said that these studies demonstrated a way to bridge the traditional gap between pre- and postmarketing data collection.

Gibson and Kane emphasized that the keys to the success of these trials were coordinating and connecting patient care in the hospital with care in community health centers and planning the trials from the beginning to answer questions from regulators, payers, and researchers. This required working before the studies began to engage or train the health system, sponsor, regulators, providers, pharmacists, and information technology departments. Kane said that these studies required far more investment in data processing and error management than traditional clinical studies. This was due to variability in the data as they were collected as well as the scale and complexity of the data linkages required to determine patient outcomes. Kane said that it was important to focus on collecting the right data rather than on collecting more data. Kane and Gibson reported that Northwest eHealth has converted the platform developed for the Salford studies into individual, cloud-based “configurable and modular” applications to make it easier to adapt the infrastructure for future studies.

During the discussion period, participants debated how the Salford model could function as a “franchise” that could be exported to other health systems and other disease questions. Kane added that the “franchise” part of the studies would consist of the methods developed and the lessons learned from each variation. John Graham of GlaxoSmithKline (GSK) agreed that the studies had generated both a reusable infrastructure and lessons learned that could be applied in other disease areas; GSK is already applying a similar model in cardiovascular and renal disease studies in the United States. Gibson added that recruiting health systems to participate became easier with each study because of the positive experiences of the study participants and the investments made in building relationships as well as the desire of health systems to participate in exciting, new, relevant studies. He said that experience alleviated the fear of doing a trial among community hospitals and other systems where research is not the main focus.

<sup>6</sup> <https://healthpolicy.duke.edu/events/public-workshop-framework-regulatory-use-real-world-evidence> (accessed November 20, 2017).

<sup>7</sup> [https://healthpolicy.duke.edu/sites/default/files/atoms/files/rwe\\_white\\_paper\\_2017.09.06.pdf](https://healthpolicy.duke.edu/sites/default/files/atoms/files/rwe_white_paper_2017.09.06.pdf) (accessed November 20, 2017).

<sup>8</sup> *Considerations for generating RWE fit for regulatory purposes* (p. 10), [https://healthpolicy.duke.edu/sites/default/files/atoms/files/rwe\\_white\\_paper\\_2017.09.06.pdf](https://healthpolicy.duke.edu/sites/default/files/atoms/files/rwe_white_paper_2017.09.06.pdf) (accessed November 20, 2017).

## Using Sentinel to Evaluate Effectiveness or Efficacy

Richard Platt of Harvard Medical School discussed Sentinel's design and its potential for use in assessing efficacy and effectiveness in addition to safety. Sentinel<sup>9</sup> is built on a highly curated, distributed data network with a common data model used across the network (NASEM, 2017). While at this time Sentinel is used primarily for safety surveillance, monitoring effectiveness is also part of its mandate. Platt shared six examples of how Sentinel has been, or is being, used to answer questions of efficacy and effectiveness in addition to questions concerning safety.

1. Sentinel data alone were used to assess the comparative safety of rivaroxaban versus warfarin.<sup>10</sup> Comparing the results of that study with the ROCKET-AF RCT demonstrated that the observational analyses in Sentinel can correlate well with RCT results (Patel et al., 2011). Platt said that this technique may offer a mechanism to explore populations that are not well represented in RCTs.
2. Sentinel linked with adjudicated medical records was used to assess intussusception with rotavirus vaccine (Yih et al., 2014). This required expert adjudication to verify exposure to the vaccine and presentation with symptoms and it resulted in a change to the vaccine label.
3. Sentinel in combination with registries was used to link infants with their mother's health records.<sup>11</sup> Linking Sentinel data with state registry data incrementally improved the results over what was possible with Sentinel data alone. Echoing experience from the Salford studies, Platt said that this illustrates the benefit of collecting the right data.
4. Sentinel-like data are being linked to EHRs in PCORnet's ADAPTABLE trial to assess low-dose versus high-dose aspirin for the prevention of coronary artery disease.<sup>12</sup>
5. Sentinel linked to patient-generated data is being tested in the development of a mobile app for patients to collect and report information, such as information about the daily lives of pregnant women, that will later be merged with the Sentinel databases.<sup>13</sup>
6. Sentinel is being tested as a platform for randomized trials through IMPACT-Afib (Pokorney, 2017). This trial will reach out directly to Sentinel members who have atrial fibrillation and a high risk of stroke but no evidence of oral anticoagulant use.

During the discussion period, workshop participants considered which aspects of Sentinel could inform RWD/RWE use. Platt and R. Sherman pointed out that everything done in Sentinel is in the public domain and is intended as a public resource. The distributed data model allows data partners to choose whether to contribute their data to each inquiry. However, Platt said that this model is expensive to maintain and is only successful if the data partners continue to find it useful.

## Applying Lessons Learned from Device Registries to Other Treatment Types

Rachael Florence of the National Evaluation System for Health Technology (NEST) Coordinating Center spoke about using RWE in devices and how the lessons learned from such use could be applied to other treatment types.

- The processes of device approval and regulation are different from the corresponding processes for drugs in a number of ways, including that there are several different pathways to approval for devices; different devices may have different requirements for postmarket safety studies; it is difficult to track device implantation because the system of unique device identifiers only began in 2015 and these identifiers are not universally required in EHRs; adherence is not a concern for implanted devices; it can be difficult to disentangle the "learning curve effect" of providers iteratively improving their implantation techniques with practice from actual problems with the device; and the current surveillance system for problems with devices depends on voluntary reporting.

<sup>9</sup> <https://www.sentinelinitiative.org> (accessed December 4, 2017).

<sup>10</sup> <https://www.sentinelinitiative.org/sites/default/files/Communications/Publications/Sentinel-ICPE-2017-Presentation-Rivaroxaban.pdf> (accessed December 4, 2017).

<sup>11</sup> <https://www.sentinelinitiative.org/sites/default/files/Sentinel-ICPE-2017-Presentation-PRISM-Mother-Infant-Cohort.pdf> (accessed December 4, 2017).

<sup>12</sup> ClinicalTrials.gov: NCT02697916: <https://clinicaltrials.gov/ct2/show/NCT02697916> (accessed December 4, 2017).

<sup>13</sup> <https://www.sentinelinitiative.org/sites/default/files/Communications/Publications/Sentinel-ICPE-2017-Presentation-Mobile-App.pdf> (accessed December 4, 2017).

- The device community has some experience already with RWD from the widespread use of registries. Device registries have historically been important in device regulation, in part due to mandates for registries for expensive, high-risk device implants.
- Registries have been considered by some in medical product development as a potential answer to a number of questions about how to use RWD more widely, Fleurence said. The benefits of registries include the provision of high-quality, fit-for-purpose data; easy linking with other data sources through coordinated registries networks; the ability to support both pre- and postmarketing observational studies, potentially at lower costs; and the potential to serve as a platform for automated safety surveillance
- However, Fleurence said, registries are not a blanket solution to be applied to all treatment types, or even to all devices, because of several significant drawbacks that registries have. They are expensive to develop and maintain; they are not practical in some kinds of treatments and patient populations; they vary in their data quality and methods; they can pose significant administrative challenges; and, as is the case with other data sources, those who operate registries need to grapple with safeguarding patient privacy and security.

Rather than further developing registries that may become increasingly burdensome as they become larger, Fleurence suggested that the key to RWE use in the future could be to focus on developing opportunities to work more directly with the systems generating data in the course of clinical care or at home. In device development and use, RWD could be used to generate robust postmarket data to support earlier premarket decisions, help researchers recognize and assess safety problems sooner, help medical professionals determine better ways of using a device, and help researchers design rigorous studies that will be able to reliably detect safety and efficacy outcomes. NEST was established, Fleurence said, to promote these applications and “serve as a catalyst to support timely and reliable development of high-quality RWE.”

## GETTING UNSTUCK: ALIGNING INCENTIVES

In Session III, Getting Unstuck: Aligning Incentives, workshop participants discussed incentives for maintaining the current data generation processes and potential barriers to the use of new methods of evidence generation. Anna McCollister-Slipp of the Scripps Translational Research Institute and VitalCrowd, Inc., framed the discussion by saying that the biggest barriers to the system as a whole are a lack of a sense of urgency to accept RWE, which she suggested is primarily due to bias favoring the traditional evidence generation system and limiting access to data after they have been generated. Like Terry and Perfetto, she observed that there are consequences to a heavy reliance on RCTs for data generation and it is no longer reasonable to exclude the consideration of patient-generated data sources.

### **A Contract Research Organization Perspective**

John Doyle of QuintilesIMS told the workshop participants that contract research organizations (CROs) are interested in using RWD/RWE to improve the process and increase the efficiency of delivering trials as well as to design better studies. He said that CROs are interested in implementing RWE studies in a scalable and systematic way, and he added that when CROs start a new study, they consider the needs and requirements of regulators, policy makers, payers, patients, and others who must make decisions about a medical product. RWD have already been used to optimize recruitment, to shorten the time it takes to start a study, and to reduce costs through risk-based monitoring. Furthermore, Doyle said, using RWD/RWE to bridge the evidence discrepancies between clinical trial patients and real-world patients could answer questions about subpopulations for precision therapy treatments or demonstrate proof of value to payers and patients. Doyle suggested that FDA’s recently released final guidance on RWE for devices could be a source of ideas for how other therapy modalities could incorporate RWE into study designs. He offered several examples of methods that blend the RWD and RCT approaches to evidence generation, such as running single-arm open trials with historical controls rather than concordant placebo study arms or pragmatic randomization designs.

### **A Product Developer Perspective**

Elliott Levy of Amgen Inc. discussed aligning incentives from the perspective of a product developer. Companies are already taking advantage of big data and RWD internally, he said, by using methods similar to those described by Doyle to improve product development, patient experience and outcomes, and value to the health care system. Levy emphasized that RWE and clinical trial evidence are not in conflict, but rather are complementary. Clinical trials inform development of rigorous RWE generation and RWE improves the pragmatism and relevance of clinical trials. Levy identified the barriers within companies that impede change as (1) a lack of knowledge and awareness of RWE methods, because product teams

are usually led and staffed by scientists without training or trust in those approaches, (2) a lack of talent and capabilities in the relevant areas, because there are few individuals in organizations who have experience in observational research and these tend to be found on safety or health economics teams rather than in product development, and (3) systems and processes are generally not set up for RWE, because organizations tend to be optimized around generating RCT data and new approaches can become unnecessarily complicated once they are fit with existing company processes, such as procurement. Levy emphasized that these barriers are all addressable. He suggested that senior leadership promote and support RWE adoption, companies invest in training team members in RWD management and analytics capabilities, and the company leadership directly address challenges in organization processes.

## **An Academic Researcher Perspective**

Ford discussed the misaligned incentives and barriers to RWE use from an academic researcher perspective. Ford said that younger investigators are often less willing to try new methods because they are more risk averse as they are establishing their careers. Furthermore, the career incentives for established investigators favor performing high-quality RCTs. RWD requires that researchers relinquish some control over a study, Ford said, because the analytical methods require more dependence on statistician colleagues and the nature of the data requires a willingness to accept less precise data as well as to discard extraneous data in EHRs. These practicalities of RWD use can be difficult to accept for researchers trained in RCT methods.

Ford suggested that pragmatic study designs might be a good way for traditional clinical trialists to gain experience with RWD. He acknowledged, however, that in addition to the initial time investments described by other presenters, it may be difficult to find partnering health systems because they are probably capable of accommodating only a few trials at a time and demand is increasing. In a different approach, Johns Hopkins is beginning to develop capabilities for physicians to query their own patients' EHR data, Ford said; the goal is to encourage greater interest in data collection.

## **A Big Data Perspective**

Marcus Wilson of Anthem's HealthCore observed that most patients get care from a highly fragmented health system. Payer companies often do not have the relevant evidence available to determine how RCT populations relate to their own patient populations, nor is the evidence generated until after a product is marketed. Wilson emphasized that the gravitational pull back to the familiar is a major underestimated systemic obstacle to overcoming this fragmentation. In organizations with large data sources like HealthCore, as well as in the developer companies as described by Levy, this pull can influence and affect decisions at every step. McCollister-Slipp added that this particular barrier also affects funding decisions made by reviewers at funding agencies, so the bias often extends to what types of studies can be run.

Wilson argued that the solution is to defragment the patient view by sharing data responsibly and creating value by linking data from disparate parts of the health system as well as patient-provided information. He said that institutions that collect and share these data should adhere to core principles, including protecting patient privacy and security, using data only for those purposes for which they are fit, and actively creating a learning health system. This data work, and its associated cost, can be planned for prior to marketing approval so that better decisions can be made by all stakeholders earlier. Wilson argued that these steps would benefit both patients and the business interests of the data sources.

## **The Perception of Evidence Hierarchies**

Workshop attendee Hui Cao of Novartis observed that the hierarchical rating of evidence by data source begins in medical school training and is further promoted in the peer-reviewed literature and asked workshop presenters for comments and possible solutions to this phenomenon. Wilson suggested a system that grades evidence by analytical method rather than by data source, and Waldstreicher commented that observational studies should incorporate higher standards for rigor, transparency, and reproducibility. Ford suggested publishing the costs of RCTs along with the resulting data to encourage consideration about whether the extra investment required is justified by the perceived improvement in data quality. Many individual workshop participants emphasized that the most important point was to find the right method, whether RCT or RWD techniques and data sources, for addressing each question. During his keynote address, Gottlieb said that the historical hierarchy of evidence is changing as the reliability of forms of evidence other than fully randomized, prospective, placebo-controlled trials increases with improvements in the methods used to evaluate them. He said that FDA could therefore support changes by releasing consensus definitions of terms and describing RWE and its applications for satisfying FDA requirements as part of a developing guidance document.



## The Potential of Building Trust

Wilson discussed the reluctance of health care systems to share data with other health care systems as one barrier to defragmentation. Gibson, Horberg, and Wilson attributed this reluctance to a lack of trust between the health systems and they emphasized the importance of establishing relationships to facilitate data sharing. Deven McGraw of the Office for Civil Rights at the U.S. Department of Health and Human Services agreed that creating trust was paramount for data sharing to develop RWE. She said that a framework could be in place to facilitate the development of trust and responsible data sharing and to reduce the uncertainty around putting patients at risk or violating the Health Insurance Portability and Accountability Act of 1996 (HIPAA). She suggested considering some type of credit or reward for organizations that were already doing this well.

## GETTING UNSTUCK: MYTHBUSTING

On the second day of the workshop, participants examined ideas and misconceptions about established evidence-generation practices. Califf opened the session with a keynote address about false precision and estimating the reliability of the evidence-generating process. He encouraged changing the goal of evidence generation from precision to reliability, which will require focusing on shedding practices and portions of the old system that increase cost without improving evidence quality and emphasizing rigorous science over standard operating procedures. Califf listed four key principles that could underlie any such evidence-generating system:

1. Build a reusable system embedded in clinical practice and learn from every encounter, but also ensure that lessons learned are actually spread to the point of care. Califf observed that several of the public–private partnerships and integrated health systems discussed on the first day of the meeting are developing this capability.
2. Use quality by design to eliminate errors that bias results and ignore those errors that do not affect the outcome so that effort and resources are spent efficiently.
3. Use automation for repetitive tasks, real-time analysis of comparison data embedded within health care, and infrastructure to share the results with practitioners to support a constantly learning system.
4. Operate from basic principles rather than merely establishing different standard operating procedures.

## A Data Aggregator Perspective

Patrick Ryan of Janssen Research and Development discussed some of the methods being developed at Observational Health Data Sciences and Informatics (OHDSI) (NASEM, 2017).<sup>14</sup> These methods and the related OHDSI databases support three types of analytic use cases: (1) clinical characterization to describe outcomes in a specific population, (2) patient-level predictions to anticipate what will happen for an individual patient, and (3) estimates of population-level effects for safety surveillance, comparative effectiveness, or causal inference.

Ryan performed a live demonstration of an analysis of results from published, peer-reviewed literature to show that even after years of research, meta-analyses conclude that the answers to many clinical questions are unknown. Ryan said that the purpose of this demonstration was to show that “we can’t necessarily trust the process that we are using to generate evidence as a community,” regardless of the data source. OHDSI databases contain raw data from four different sources in order to minimize bias and allow for easy comparisons. This systematic standardized approach of analyzing multiple data sources simultaneously and asking specific, fit-for-purpose questions can generate more trustworthy answers and potentially answer patients’ question of “What will work for me?”

## A Medical Product Developer Perspective

Graham said that the goal for developers is to have the right answers to the right questions at the right time. He emphasized that RWE is now necessary to answer many questions, but it is not a replacement for traditional research. He advocated for focusing on a challenge-based, holistic thinking process about what will improve an individual patient’s outcome rather than the traditional assessment of individual studies. RWE can help do this by informing which disease states to focus on in development, how to develop a particular treatment, or how to explain benefits and risks to patients in a meaningful way.

---

<sup>14</sup> <https://ohdsi.org> (accessed December 4, 2017).

## An Academic Researcher Perspective

Rory Collins of the University of Oxford said that randomized trials are necessary to detect moderately beneficial or adverse effects of new treatments and establish causality, particularly when trial populations are widely diverse. He argued, however, that under current regulations the burden of conducting randomized trials is too cumbersome, in large part because of the widespread misapplication of the good clinical practice (GCP) guidelines issued by the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. Collins said that these guidelines incentivize a focus on standard operating procedures, rather than on innovation in the design and conduct of studies, and on verifying and adjudicating source data rather than on generating reliable results. He therefore advocated for a focus on developing evidence-based strategies to make conducting randomized trials easier. Other individual workshop participants noted that a purpose of GCP was to give providers guidelines for how to conduct research well. Those participants said that in the context of pragmatic or real-world trials, this becomes a mission to train providers to recognize whether to intervene with an individual patient's treatment course.

## A Regulator Perspective

Janet Woodcock of FDA reiterated that current data generation methods are costly and time consuming and often cannot generate the evidence necessary to answer the relevant question. Woodcock said that FDA is committed to exploring the use of RWE in regulatory decisions and she cited the recently issued guidance for devices. She said that the incentive embedded within that guidance was for developers to invest in developing RWE methods for regulatory purposes. As drug development changes with the onset of precision medicine and there is a push for accelerated development programs, the need for innovation in study methods is becoming more apparent. Woodcock emphasized the promise of hybrid RCT–RWD approaches to studying investigational drugs and she pointed to the NIH collaboratories for examples of using innovative method applications. She also discussed innovative trial designs, such as master protocols or platform trials, as promising ways to incorporate RWD. To carry out innovative trial designs will require additional work in standardization, verification, and training as well as potentially different strategies for development or funding structures and academic rewards, but these designs offer tangible opportunities to more easily adopt new practices in the clinic; to answer multiple questions simultaneously, including comparative effectiveness; and to maintain the focus on patients.

## FINAL REMARKS

Simon concluded by commenting that change is held back not by greed, but by fear. Traditional evidence-generating methods fail in familiar ways and so are perceived as more reliable than RWE. The benefits to the broader community may not translate to benefits for the data generator. Finally, Simon echoed the discussion about the difficulty of building trust across stakeholders. He listed potential next steps discussed by some workshop participants which could help address these challenges and encourage the wider use of RWE. First, Simon said that asking fit-for-purpose questions will be critical in determining context-specific value in the health care system. Second, he stressed the importance of using appropriate methods, including making choices about when randomization is needed to answer a particular question. Regarding observational studies, he said that elevating the rigor of trial designs through transparency and sharing of methods and data will become more important. Finally, Simon suggested accommodating diverse evidence needs across stakeholders, defining smaller studies based on simple questions and sound research design, and reconsidering the delineation between the preapproval and postapproval process.◆◆◆

## REFERENCES

- NASEM (National Academies of Sciences, Engineering, and Medicine). 2017. *Real-world evidence generation and evaluation of therapeutics: Proceedings of a workshop*. Washington, DC: The National Academies Press.
- Patel, M. R., K. W. Mahaffey, J. Garg, G. Pan, D. E. Singer, W. Hacke, G. Breithardt, J. L. Halperin, G. J. Hankey, J. P. Piccini, R. C. Becker, C. C. Nessel, J. F. Paolini, S. D. Berkowitz, K. A. A. Fox, R. M. Califf, and the ROCKET AF Steering Committee. 2011. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *New England Journal of Medicine* 365(10):883–891.
- Pokorney, S. 2017. *Sentinel IMPACT-AFib: Transforming pragmatic clinical trials using a nationwide distributed claims database*. Presented at the Clinical Effectiveness Research Innovation Collaborative Meeting, Washington, DC.
- Vestbo, J., D. Leather, N. Diar Bakerly, J. New, J. M. Gibson, S. McCorkindale, S. Collier, J. Crawford, L. Frith, C. Harvey, H. Svedsater, and A. Woodcock. 2016. Effectiveness of fluticasone furoate–vilanterol for COPD in clinical practice. *New England Journal of Medicine* 375(13):1253–1260.



- Woodcock, A., J. Vestbo, N. D. Bakerly, J. New, J. M. Gibson, S. McCorkindale, R. Jones, S. Collier, J. Lay-Flurrie, L. Frith, L. Jacques, J. L. Fletcher, C. Harvey, H. Svedsater, D. Leather, D. Adams-Strump, L. S. Addlestone, A. Afshar, J. Amin, R. Archer, M. Austin, A. Bakhat, J. Behardien, J. M. Borg-Costanzi, G. Breen, N. Browne, C. Brunt, K. H. Buch, P. Budden, J. Chandy, A. Chaudhry, L. Cheema, N. Chennupati, S. Coulson, L. Cribbin, D. Dillon, A. El-Kafrawy, E. Elliott, B. Farooq, N. A. Finegan, P. Fink, A. Fletcher, S. Frier, C. Gibbons, L. Gill, D. Herron, B. Hope, R. E. Howard, C. Hughes, S. Iles, P. Jackson, M. Jarvis, V. Joshi, N. Kanumilli, R. A. Khan, M. Khan, S. Kwok, N. Lord, C. Mafunga, C. I. Malcomson, D. K. McCarthy, H. S. Milligan, P. Patel, S. J. Patel, V. B. Raj, K. A. Richardson, R. Salim, R. B. Seaton, D. Shah, M. Sharma, H. Singh, N. Smith, N. N. Smyrniou, M. Stamp, P. Stratford-Smith, M. Sultan, R. S. Sumra, J. Tankel, U. I. N. Umeadi, C. Westwood, J. White, H. C. Wilkinson, R. G. Wilson, S. A. Wright, and A. T. Wright. 2017. Effectiveness of fluticasone furoate plus vilanterol on asthma control in clinical practice: An open-label, parallel group, randomised controlled trial. *The Lancet* 390(10109):2247–2255.
- Yih, W. K., T. A. Lieu, M. Kulldorff, D. Martin, C. N. McMahill-Walraven, R. Platt, N. Selvam, M. Selvan, G. M. Lee, and M. Nguyen. 2014. Intussusception risk after rotavirus vaccination in U.S. infants. *New England Journal of Medicine* 370(6):503–512.

**DISCLAIMER:** This Proceedings of a Workshop—in Brief was prepared by **Amanda Wagner Gee, Ben Kahn, and Carolyn Shore** as a factual summary of what occurred at the workshop. The statements made are those of the rapporteurs or individual workshop participants and do not necessarily represent the views of all workshop participants; the planning committee; or the National Academies of Sciences, Engineering, and Medicine.

**REVIEWERS:** To ensure that it meets institutional standards for quality and objectivity, this Proceedings of a Workshop—in Brief was reviewed by **John Doyle**, QuintilesIMS; and **Rachael Fleurence**, National Evaluation System for Health Technology (NEST) Coordinating Center. **Lauren Shern**, National Academies of Sciences, Engineering, and Medicine, served as the review coordinator.

**SPONSOR:** This workshop was partially supported by the U.S. Food and Drug Administration.

For additional information regarding the workshop, visit <http://nationalacademies.org/hmd/Activities/Research/DrugForum/2017-SEP-19.aspx>.

Suggested citation: National Academies of Sciences, Engineering, and Medicine. 2018. *Examining the impact of real-world evidence on medical product development: I. incentives: Proceedings of a workshop—in brief*. Washington, DC: The National Academies Press. doi: <https://doi.org/10.17226/25024>.

## Health and Medicine Division

*The National Academies of*  
SCIENCES • ENGINEERING • MEDICINE

The nation turns to the National Academies of Sciences, Engineering, and Medicine for independent, objective advice on issues that affect people's lives worldwide.

[www.national-academies.org](http://www.national-academies.org)

July 2018

## Examining the Impact of Real-World Evidence on Medical Product Development: II. Practical Approaches

Proceedings of a Workshop—in Brief

### INTRODUCTION

On March 6–7, 2018, the National Academies of Sciences, Engineering, and Medicine held the second workshop of a three-part series titled Examining the Impact of Real-World Evidence on Medical Product Development. The workshops are convened under the auspices of the Forum on Drug Discovery, Development, and Translation and are sponsored by the U.S. Food and Drug Administration (FDA). The workshops are intended to advance discussions and common knowledge among key stakeholders about complex issues relating to the generation and use of real-world evidence (RWE). The second workshop focused on practical approaches for the collection of real-world data (RWD)—data generated outside of the traditional clinical trial setting—and the use of RWE.

Workshop discussions centered around three framing questions to help shift the discussion toward practical and generalizable considerations for those embarking on a study that would use RWE to inform decision making in the development and evaluation of medical products:

1. When can decision makers rely on real-world data?
2. When can decision makers rely on real-world treatment?
3. When can decision makers learn from real-world treatment assignment?

Gregory Simon of the Kaiser Permanente (KP) Washington Health Research Institute opened the workshop by explaining that while RWE may mean something different to distinct stakeholders, the goal of the workshop was to discuss possible questions for stakeholders to refer to when considering the use of RWE in a study. The questions suggested by individual workshop participants on Day 2 of the workshop (March 7, 2018) following discussions on Day 1 (March 6, 2018) are listed in Boxes 1–3. The statements, recommendations, and opinions expressed are those of individual presenters and participants and they should not be construed as reflecting any group consensus.

### WHEN CAN DECISION MAKERS RELY ON REAL-WORLD DATA?

#### **Novel Oral Anticoagulants (NOACs) in Comparison with Warfarin**

Adrian Hernandez of the Duke University School of Medicine presented on a suite of trials that investigate the use of NOACs compared to warfarin, a drug typically used in patients to prevent atrial fibrillation and reduce the risk of stroke. The following trials, which used RWD to compare treatments, served as a starting point for discussion:

- Randomized Evaluation of Long-Term Anticoagulation Therapy (RE-LY) (Connolly et al., 2009);
- Rivaroxaban Once Daily Oral Direct Factor Xa Inhibition Compared with Vitamin K Antagonism for Prevention of Stroke and Embolism Trial in Atrial Fibrillation (ROCKET AF) (Patel et al., 2011);
- Apixaban for Reduction in Stroke and Other Thromboembolic Events in Atrial Fibrillation (ARISTOTLE) (Granger et al., 2011); and
- Effective Anticoagulation with Factor Xa Next Generation in Atrial Fibrillation–Thrombolysis in Myocardial Infarction 48 (ENGAGE AF-TIMI 48) (Giugliano et al., 2013).

Hernandez explained that a 2014 meta-analysis of these trials demonstrated that NOAC treatment led to better outcomes for stroke and thromboembolic events and showed non-inferiority to warfarin. Hernandez said that all four NOAC trials favored NOACs over warfarin when it came to risk of stroke and systemic embolic events, as well as secondary outcomes, such as ischemic stroke, hemorrhagic stroke, myocardial infarction (MI), and all-cause mortality (Ruff et al., 2014). Using the ROCKET trial as an example, Hernandez emphasized the importance of assessing trial quality in order to judge the reliability of the associated outcome measurements and pointed to the frequency of enrolled patients in the therapeutic range for the various treatments as a possible surrogate for trial quality.

In conclusion, Hernandez emphasized that study cost is an important factor. Trials assessing a product early in its lifecycle include significant data collection and thus are expensive, whereas studies assessing a product later in its lifecycle—but still aiming to include an indication—can streamline study protocol and data collection, leading to lower costs. Hernandez ended by posing a question for consideration: What questions characterize the utility of an RWD source and signal reliability before a study is performed?

## Discussion

During the discussion session, Hernandez and Jesse Berlin of Johnson & Johnson emphasized the importance of predetermining the population, exposure, and outcome of interest when conducting a study using RWD. Hernandez argued that careful data curation and characterization are necessary to reduce systematic bias. Simon agreed, calling for ongoing data curation and monitoring. Similarly, some panelists and participants discussed validating data as a crucial step in using RWD for research. Berlin said that coding is not always accurate, so validated algorithms and methods should be used to accurately identify patient cohorts. He also suggested that a library of validated algorithms to use during large-scale studies would be a useful research tool. Berlin and other participants also raised the point that algorithms may change over time because coding may change over time (including the introduction of new versions of coding and the implementation of new coding terminology). Robert Califf of Verily Life Sciences and Duke University said that when validating the quality of data researchers must be aware that there are multiple sources of data for one individual and it may actually be irresponsible to ignore RWD because non-traditional sources could actually capture more accurate information about a person (e.g., such as in the case of sensitive information he or she may not share with study coordinators).

Hui Cao of Novartis noted that there are some common “rules” that could help assess uncertainty in RWD sources. For example, for certain diseases there are established algorithms that identify data for a given patient population. She pointed out that exposure can be predicted with a high level of confidence for injection or intravenous administration, but for oral drugs or inhaled products—for which exposure may be harder to determine—researchers may need to rely more on larger sample sizes. Cao pointed out that different metrics for determining outcome will vary in terms of accuracy. For example, certain events or lab measures—such as hospitalization for MI and HbA1c measure—can be identified with higher accuracy than other events or measures such as moderate asthma exacerbations and lung function.

Khaled Sarsour of Genentech commented on the lack of alignment between randomized controlled trial (RCT) endpoints and RWE endpoints, noting that RCT endpoints are not always relevant to real-world practice. Hernandez highlighted that there should be more consideration for patient preferences and experience in medical practice. Gregory Daniel of the Duke-Margolis Center for Health Policy remarked that this misalignment only highlights the importance of RWE as a tool for a diverse group of stakeholders, including patients, payers, and providers, and the need to utilize the full suite of available evidence to evaluate a medical product.

Daniel brought to attention the use of data from mobile apps and other new sources. Hernandez said that patient-imported data require further consideration and integration in the evidence generation system. However, Daniel and Christina Stephan of the American Medical Informatics Association warned that it is hard to rely on mobile data in particular when social determinants and lifestyle factors, such as the presence of pets or the misuse of mobile devices, affect the accuracy of the data. John Burch of the Mid-America Angels Investment Group noted that existing data sources will continue to change over time and the methods used to extract and analyze data will have to change as well.

Cathy Critchlow of Amgen asked about the role of pre-specifying data analysis methodology and its necessity during the collection and analysis of RWD. Hernandez suggested that this should be done as much as possible. He emphasized that pre-specification and other efforts to create transparency around data collection make RWD much more trustworthy.

## Discussion of Potential Questions Around Real-World Data

On Day 2 of the workshop, individual participants discussed potential questions for consideration when contemplating the use of RWD in a study. The starting point for the discussion was the central question from session 1: When can decision makers rely on real-world data?

### BOX 1

#### When Can Decision Makers Rely on Real-World Data? Questions Raised by Individual Participants on Day 2 of the Workshop<sup>a</sup>

##### Data Collection and Evaluation

- What questions can characterize the utility of any RWD source and signal reliability before a study is performed?
- Is information recorded along the spectrum of care (e.g., data collected upon patient presentation and subsequent treatment, provider assessment of the condition, clinical outcome)?
- Does the record system accurately capture information (e.g., consider the transmission from the source record to a database)?
- What are the standards for data collection and how do these compare to traditional RCT methods?
- What are potential sources of systemic bias? Are there other concerns with data entry and collection?
- Are the data tailored for a particular stakeholder, such as a payer or a regulator?
- What steps have been taken to address data deficits (such as missing data) for prospective studies?
- What guidelines for data curation could be applied at the source of data generation?
- Does the RWD source capture clinically meaningful outcomes (particularly when it comes to data from new technology sources)?
- Can community providers—based on their training and ability, the *International Classification of Diseases* (ICD) code definition, general practice, and other factors—accurately assess the clinical phenomenon of interest? Is there an incentive to do so?
- Could the relevant outcome be captured in a traditional clinical trial?
- How accurate is the final diagnosis of a condition as captured in the data, given that a diagnosis can change over time as more information is gathered about a patient?
- How does the data source affect the definition of outcomes, because any data source may be tailored to the needs of the primary group using it?
- Are there concerns about data collection or entry, particularly in relation to creating systemic bias?
- Are there any major safety issues that would not be captured by the RWD source?

##### Transparency

- With consideration for transparency, what is the path to generating data?
- What information is needed about data extraction methods to be able to evaluate them? Is it possible and appropriate to pre-specify?
- When are the standards for data collection and evaluation “good enough,” and/or “transparent enough” to consistently produce reliable results?
- What measures have been implemented to account for data privacy concerns (e.g., how are data handled for multi-site studies)?

<sup>a</sup> DISCLAIMER: This document represents discussion by individual workshop participants of the Examining the Impact of Real-World Evidence on Medical Product Development—Workshop II: Practical Approaches. The statements made are those of the individual workshop participants and do not necessarily represent the views of all workshop participants; the planning committee; or the National Academies of Sciences, Engineering, and Medicine.

### Validation and Accuracy

- When is expert adjudication necessary to evaluate whether the recorded data are reliable and/or reasonably complete?
- When is accuracy good enough to reasonably and consistently identify the right patient population?
- When is accuracy good enough to reasonably and consistently assess treatment exposure or intervention, understanding there is variation across treatment types?
- When is accuracy good enough to reasonably and consistently assess clinical outcome?
- Are there expected outcomes that could be used to validate the results of an RWD study?
- What building blocks for data curation can also be applied at the source of data generation?
- Are the results of the RWD study too remarkable to be true?

### Other

- Where is the product in its lifecycle of development and is there enough information known about it to be able to assess safety using RWD?
- Could RWD sources shed light on the public health impact of the disease or intervention (consider the need for long-term assessment, social determinants of health, and lifestyle influences)?
- What are additional considerations when it comes to the use of consumer-generated data and data collected through the “Internet of things”—the network of mobile devices, vehicles, home appliances, and other items embedded with electronics, software, sensors, and connectivity that enables these objects to connect and exchange data?

During the Day 2 discussion, several participants discussed incomplete ascertainment, due to health system issues (e.g., data fragmentation) and the likelihood of certain events occurring outside of a patient’s typical health system, as a possible source of bias when collecting data. Other participants mentioned the direction of possible misclassification, given the knowledge that random misclassification biases result in different directions based on the study type (non-inferiority versus superiority), as another potential source of bias during data collection.

On the topic of expert adjudication, Hernandez mentioned that tracking minor events such as nosebleeds or bruising is a much different task than tracking significant events like major bleeding. He noted that the data collection and reporting process is often much more uncertain and variable across institutions for minor events. Joanne Waldstreicher of Johnson & Johnson pointed out there are many examples of the use of expert adjudication in the cardiovascular field. Grazyna Lieberman of Genentech commented on the importance of transparency around data sources, including the source itself, the procedures used to extract data, the completeness and accuracy of the data, and the consistency of the data collection methodology used across research.

Elise Berliner from the Agency for Healthcare Research and Quality addressed the collective need for better data and data infrastructure, recognizing that this affects not only industry and FDA but also payers. Simon agreed that even with different uses for data—clinical, quality improvement, payment, product comparison, etc.—there is shared interest and alignment across the health care system in the quality of original data. Echoing comments from Day 1, Andy Bindman of the University of California, San Francisco, said that certifying rules around the curation of data at the front end of collection could be useful. Jacqueline Corrigan-Curay of FDA commented that there should be a greater focus on prospective data collection and analysis, rather than just retrospective data.

Last, Frank Sifakis of AstraZeneca noted the importance of increased data collection for the purpose of routine business decisions, as well as supporting providers and patients in making treatment decisions. Sheila Weiss from Evidera said that while many companies already engage in such collection, the next step should be leveraging that data to identify appropriate situations when observational data can be used to make decisions about a product’s efficacy or effectiveness.

## WHEN CAN DECISION MAKERS RELY ON REAL-WORLD TREATMENT?

### Lithium for Suicidal Behavior in Mood Disorders: A Classic RCT Addressing a Real-World Question

Ira Katz of the Corporal Michael J. Crescenz Veterans Affairs Medical Center presented on the U.S. Department of Veterans Affairs (VA) study investigating the use of lithium for the prevention of suicide among VA patients with major depression.<sup>1</sup>

<sup>1</sup> More information about the study is available at <https://ClinicalTrials.gov/ct2/show/NCT01928446> (accessed May 16, 2018).

Katz explained that the use of lithium to prevent suicide has long been hypothesized, but it has been difficult to carry out an RCT due to safety and treatment control concerns. Previous RCTs were subject to a negative indication bias due to the sensitivity of lithium dosing. Katz said at least two prior studies were terminated because of issues with recruitment, but previous research indicated support for the hypothesis that lithium can prevent suicide among patients with bipolar disorder and depression.

Katz explained that the VA is the ideal setting for a trial on this topic because of its large population of study-qualified patients and numerous study centers dispersed nationwide. While depression is required in the inclusion criteria for the study, patients with comorbidities—including histories of substance abuse and posttraumatic stress disorder—are allowed as well. Doing so, Katz said, is an admission that most patients at risk for suicide experience co-morbidities and investigators should consider how previous RCTs in this field have filtered out patients who are at risk for suicide primarily because of comorbidities.

To determine the effect size needed for the study, Katz said that study investigators surveyed health care providers about the necessary effect size to modify practice, which indicated that a fairly low effect size would still be an improvement. Even with this threshold, monitoring patients can be complex. For example, Katz cited a patient who was recruited for the trial and who received shipments of the medication despite not submitting monthly blood samples. In this case, the study care team followed good clinical management for the patient rather than the requirements of the RCT. Katz concluded by emphasizing a key monitoring question for studies being conducted in real-world settings: How can investigators reconcile the need to remain flexible while caring for patients with complex needs and the need to adhere to strict protocol in an RCT setting?

## Discussion

During the subsequent discussion, panelists and participants considered the lithium use case presented by Katz as well as the NOACs example presented earlier by Hernandez, which also illuminated considerations for real-world treatment, such as choosing an appropriate standard of care to compare to an investigational treatment, in addition to the data reliability issues discussed previously.

Workshop participants discussed desired standards of care in real-world treatment studies. Simon said that generally, care used in research should be at least as good as the typical standard of care in the real world. However, Simon, Hernandez, and Michael Horberg of the KP Mid-Atlantic Permanente Medical Group pointed out that “standard of care” varies significantly across different health systems and regions in the United States as well as abroad. Ultimately, Simon noted, it is the responsibility of researchers to identify and implement a reasonable standard of treatment for RWE studies. Simon and Jonathan Watanabe of the University of California, San Diego, also emphasized that the control treatment should be “better than average” in the real world. Watanabe noted that the real world is not static and therefore the comparator treatment should include a “floor” that is up to par with an ever-changing treatment environment, especially considering the variability in standards across clinical environments.

Sarsour addressed the topic of safety monitoring in real-world trials and noted that the value of safety monitoring is dependent on the lifecycle phase of the drug of interest. Lithium, he noted, has been used for decades and it is unlikely any major new safety information will come out of the VA’s study. Although Califf pointed out that collecting data on non-serious adverse events for treatments already used widely can be costly, Katz said that collecting that information—including events not related to the study outcomes—can still be useful because it may serve as an internal control indicative of the level of care in a particular study.

Participants grappled with concerns about patient behavior and adherence to treatment framed by Katz’s presentation. Expounding on the topic of complex patients, Simon and Horberg discussed non-adherent behavior in trials. Simon said that while human behavior, behavioral error, and comorbidities make science harder, they are also components of the real world that can make results more generalizable and relevant.

Horberg pointed to an example from a recent trial for pre-exposure prophylaxis (PrEP) in HIV prevention used by sero-discordant couples, during which couples were consistently advised to use condoms, but more than 100 pregnancies still occurred. The researchers took advantage of these situations to measure the couples for HIV transmission and assess the real-world effectiveness of the treatment. He said that this type of human behavior should generally be expected in real-world trials and suggested that real-world patient behavior should be incorporated into RWE studies and assessed as an endpoint to judge the success of a particular treatment.

While discussing NOACs, Robert Temple of FDA said that simpler care—as long as it does not come at the expense of quality—can help diminish problems around adherence. He highlighted long-acting injectable anti-psychotic drugs in contrast to oral anti-psychotic medication as an example. Sebastian Schneeweiss of Brigham and Women’s Hospital and



Harvard Medical School asked how much discretion regulators have to regulate clinical strategies around certain products that could be observed with RWD versus a specific molecule in a narrow sense. Ultimately, Temple and Horberg said the health care system is responsible for problems around patient adherence and compliance, a reality that is not acknowledged often enough. Sarsour suggested that adherence and compliance should be prioritized as the research community moves toward a more patient-centered approach.

Building off of the theme of compliance, Simon mentioned the concept of a “randomized nudge” trial as a possible solution for poor patient adherence. The trial framework would allow investigators to “nudge” prospective patients and providers toward a certain treatment, notifying them that a successful treatment is available for use without necessarily assigning them to a specific treatment group. Simon referenced the IMPACT-AFib trial<sup>2</sup> (Implementation of an RCT to Improve Treatment with Oral Anticoagulants in Patients with Atrial Fibrillation), which leverages this concept. Simon and Temple also debated the use of secondary randomization, in which a randomized group of “nudged” patients is compared to a group that is not “nudged,” and whether these types of randomized trials could help untangle confusion around confounding by indication versus compliance.

Several panelists and participants discussed the importance of blinding during different stages of the study cycle, including patient and/or provider blinding during treatment assignment, blinding of treatment group during analysis, and the blinding of assessors during outcome assessment. Simon argued that blinding of outcome assessment is critical, because the potential introduction of bias during outcome assessment is detrimental to trial results. Blinding during analysis is also important and could be done better, he said. Blinding patients and providers, however, may not be necessary and could introduce distortion of both the patient population in the trial and the true nature of the treatment. Katz commented that partially randomized patient preference designs would alleviate those distortions, offering an opportunity for those who declined participation in a blinded study the ability to choose a treatment, after which investigators could compare the outcome of the randomized group and the non-randomized group.

Critchlow asked how blinding affects clinical equipoise and whether the perception of equipoise could shift over the course of a study. She pointed out that accumulating information and potentially attributing it to one treatment could be particularly biased because the perceived effects may or may not be real. Simon related this question to the concept of unbiased outcome ascertainment, stating that if beliefs about a treatment would influence the reporting or ascertainment of the outcome, blinding becomes absolutely necessary. Overall, Simon and Temple both stressed the importance of blinding during outcome assessment but questioned its value for patients and providers during treatment, especially if it alters behavior. Watanabe pushed back against the value of the limited use of blinding. He said that if investigators make the effort to blind analysts and raters during outcome assessment to reduce bias, it seems like minimal investment to blind prescribers as well, because one could imagine a prescriber preferring a NOAC to warfarin because NOACs are less labor intensive to administer, thus biasing the results.

Individual workshop participants discussed some parameters of studies or treatments that could suggest when blinding would be helpful. Bill Potter of the National Institute of Mental Health commented that the need for blinding may change based on the subjectivity of the outcome being measured; for example, amyloid deposition would not change regardless of blinding. Hernandez observed that while blinding can ensure consistent behavior by patients and providers, it can also create a more homogeneous study population because of the burden of treatment and the possibility of receiving a placebo. Jennifer Graff of the National Pharmaceutical Council, discussing systemic bias in real-world treatment, mentioned that health care systems can encourage patients to behave a certain way or select certain treatments for non-clinical purposes. For example, reimbursement rules could nudge patients to choose one treatment option over another. Temple suggested there may be different considerations around blinding in high-risk versus low-risk populations.

## **Discussion of Potential Questions Around Real-World Treatment**

On Day 2 of the workshop, individual participants discussed potential questions for consideration when contemplating the role of real-world treatment in research. The starting point for the discussion was the central question from session 2: When can decision makers rely on real-world treatment?

During the discussion on real-world treatment during Day 2 of the workshop, several participants discussed possible lessons and considerations about study conditions in which patient self-monitoring is acceptable and safe. Some mentioned the potential effect of research location and control group selection (including the average quality of care for the control group) on results. Others mentioned the importance of distinguishing between bad outcomes and bad treatment and the likely role of simpler care protocols making trial results easier to interpret. Some participants discussed the ascertainment of treatment effects from different populations and commented that ascertainment is often easier from

---

<sup>2</sup> More information about the study is available at <https://clinicaltrials.gov/ct2/show/NCT03259373> (accessed May 17, 2018).



## BOX 2

### When Can Decision Makers Rely on Real-World Treatment? Questions Raised by Individual Participants on Day 2 of the Workshop<sup>a</sup>

#### Monitoring and Self-Reporting

- When conducting research in a real-world setting, are there situations that would require special guidance, knowledge, or experience in order for clinicians to adequately monitor participant safety and respond appropriately to adverse events?
- What are the conditions under which patient self-monitoring and reporting would be acceptable? Does this vary based on the stage of product development or on the baseline clinical knowledge regarding the use of a given medical product to treat particular patient types and/or treatment conditions?
- Are there generalizable lessons learned based on cases in which self-monitoring is acceptable and safe?
- Given that safety monitoring needs will depend on where a product is in its lifecycle, would the real-world study change understanding of the safety of the treatment?

#### Blinding

- When does blinding create noise and when does it create signal? For example, in a new formulation of a treatment that requires less frequent dosing than the standard formulation, is requiring identical dosing regimens to preserve blinding appropriate or would it obscure signal by masking a true difference in the treatment?
- How might knowledge of treatment assignment influence ascertainment?
- Under what circumstances is blinding necessary and would it be necessary for particular components of the study (e.g., patients, providers, outcomes)?
- How are real-world study outcomes influenced by the desire to achieve good results and how could blinding of outcome measures potentially affect this? How can we learn more about this connection?
- How does blinding affect clinical equipoise? Consider that blinding is harmful if it distorts treatment in a negative way, but beneficial if it helps ascertain the true nature of the effect.

#### Adherence

- What guidelines could help providers balance valid inference with conflicts of interest?
- What conditions and training would be needed to prepare clinical care providers to monitor patient safety outside of a tightly controlled RCT environment?
- How do investigators decide which variables require strict adherence to study protocol and which can be allowed to vary?
- How could treatments be adjusted to suit patients' needs, given that adherence can be difficult to achieve and individuals may not behave as investigators would hope in the real world?

#### Other

- When does variation between comparison groups (socioeconomic, demographic, etc.); in treatment fidelity; in provider behavior and preferences; or in adherence yield a valid signal about real-world effectiveness and when is it just noise?
- How could the definition of a "randomized trial" become broader and more adaptable for the real world (e.g., enrichment trials in the context of precision medicine)?

<sup>a</sup> DISCLAIMER: This document represents discussion by individual workshop participants of the Examining the Impact of Real-World Evidence on Medical Product Development—Workshop II: Practical Approaches. The statements made are those of the individual workshop participants and do not necessarily represent the views of all workshop participants; the planning committee; or the National Academies of Sciences, Engineering, and Medicine.

low-risk populations despite the possible ethical necessity of including high-risk populations in research. Last, several speakers said that from a research or regulatory perspective, collecting data on non-serious adverse events is not useful late in a product's lifecycle, but it can serve as a built-in checkpoint that patients are being cared for appropriately.

## WHEN CAN DECISION MAKERS LEARN FROM REAL-WORLD TREATMENT ASSIGNMENT

### Health Care Database Analyses of Medical Products for Regulatory Decision Making

Schneeweiss presented on non-interventional health care database analyses and their potential for the development of RWE. Schneeweiss described that within the world of biomedical data, there are broad categories, including RCT-derived data and non-interventional data. Non-interventional data can be broken down further into research data and transactional data (e.g., data from a claims database). Schneeweiss explained that transactional data is crucial to understand the benefits and risks of medical products outside of controlled research investigations in routine care settings.

Schneeweiss explained that given the expanding role of RWE in regulatory decision making—supporting secondary indications, adaptive pathways, safety, and other areas—the use of RWD analyses is expanding; in a project among Aetion, Inc.; Brigham and Women’s Hospital; and FDA, 30 RCTs that were designed for drug approvals are in the process of being identified for reproduction by database studies to understand how well the results can be predicted with RWD. Schneeweiss stated that when conducting RWD studies, discovering the unintended—including harmful—effects of a product can usually be executed with higher confidence than analyzing intended product effects because there is less patient selection regarding the unknown unintended effect, which reduces confounding.

Schneeweiss acknowledged that investigators prefer RCTs because of the amount of control and transparency they offer: RCTs include random treatment assignment, a controlled outcome measurement, and often clear and simple study implementation. However, he emphasized that there are certain situations that have a higher likelihood of being able to be studied with database analyses. These situations include

- An active comparator (which reduces confounding by design);
- Measurable exposures and outcomes; and
- Measurable confounders.

Schneeweiss cited the ONTARGET trial (which compared telmisartan to ramipril for the reduction of cardiovascular risk) as an ideal example to meet these conditions for RWD replication (Fralick et al., 2018; Yusuf et al., 2008), as demonstrated in an article he co-authored in 2018 describing the near replication of results from the ONTARGET trial using a health care database analysis. He also noted that transparent and structured reporting of the complex methodologies used in RWD studies is a way to achieve confidence when comparing non-randomized studies to RCTs. Emerging analytic tools are now built for transparency and reproducibility; meanwhile, line programming is considered error prone, is not able to be validated at scale, lacks transparency, and is hard to reproduce by other researchers. Ultimately, Schneeweiss posed several critical questions to workshop participants:

- Which questions are answerable using RWD analyses versus only RCTs?
- Are non-interventional RWD study protocols precise enough to translate intention into action?
- What are the factors that must be present to make database analyses reproducible?

To conclude, Schneeweiss presented a list of questions to consider—which he framed as a pathway—for embarking on an RWD analysis:

- Is the setting adequate for an RWD analysis?
- Is data quality fit for purpose?
- Is the data analysis plan based on epidemiologic study design principles?
- Was balance in confounding factors between treatment groups achieved?

### Discussion

Panelists discussed the role of randomization in clinical research. Califf said that while many single-armed studies already exist, the medical research community should embrace the use of randomization more than in the current environment, contrasting the lack of randomization in medical research to its widespread use among Silicon Valley technology companies and other business environments. However, he noted that randomization is not widely used in oncology research or accelerated treatment approvals and there may be applicable lessons to learn from that work. David Madigan of Columbia University acknowledged that while randomization is a key tool in clinical trial study design, there is a desperate need for more evidence to answer questions when RCTs simply are not a realistic solution, whether due to cost or feasibility. He offered assessing the long-term effects of bisphosphonates, a class of drugs used to prevent the loss of bone density, as an example.

Additionally, Madigan noted that randomization solves the central issue of unmeasured confounding, but it cannot overcome data source–derived problems such as missing data or systemic bias. Madigan said that in observational studies, propensity score matching—as first described in work by Paul Rosenbaum and Donald Rubin (Rosenbaum and Rubin, 1983)—can be used to mimic the effect of randomization on the dataset. Propensity scoring, he said, attempts to estimate the probability of being in a certain treatment group based on a variety of covariates, such as age, sex, pre-existing conditions, or comorbidities. Madigan also noted that a propensity score analysis would ideally be paired with a gamma analysis, a secondary analysis developed by Rosenbaum that quantifies the necessary size of an unmeasured confounder to invalidate study results. This type of analysis would make it easier to interpret propensity score matching and would also help both evidence generators and evidence users think more comprehensively about a study, Madigan said. Sarsour agreed that confounding by indication is often the key barrier to successful observational studies and said that better characterization and accounting for confounders would make observational studies more feasible. Martin stated that analyses like this would enable a thoughtful and quantitative consideration of both observational study results, as well as clinical phenomena that could be relevant and important enough to change the findings of a study.

Madigan and Schneeweiss agreed that the research community must overcome the barrier of unmeasured confounding by setting standards around data use. Schneeweiss noted that FDA’s Sentinel Initiative includes tools for expedited bias sensitivity analyses. Session moderator Richard Platt of Harvard Medical School reiterated this point, highlighting that the Sentinel Initiative articulates well-established principles of data use in advance of a study, requires transparent study designs, and incorporates data interpretation measures that take into account plausible alternative explanations for findings.

Madigan, Schneeweiss, and Platt mentioned a lack of reproducibility of some study results as an important factor reducing confidence in non-interventional RWD analyses. Referring to a commentary he wrote in the *Journal of the American Medical Association* (Califf, 2018), Califf warned that the interested community will need to avoid allowing researchers to conduct observational analyses repeatedly and in isolation until they find answers that match with RCT evidence. He said that with this concern looming, it is still difficult to trust observational data analyses as independently valuable compared to RCTs. Temple wondered how one could control for this problem and David Martin of FDA said that transparency in study approach is one important way to address it. Platt also noted that pre-specifying—as he described when discussing Sentinel—is one way to mitigate this concern. Simon suggested an “all by all” approach: Rather than limiting the number of analyses, the community could include and compare all analyses conducted. Simon said this approach would act as “simulating the null distribution.”

Cao commented on the perceived value of prospective observational studies. Cao said there is value in the studies if they leverage existing platforms and do not add to the burden of data collection already in place, but otherwise questioned the value of prospective observational work. Martin responded that he views prospective observational studies as one option in a broader menu of potential study designs. He also noted that a potential strength of prospective observational studies might be that they are better able to incorporate patient preferences through the realignment of incentives.

Graff asked how the business ecosystem will be able to use and trust health data in a machine-learning environment. Madigan noted that there is a large body of work on large-scale predictive modeling in health care and there has been substantial progress in recent years. Schneeweiss mentioned an example, referring to a case in which a health plan considered switching to a more expensive anti-diabetic drug that would have reduced total cost of care based on robust RWD analyses. He argued that it is critical for health plans and other business entities to utilize health care data appropriately, because they can positively affect both a company’s bottom line and clinical outcomes. In the case of this particular anti-diabetic drug, senior decision makers relied on RWE produced with their own data using an advanced RWD analytic platform. This switch ultimately resulted in better health outcomes and reduced the total cost of care for the health plan.

Watanabe asked how pharmacy benefit managers and other stakeholders could make the outcomes of RWD studies more central to the needs of real-world populations, given that populations analyzed in studies are often not reflective of the real world. Platt said there is more variation in the data available to researchers today than there was even a decade ago. Schneeweiss argued that the heterogeneity in analytic findings from health care databases today stems mostly from the provision of health care and its recording, rather than true genetic variation; he said that treatment administration as a clinical strategy, such as deciding dose adjustments or encouraging adherence, should become a larger focus.

Martin concluded Day 1 of the workshop by comparing current work around RWD analysis to a 1984 paper on large simple trials (Yusuf et al., 1984), which hinted at underlying issues in the big data environment. He said the authors wrote that the effects of treatments on major endpoints are easier to operationalize in major trials and Martin

observed that there is a corollary now when considering outcomes based on claims and electronic health record data. Schneeweiss also emphasized that when conducting RWD analyses it is important to consider both the “when” and the “how” questions: There is established knowledge on how to conduct such studies, but the question of when such analyses will likely produce valid findings will depend on the study question and the available data.

**Discussion of Potential Questions Around Treatment Assignment and Randomization**

On Day 2 of the workshop, individual participants discussed potential questions for consideration when contemplating the use of treatment assignment. The starting point for discussion was the central question from session 3: When can decision makers learn from real-world treatment assignment?

**BOX 3**

**When Can Decision Makers Learn from Real-World Treatment Assignment? Questions Raised by Individual Participants on Day 2 of the Workshop<sup>a</sup>**

|  |
|--|
| <p><b>Observational Studies</b></p> <ul style="list-style-type: none"> <li>• Random assignment is always preferable, but when is the cost—in time, money, infrastructure, and patient exposure—truly necessary? Several participants noted that candidate conditions that may benefit from random assignment could include assessing continuously updated data, determining an overall clinical strategy rather than assessing an individual treatment, or for trials that require fewer direct interactions between patients and providers.</li> <li>• Under what conditions could there be increased confidence in inference from non-randomized comparisons?</li> <li>• How could decision makers judge the validity of observational comparisons in advance, rather than waiting until after the results have been observed?</li> <li>• How can decision makers know the effects from unmeasured confounders are not so large that they would change a decision based on information from an observational study?</li> <li>• How can stakeholders interested in the use of RWE guard against the scenario in which researchers are able to redo their analysis in an observational study until they reach conclusions they would like to see? Could potential solutions involve the separation of research and analysis, replication by separate groups, and audit trail requirements?</li> <li>• Should stakeholders interested in RWE register all analyses or trust only results from pre-registered analyses? Should they include positive/negative controls? Should they do more analyses (“all by all” paradigm)?</li> </ul> |
| <p><b>Alternative Study Designs</b></p> <ul style="list-style-type: none"> <li>• Could randomized trials that “nudge” patients and providers toward a certain treatment, without necessarily assigning them to a specific treatment group, help improve patient compliance/adherence?</li> <li>• When can we have confidence in inference from cluster-randomized or stepped-wedge study designs?</li> </ul>   |

<sup>a</sup>DISCLAIMER: This document represents discussion by individual workshop participants of the Examining the Impact of Real-World Evidence on Medical Product Development—Workshop II: Practical Approaches. The statements made are those of the individual workshop participants and do not necessarily represent the views of all workshop participants; the planning committee; or the National Academies of Sciences, Engineering, and Medicine.

During the question-generating discussion on Day 2 of the workshop, several workshop participants considered the question of when there could be confidence in inference from non-randomized comparisons and discussed possible conditions, including the expectation of large effects, long study durations and the assessment of long-term outcomes, and other situations when RCTs are not practical. Other speakers pointed out additional conditions, such as when the reliability of the effect is dependent on clinical observation and the selection of historical comparators; when outcome is proximal to treatment; when there is a high degree of similarity between comparison groups; when the pathway from treatment to outcome is relatively clear (without complexity or reciprocal effects); and when treatment allocation methods are relatively transparent.

Simon mentioned prospective treatment assignment and unmeasured confounding in the absence of randomization as high-priority concerns. He also acknowledged that randomization is difficult to implement because it is expensive, time consuming, and often unpopular among both patients and providers. Building on a topic from Day 1 of the workshop, Temple and Mark McClellan of the Duke-Margolis Center for Health Policy discussed the potential use of “randomized nudge” trials as a way to improve patient adherence during studies.

## CLOSING REMARKS

McClellan concluded the workshop by thanking the participants and reminding them that this workshop, the second in a three-part series, is part of a larger activity examining the impact of RWE on medical product development. He noted that the third workshop, scheduled to take place on July 17–18, 2018, will move the conversation forward from practical approaches to applications for the collection and use of RWE. ♦♦

## REFERENCES

- Califf, R. M. 2018. Comparison of observational data and the ONTARGET results for telmisartan treatment of hypertension: Bull’s-eye or painting the target around the arrow? *JAMA Internal Medicine* 178(1):63–65.
- Connolly, S. J., M. D. Ezekowitz, S. Yusuf, J. Eikelboom, J. Oldgren, A. Parekh, J. Pogue, P. A. Reilly, E. Themeles, J. Varrone, S. Wang, M. Alings, D. Xavier, J. Zhu, R. Diaz, B. S. Lewis, H. Darius, H.-C. Diener, C. D. Joyner, and L. Wallentin. 2009. Dabigatran versus warfarin in patients with atrial fibrillation. *New England Journal of Medicine* 361(12):1139–1151.
- Fralick, M., A. S. Kesselheim, J. Avorn, and S. Schneeweiss. 2018. Use of health care databases to support supplemental indications of approved medications. *JAMA Internal Medicine* 178(1):55–63.
- Giugliano, R. P., C. T. Ruff, E. Braunwald, S. A. Murphy, S. D. Wiviott, J. L. Halperin, A. L. Waldo, M. D. Ezekowitz, J. I. Weitz, J. Špinar, W. Ruzyllo, M. Ruda, Y. Koretsune, J. Betcher, M. Shi, L. T. Grip, S. P. Patel, I. Patel, J. J. Hanyok, M. Mercuri, and E. M. Antman. 2013. Edoxaban versus warfarin in patients with atrial fibrillation. *New England Journal of Medicine* 369(22):2093–2104.
- Granger, C. B., J. H. Alexander, J. J. V. McMurray, R. D. Lopes, E. M. Hylek, M. Hanna, H. R. Al-Khalidi, J. Ansell, D. Atar, A. Avezum, M. C. Bahit, R. Diaz, J. D. Easton, J. A. Ezekowitz, G. Flaker, D. Garcia, M. Geraldes, B. J. Gersh, S. Golitsyn, S. Goto, A. G. Hermosillo, S. H. Hohnloser, J. Horowitz, P. Mohan, P. Jansky, B. S. Lewis, J. L. Lopez-Sendon, P. Pais, A. Parkhomenko, F. W. A. Verheugt, J. Zhu, and L. Wallentin. 2011. Apixaban versus warfarin in patients with atrial fibrillation. *New England Journal of Medicine* 365(11):981–992.
- Patel, M. R., K. W. Mahaffey, J. Garg, G. Pan, D. E. Singer, W. Hacke, G. Breithardt, J. L. Halperin, G. J. Hankey, J. P. Piccini, R. C. Becker, C. C. Nessel, J. F. Paolini, S. D. Berkowitz, K. A. A. Fox, and R. M. Califf. 2011. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *New England Journal of Medicine* 365(10):883–891.
- Rosenbaum, P. R., and D. B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1):41–55.
- Ruff, C. T., R. P. Giugliano, E. Braunwald, E. B. Hoffman, N. Deenadayalu, M. D. Ezekowitz, A. J. Camm, J. I. Weitz, B. S. Lewis, A. Parkhomenko, T. Yamashita, and E. M. Antman. 2014. Comparison of the efficacy and safety of new oral anticoagulants with warfarin in patients with atrial fibrillation: A meta-analysis of randomised trials. *The Lancet* 383(9921):955–962.
- Yusuf, S., R. Collins, and R. Peto. 1984. Why do we need some large, simple randomized trials? *Statistics in Medicine* 3(4):409–420.
- Yusuf, S., P. Sleight, C. Anderson, K. Teo, et al. for the ONTARGET Investigators. 2008. Telmisartan, ramipril, or both in patients at high risk for vascular events. *New England Journal of Medicine* 358(15):1547–1559.

**DISCLAIMER:** This Proceedings of a Workshop—in Brief was prepared by **Ben Kahn, Amanda Wagner Gee, and Carolyn Shore** as a factual summary of what occurred at the workshop. The statements made are those of the rapporteurs or individual workshop participants and do not necessarily represent the views of all workshop participants; the planning committee; or the National Academies of Sciences, Engineering, and Medicine.

**REVIEWERS:** To ensure that it meets institutional standards for quality and objectivity, this Proceedings of a Workshop—in Brief was reviewed by **Elise Berliner**, Agency for Healthcare Research and Quality; and **Hui Cao**, Novartis. **Lauren Shern**, National Academies of Sciences, Engineering, and Medicine, served as the review coordinator.

**SPONSOR:** This workshop was partially supported by the U.S. Food and Drug Administration.

For additional information regarding the workshop, visit <http://nationalacademies.org/hmd/Activities/Research/DrugForum/2018-MAR-06.aspx>.

Suggested citation: National Academies of Sciences, Engineering, and Medicine. 2018. *Examining the impact of real-world evidence on medical product development: II. Practical approaches: Proceedings of a workshop—in brief*. Washington, DC: The National Academies Press. doi: <https://doi.org/10.17226/25176>.

## Health and Medicine Division

*The National Academies of*

SCIENCES • ENGINEERING • MEDICINE

The nation turns to the National Academies of Sciences, Engineering, and Medicine for independent, objective advice on issues that affect people's lives worldwide.

[www.national-academies.org](http://www.national-academies.org)