Improving Health Research for Small Populations: A Workshop

January 18-19, 2018

Meeting Location
NAS Building
2101 Constitution Avenue NW
Washington, DC
Lecture Room

The National Academies of
SCIENCES · ENGINEERING · MEDICINE

Improving Health Research for Small Populations: A Workshop

Table of Contents

Tabs

- Kathi Mooney, *Using Technology for Recruitment, Retention, Data Collection, and Intervention Delivery*
- Tracy Onega, *Recruitment, Retention, and Collection of Data with a Focus on Small or Hard to Reach Populations*

H. Session 6 Presentations

- Rick H. Hoyle, *Design and Analysis Considerations in Research with Small Samples*
- Thomas A. Louis, *Bayesian Methods for Small Population Analysis*
- Katherine R. McLaughlin, *Estimating the Size of Hidden Populations*

A

TAB A

Improving Health Research for Small Populations: A Workshop

Materials in this tab:

Workshop Agenda

Workshop Participant List

Biographies for Workshop Steering Committee

Biographies for Workshop Presenters

Improving Health Research for Small Populations
A Workshop

January 18 and 19, 2018

The NAS Building
2101 Constitution Ave. NW
Washington, DC 20418

NAS Lecture Room

AGENDA

Brief Statement of Task

A National Academies of Sciences, Engineering, and Medicine ad hoc planning committee will organize a public workshop to discuss the methodological challenges of conducting research on small, underrepresented population subgroups in health research. The workshop will consider ways of addressing the challenges of conducting epidemiological studies or intervention research with small population groups, including alternative study designs, innovative methodologies for data collection, and innovative statistical techniques for analysis. Specifically, the workshop will address approaches for identification, recruitment, and retention of study participants to maximize the sample sizes of small groups in research studies; epidemiological design and analytics approaches for small samples; and intervention study design and analytic approaches for subpopulations.

This workshop is co-sponsored by the National Cancer Institute (NCI), the National Institute on Minority Health and Disparities (NIMHD), and the Office of Behavioral and Social Sciences Research (OBSSR) of the National Institutes of Health. Support for this workshop was also provided, in part, by the Robert Wood Johnson Foundation. The views expressed here do not necessarily reflect the views of the Foundation.

DAY 1: Thursday, January 18, 8:30 AM - 5:00 PM (NAS Lecture Room)

8:00 - 8:30 AM Registration (East Court); Refreshments Available

WELCOME AND INTRODUCTIONS

8:30 MODERATOR: **Graham Colditz**, *Committee Chair, Washington University in St. Louis, Opens the Workshop*

8:35 Welcome to the National Academy of Sciences
Brian Harris-Kojetin, *Director, Committee on National Statistics, National Academy of Sciences, Engineering, and Medicine*

8:45 Motivation and Objectives for the Workshop
Robert T. Croyle, *Director, Division of Cancer Control and Population Sciences, National Cancer Institute*

9:00 **SESSION I:** What do we Mean by Small Populations? How to Decide when a Small Population is Important or Meaningfully Different Enough to Study? Why did we Structure the Workshop this Way?

MODERATOR: **Graham Colditz**, *Committee Chair, Washington University in St. Louis*

9:05 **Howard Koh**, Harvard T.H. Chan School of Public Health, *The Importance of Health Research on Small Populations*
9:35 **Scarlett Lin Gomez**, University of California, San Francisco, *Data Issues in Studying Small Populations: Challenges, Opportunities, and a Case Study*
9:50 **Lisa Signorello**, Division of Cancer Prevention, National Cancer Institute, *Fielding Studies in Underrepresented Populations: Challenges and Considerations*
10:05 **Floor Discussion**

10:20 ***BREAK*** (refreshments available in East Court)

10:40 **SESSION 2**: Challenges in Using Available Data for Small Population Health Research

MODERATOR: **Lance Waller**, *Committee Member*, *Emory University*

10:45 **Kelly Devers**, NORC, *The Feasibility of Using Electronic Health Records and Electronic Health Data for Research on Small Populations*
11:05 **Chris Fowler**, Pennsylvania State University, *Using Geospatial Methods with Demographic Data to Identify Populations*
11:25 **Ellen Cromley**, Consultant, *Using Geospatial Methods with Other Health and Environmental Data to Identify Populations*

11:45 **Floor Discussion**

12:00 PM ***LUNCH*** (available in East Court)

1:00 **SESSION 3**: Techniques Used in Survey Research to Identify and Find Small Populations for Health Research

MODERATOR: **Graham Kalton**, *Committee Member*, *Westat*

1.05 **Marc Elliot**, Rand, *Probability Sampling Methods for Small Populations*
1:25 **Sunghee Lee**, University of Michigan, *Two Applications of Respondent Driven Sampling: Ethnic Minorities and Illicit Substance Users*
1:45 **Patrick Sullivan**, Emory University, *Venue-Based and On-line Sampling*
2:05 **Krista Gile**, University of Massachusetts, Amherst, Invited Discussant
2:25 **Floor Discussion**

2:45 ***BREAK*** (refreshments available in East Court)

3:00 **SESSION 4**: New and Emerging Designs for Intervention Studies

MODERATOR: **James Allen**, *Committee Member*, *University of Minnesota Medical School*

3:05 **Amy M. Kilbourne,** University of Michigan, *Designs for Dissemination and Implementation Research for Small Populations*
3:35 **Christine Lu**, Harvard Medical School, *Quasi-experimental Designs with Application to Small Populations*
3:55 **Diane Korngiebel,** University of Washington, *Addressing the Challenges of Research with Small Populations*
4:15 **Patrick H. Tolan,** University of Virginia, Invited Discussant
4:35 **Floor Discussion**

5:00 PM ***PLANNED ADJOURNMENT***

DAY 2: Friday, January 19, 8:30 AM - 2:00 PM (NAS Lecture Room)

(Refreshments available in East Court from 8:00 AM)

8:30 AM MODERATOR: **Graham Colditz**, *Committee Chair, Washington University in St. Louis*
Welcome and Introductions to Second Day

8:40 SESSION 5: Recruitment, Retention, and Collection of Data with a Focus on Small or Hard to Reach Populations
MODERATOR: **Jan Probst**, *Committee Member*, *University of South Carolina*

- 8:45 **Vetta Sanders-Thompson**, Washington University in St. Louis, *Issues and Challenges Associated with Recruitment and Retention for Health Research*
- 9:05 **F. Douglas Scutchfield,** University of Kentucky**,** *Improving Health Research in Rural Areas*
- 9:25 **Kathi Mooney**, University of Utah, *Using Technology for Recruitment, Retention. Data Collection, and Intervention Delivery*
- 9:45 **Tracy L. Onega,** Dartmouth University, Invited Discussant
- 10:05 **Floor Discussion**

10:20 ***BREAK*** (refreshments available in East Court)

10:40 **SESSION 6**: Analysis Techniques for Small Population Research
MODERATOR: **Lance Waller**, *Committee Member*, *Emory University*

- 10:45 **Rick H. Hoyle**, Duke University, *Design and Analysis Considerations in Research with Small Samples*
- 11:05 **Thomas A. Louis**, Johns Hopkins Bloomberg School of Public Health, *Bayesian Methods for Small Population Analysis*
- 11:25 **Katherine R. McLaughlin**, Oregon State University, *Estimating the Size of Hidden Populations*
- 11:45 **Floor Discussion**

12:00 PM ***LUNCH*** (available in East Court; seating available in Great Hall)

1:00 **SESSION 7**: Wrap-up and Concluding Remarks
MODERATOR: **Gordon Willis**, *National Cancer Institute*

- 1:00 **Steering Committee**. Panel Discussion – Next Steps in Improving Health Research for Small Populations
- 1:30 **Floor Discussion**

2:00 PM ***ADJOURNMENT***

Participant List

STEERING COMMITTEE MEMBERS

Graham Colditz (Chair), *Niess-Gain Professor of Surgery, Washington University School of Medicine*

James Allen, *Professor, Department of Biobehavioral Health and Population Sciences, University of Minnesota Medical School*

G. Graham Kalton, *Chairman and Senior Vice President, Westat, Inc.*

Janice Probst, *Professor, Department of Health Services Policy and Management, University of South Carolina, Arnold School of Public Health*

Lance Waller, *Rollins Professor and Chair, Department of Biostatistics and Bioinformatics, Emory University, Rollins School of Public Health*

SPONSOR

David Berrigan, *Program Director, Behavioral Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, NIH*

Robert Croyle, *Director, Division of Cancer Control and Population Sciences, National Cancer Institute, NIH*

Richard Moser, *Training Director, Division of Cancer Control and Population Sciences, National Cancer Institute, NIH*

Alonzo Plough, *Vice-President, Research Evaluation Learning and Chief Science Officer, Robert Wood Johnson Foundation*

Shobha Srivivasan, *Health Disparities Research Coordinator, National Cancer Institute, NIH*

Gordon Willis, *Program Director, Behavioral Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, NIH*

PRESENTERS

Ellen Cromley, *Consultant, Washington State*

Kelly Devers, *Senior Fellow, Health Care, NORC at the University of Chicago*

Marc Elliott, *Senior Principal Researcher, RAND Corporation*

Chris Fowler, *Assistant Professor of Geography and Demography, Penn State University*

Krista Gile, *Associate Professor, Department of Mathematics and Statistics, University of Massachusetts*

Scarlett Lin Gomez, *Adjunct Professor, UCSF School of Medicine*

Rick Hoyle, *Professor, Department of Psychology and Neuroscience, Duke University*

Amy Kilbourne, *Professor, Institute for Healthcare Policy and Innovation, University of Michigan*

Howard Koh, *Harvey V. Fineberg Professor of the Practice of Public Health Leadership, Harvard T.H. Chan School of Public Health*

Diane Korngiebel, *Assistant Professor, Department of Biomedical Informatics and Medical Education, University of Washington*

Sunghee Lee, *Associate Research Scientist, ISR, University of Michigan*

Tom Louis, *Professor, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health*

Christine Lu, *Associate Professor of Population Medicine, Harvard Medical School*

Katherine McLaughlin, *Assistant Professor of Statistics, Oregon State University*

Kathi Mooney, *Distinguished Professor, University of Utah College of Nursing*

Tracy Onega, *Associate Professor, Department of Biomedical Data Science, and of Epidemiology, and The Dartmouth Institute for Health Policy and Clinical Practice*

F. Douglas Scutchfield, *University of Kentucky College of Public Health*

Lisa Signorello, *Senior Biomedical Scientist, National Cancer Institute, NIH*

Patrick Sullivan, *Rollins School of Public Health, Emory University*

Vetta Sanders Thompson, *Professor, Brown School, Washington University in St. Louis*

Patrick H. Tolan, *Director Emeritus, Youth-Nex Center, The UVA Center to Promote Effective Youth Development*

GUESTS

Alaba Oluwatoyin Adedoyin, *Industrial Program Officer, Obafemi Awolowo University, Nigeria*

Brenda Adjei, *Program Director, National Cancer Institute (joining via Webinar)*

Jennifer Alvidrez, *Program Official, Division of Scientific Programs, National Institute on Minority Health and Health Disparities, NIH (joining via Webinar)*

Sean Arayasirikul, *Assistant Professor, Division of Developmental Medicine, Department of Pediatrics, School of Medicine, University of California, San Francisco (joining via Webinar)*

Jessica Athens, *Assistant Professor, Department of Population Health, New York University School of Medicine*

Kellan Baker, *Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health*

Christopher Barnhart, *AAAS Science and Technology Policy Fellow, Health Scientist, National Institutes of Health (NIH) Office of the Director (OD) (joining via Webinar)*

Annie Beach, *Research Fellow, Division of Cancer Control and Population Sciences, National Cancer Institute, NIH*

Inna Belfer, *Health Scientist Administrator/Project Officer, Office of Research on Women's Health, DHHS/NIH/OD/DPCPSI (joining via Webinar)*

Kara Bensley, *Postdoctoral Research Fellow | University of California, Berkeley Associate Scientist | Alcohol Research Group (joining via Webinar)*

Lew Berman, *Vice President, Social and Analytic Solutions, Maryland*

Sarah Bruce Bernal, *Implementation Science Team, Division of Cancer Control and Population Sciences, National Cancer Institute, NIH*

Obasanjo Bolarinwa, *Post-graduate, Obafemi Awolowo University, Nigeria (joining via Webinar)*

Abee Boyles, *Population Health Branch, Division of Extramural Research & Training, National Institute of Environmental Health Sciences, NIH (joining via Webinar)*

Cheryl Boyce, *Chief, Implementation Science Branch, Center for Translation Research and Implementation Science (CTRIS), National Heart, Lung, and Blood Institute (NHLBI), National Institutes of Health (NIH), Department of Health and Human Services (DHHS) (joining via Webinar)*

Kathryn Brignole, *Cancer Research Training Award Fellow, Genomic Epidemiology Branch, Epidemiology and Genomics Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute*

Victoria Cargill, *Medical Officer, Associate Director, Interdisciplinary Research, Office of Research on Women's Health, National Institutes of Health (joining via Webinar)*

Dorothy Castille, *Health Scientist Administrator, National Institute on Minority Health and Health Disparities, National Institutes of Health*

Liza Catucci, *Administrative Officer, Health Services Research and Development, U.S. Dept. of Veterans Affairs*

Christine Chang, *Scientific Program Coordinator, National Human Genome Research Institute, NIH (joining via Webinar)*

Mandi Pratt Chapman, *Associate Center Director, Patient-Centered Initiatives and Health Equity, GW Cancer Center*

Anita Cheung, *Johns Hopkins University (joining via Webinar)*

Dietta Chihade, *Mirzayan Fellow, The National Academies of Sciences, Engineering, and Medicine*

Juanita Chinn, *Program Director, Population Dynamics Branch, Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD)*

Sarah Conderino, *Surveillance Data Scientist, Department of Population Health, NYU School of Medicine*

Allison Cook, *Program Specialist, National Cancer Institute, NIH*

Danielle Daee, *Program Director, Division of Cancer Control and Population Sciences, National Cancer Institute, NIH (joining via Webinar)*

Jyoti Dayal, *Program Officer, National Human Genome Research Institute (NHGRI), NIH (joining via Webinar)*

Rina Das, *Program Director, Division of Extramural Scientific Programs, National Institute on Minority Health and Health Disparities, National Institutes of Health*

Roberto Delgado, *National Institute of Mental Health, NIH*

Christina N. Dragon, *Statistician, Office of Minority Health, CMS*

Kevin English, *Director, Albuquerque Area Southwest Tribal Epidemiology Center (AASTEC), Albuquerque Area Indian Health Board, Inc. (AAIHB), New Mexico (joining via Webinar)*

Sarah Fagan, *Fellow, Division of Cancer Control and Population Sciences, National Cancer Institute, NIH*

Angela Falisi, *Program Specialist, Division of Cancer Control and Population Sciences, National Cancer Institute, NIH (joining via Webinar)*

Margaret Farrell, *Public Health Advisor, Division of Cancer Control and Population Sciences, National Cancer Institute, NIH*

Jessica Fish, *Postdoctoral Fellow, Population Research Center, University of Texas at Austin (joining via Webinar)*

Eleanor Fleming, *Dental Public Health Resident, Boston University*

Gonçalo Forjaz de Lacerda, *Guest Researcher, Data Analytics Branch, National Cancer Institute*

Lisa Gallicchio, *Program Director, Division of Cancer Control and Population Sciences, National Cancer Institute, NIH*

Silvia Garvey, *Scientific Program Specialist, Division of Genomic Medicine, National Human Genome Research Institute, NIH (joining via Webinar)*

Elizabeth Gillanders, *Branch Chief, Division of Cancer Control and Population Sciences, National Cancer Institute, NIH (joining via Webinar)*

Kimberly Ann Gray, *Program Director Children's Environmental Health, Population Health Branch, Division of Extramural Research and Training, National Institute of Environmental Health Sciences, NIH*

Michaela Gross, *Research Assistant, University of South Florida (joining via Webinar)*

Paul Guerino, *Statistician and Poverty Data Lead, Office of Minority Health, CMS (joining via Webinar)*

Elina Guralnik, *Health Consultant, George Mason University*

Craig Hales, *Division of Health and Nutrition Examination Surveys, National Center for Health Statistics, Centers for Disease Control and Prevention*

Emily Haozous*, Associate Professor, Regent's Professor, UNM College of Nursing (joining via Webinar)*

Jody L. Herman, *Scholar of Public Policy, Williams Institute, UCLA School of Law (joining via Webinar)*

Kelly Hirko, *Assistant Professor, Department of Epidemiology & Biostatistics, College of Human Medicine, Michigan State University (joining via Webinar)*

Kristen Hudgins, *Social Science Analyst, Office of Performance and Evaluation, U.S. Department of Health and Human Services*

Genevieve Hugenbruch, *Graduate Research Assistant, Maryland Department of Health (joining via Webinar)*

Chandra Jackson, *Earl Stadtman Investigator, Epidemiology Branch, Social and Environmental Determinants of Health Equity, National Institute of Environmental Health Sciences, National Institutes of Health (joining via Webinar)*

Aimee James, *Associate Professor and T32 Training Director, Division of Public Health Sciences –Department of Surgery, Washington University in St Louis School of Medicine*

Anuradha Jetty, *Research Associate, Robert Graham Center for Policy Studies in Family Medicine and Primary Care, American Academy of Family Physicians*

Bonnie Joubert, *Program Director, Population Health Branch, Division of Extramural Research and Training, National Institute of Environmental Health Sciences, NIH (joining via Webinar)*

Edmund Keane, *Health Science Policy Analyst, Tribal Health Research Office, NIH*

Amy Kennedy, *Center for Research Strategy, National Cancer Institute, NIH*

Min Kim, *Senior Program Specialist, Office of Minority Health and Health Disparities, Maryland Department of Health (joining via Webinar)*

Imke Kirste, *Study Manager, NIEHS Clinical Research Unit, NIH (joining via Webinar)*

Lisa Klesges, *Senior Advisor, Implementation Science and Public Health (Consultant), Division of Cancer Control and Population Sciences, National Cancer Institute*

Sarah Kobrin, *Chief (acting), HSIRB, National Cancer Institute*

Julie Kranick, *Center for the Study of Asian American Health, NYU School of Medicine (joining via Webinar)*

Poorna Kushalnagar, *Director, Deaf Health Communication and Quality of Life Research Lab, Gallaudet University (joining via Webinar)*

Bryan Kutner, *Postdoctoral Research Fellow, HIV Center for Clinical and Behavioral Studies at the NY State Psychiatric Institute and Columbia University (joining via Webinar)*

Simona Kwon, *Assistant Professor, NYU School of Medicine, (joining via Webinar)*

Karen Lee, *Medical Officer, Director, Behavioral Pediatrics and Health Promotion Program, Child Development and Behavior Branch, National Institute of Child Health and Human Development, National Institutes of Health (joining via Webinar)*

Eugene Lengerich, *Professor of Public Health Sciences, The Pennsylvania State University*

Alicia LePard, *Acute Care Nurse Practitioner, High Desert Healthcare, LLC, Wyoming (joining via Webinar)*

Benmei Liu, *Mathematical Statistician and Program Director, Division of Cancer Control and Population Sciences, National Cancer Institute, NIH*

Ryan Mahon, *Program Analyst (Contractor), Division of Program Coordination, Planning, and Strategic Initiatives, Office of the Director, National Institutes of Health (joining via Webinar)*

Andrea Horvath Marques, *Chief, Office for Research on Disparities and Global Mental Health, National Institute of Mental Health, NIH*

Heather Menne, *Social Science Analyst, Office of Performance and Evaluation, Center for Policy and Evaluation, Administration for Community Living, U.S. Department of Health & Human Services (joining via Webinar)*

Nancy Miller, *Senior Advisor to the Associate Director, Division of Cancer Control and Population Sciences, National Cancer Institute, NIH (joining via Webinar)*

Shabab Ahmed Mirza, *Research Assistant, LGBT Progress, Center for American Progress (joining via Webinar)*

Florence Momplaisir, *Assistant Professor, Drexel University College of Medicine (joining via Webinar)*

Jennifer Moss, *Cancer Research Fellow, Division of Cancer Control and Population Sciences, National Cancer Institute, NIH (joining via Webinar)*

Arnab Mukherjea, *Assistant Professor of Health Sciences, California State University, East Bay*

Hervette Nkwihoreze, *Research Assistant, Drexel University College of Medicine (joining via Webinar)*

Angele Nyinawingabo, *Research Assistant, Drexel University College of Medicine (joining via Webinar)*

Ayokunle Omoniyi, *Program Officer, Huklyn Global Fund, Nigeria*

Chris Obermeyer, *Fulbright Public Policy Fellow, Ministry of Health of Ukraine (joining via Webinar)*

Karen Parker, *Director, Sexual & Gender Minority Research Office, Division of Program Coordination, Planning, and Strategic Initiatives, Office of the Director, NIH (joining via Webinar)*

Dolly Penn, *Medical Officer, Healthcare Assessment Research Branch, Healthcare Delivery Research Program, Division of Cancer Control & Population Sciences, National Cancer Institute (joining via Webinar)*

Antoinette Percy-Laurry, *Health Scientist, Implementation Science, Division of Cancer Control and Population Sciences, National Cancer Institute (joining via Webinar)*

Mackenzi Pergolotti, *Assistant Professor, Department of Occupational Therapy, Colorado State University (joining via Webinar)*

Courtney Pinard, *Senior Research Scientist, Center for Nutrition, Nebraska (joining via Webinar)*

Stephanie Jilcott Pitts*, Associate Professor, Department of Public Health, East Carolina University*

Mona Puggal, *Division of Cardiovascular Sciences, NHLBI/NIH (joining via Webinar)*

Gita Rampersad, *Strategic Engagement Consultant, Healthcare Ready, Washington, DC (joining via Webinar)*

Jennifer Rankin, *Senior Manager, Research and Product Services, American Academy of Family Physicians*

Sarah E. Raskin, *Assistant Professor, iCubed Initiative, Oral Health Core, L. Douglas Wilder School of Government and Public Affairs, Virginia Commonwealth University*

Nicole Redmond, *Medical Officer, Division of Cardiovascular Sciences, NIH*

Mary Roary, *Program Director, Division of Extramural Science Programs, National Institute of Nursing Research, National Institute of Health (joining via Webinar)*

Carol Robbins, *Public Health Analyst, Office of Planning, Analysis and Evaluation Health Resources and Services Administration*

Megan Roberts, *Division of Cancer Control and Population Sciences, NCI, NIH*

Eduardo Romano, *Senior Research Scientist, Pacific Institute for Research and Evaluation (PIRE), (joining via Webinar)*

Adelaida Rosario*, Health Specialist, National Institute on Minority Health and Health Disparities, National Institutes of Health*

Sonia Rosenfield, *Center for Research Strategy, National Institute on Alcohol, NIH*

Melissa Rotunno, *Program Director, NIH/NCI, Division of Cancer Control and Population Sciences (joining via Webinar)*

Stefanie Russell, *Associate Professor, Department of Epidemiology & Health Promotion, NYU College of Dentistry*

Zeina Saliba, *Assistant Professor, The George Washington University (joining via Webinar)*

Ramzi Salloum, *Assistant Professor, Health Outcomes and Policy, University of Florida*

Parth Saraiya, *MPH student, The George Washington University (joining via Webinar)*

Ayden Scheim, *Division of Infectious Diseases and Global Public Health, School of Medicine, University of California San Diego (joining via Webinar)*

Marcia Scott, *Division of Epidemiology and Prevention Research, National Institute on Alcohol*

Abuse and Alcoholism, NIH (joining via Webinar)

Katrina Serrano, *Health Scientist Administrator/Program Officer, Interdisciplinary Research, Office of Research on Women's Health (ORWH), Division of Program Coordination, Planning, and Strategic Initiatives, Office of the Director, NIH (joining via Webinar)*

Betsy Shenkman, *Associate Director, Population Sciences, Health Cancer Center, University of Florida*

Lisa Shook, *Assistant Professor of Pediatrics, Administrative Director, Cincinnati Comprehensive Sickle Cell Center (joining via Webinar)*

Monica Sierra, *Division of Cancer Prevention and Control, Centers for Disease Control and Prevention (joining via Webinar)*

Alena Smith, *Graduate Student (joining via Webinar)*

Michael Spittel, *National Institutes of Health (joining via Webinar)*

David Stinchcomb, *Westat, Inc., Maryland*

Denise Stredrick, *Health Scientist Administrator/Project Officer, Office of Research on Women's Health, DHHS/NIH/OD/DPCPSI (joining via Webinar)*

Eward Strickler, *School of Medicine, University of Virginia (joining via Webinar)*

Jariah Strozier, *Doctoral Student, Certified Health Education Specialist, Advisor of Black Organizations Council, Department of Sociology, Virginia Tech (joining via Webinar)*

L. Jospeh Su, *Department of Epidemiology, University of Arkansas for Medical Sciences (joining via Webinar)*

Bob Sun, *Fellow, NIEHS, NIH*

Megan Sutter, *Postdoctoral Research Fellow, Moffitt Cancer Center, Florida (joining via Webinar)*

Catherine Tallant, *Doctoral Candidate, Florida School of Professional Psychology at Argosy University (joining via Webinar)*

Kosuke Tamura, *Postdoctoral Fellow, National Heart, Lung, and Blood Institute, National Institutes of Health*

Caitlin Turner, *Data Manager, Center for Public Health Research (CPHR), San Francisco Department of Public Health (joining via Webinar)*

Kathleene Ulanday, *Pre-doctoral Student, Graduate Research Assistant, Department of Epidemiology, Mailman School of Public Health, Columbia University*

Robin Vanderpool, *Associate Professor, Department of Health, Behavior & Society, University of Kentucky College of Public Health (joining via Webinar)*

Cindy Veldhuis, *Ruth Kirschstein Postdoctoral Research Fellow, School of Nursing, Columbia University (joining via Webinar)*

Amy Veney, *Lecturer, Kent State University, College of Nursing*

Cynthia Vinson, *Senior Advisor for Implementation Science, Implementation Science Team, Division of Cancer Control and Population Sciences, National Cancer Institute (joining via Webinar)*

Ed Villalaz, *(joining via Webinar)*

Caroline Voyles, *Director, Student Placement & Partnership Development, Dornsife School of Public Health, Drexel University (joining via Webinar)*

Paul Wesson, *Postdoctoral Fellow, Center for AIDS Prevention Studies, University of California, San Francisco (joining via Webinar)*

Chris Wheldon, *Cancer Prevention Fellow, National Cancer Institute*

Jamie White, *Presidential Management Fellow – STEM (Special Assistant), Office of Research on Women's Health (ORWH), National Institutes of Health*

David Wilson, *Director, Tribal Health Research Office, NIH*

Andrea Wirtz, *Assistant Scientist, Johns Hopkins Bloomberg School of Public Health*

Diana Withrow, *Postdoctoral Fellow, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH*

Behailu Woldegiorgis, *Researcher, Ethiopia*

Laura Wyatt, *Research Data Manager, Section for Health Equity, Department of Population Health, NYU School of Medicine (joining via Webinar)*

Jessica Xavier, *Health Scientist, HRSA\HAB\OTCD\DEB*

Sheldwin Yazzie, *Exposure Scientist/Epidemiologist, Albuquerque Area Southwest Tribal Epidemiology Center (AASTEC), New Mexico*

Melanie Young, *Biological Scientist, U.S. Environmental Protection Agency*

Mandi Yu, *Surveillance Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute (joining via Webinar)*

Xingyou Zhang, *Senior Mathematical Statistician, Economic Research Service*

NATIONAL ACADEMIES STAFF

Hermann Habermann, *Senior Program Officer, Committee on National Statistics, The National Academies of Sciences, Engineering, and Medicine*

Brian Harris-Kojetin, *Director, Committee on National Statistics, The National Academies of Sciences, Engineering, and Medicine*

Nancy Kirkendall, *Study Director, Committee on National Statistics, The National Academies of Sciences, Engineering, and Medicine*

Anthony Mann, *Program Coordinator, Committee on National Statistics, The National Academies of Sciences, Engineering, and Medicine*

Sharyl Nass, *Board Director, Health and Medicine Division, The National Academies of Sciences, Engineering, and Medicine*

Jordyn White, *Program Officer, Committee on National Statistics, The National Academies of Sciences, Engineering, and Medicine*

Steering Committee Biographies

Chair

GRAHAM A. COLDITZ (NAM) is Neiss-Gain professor in the School of Medicine at Washington University, St. Louis (WUSTL), where he also serves as chief of the Division of Public Health Sciences in the Department of Surgery and program director of the master of population health sciences degree program. Dr. Colditz is the associate director of prevention and control at the Siteman Cancer Center at WUSTL and deputy director of the Institute for Public Health at WUSTL, where he holds leadership roles in education and fostering transdisciplinary research to address local and global public health challenges. His longstanding research focuses on causes and prevention of chronic diseases, and strategies to implement prevention based on what we already know.

Members

JAMES ALLEN is professor in the Department of Family Medicine and Biobehavioral Health, University of Minnesota Medical School, Duluth Campus. He is also a licensed psychologist with interests in individual and family therapy. His current work focuses on multi-level community intervention promotion for American Indian/Alaska Native, and rural community health. Dr. Allen received a Fulbright lecturer/research fellowship in 2003 and the Martin Mayman award for distinguished contribution to personality assessment. He currently teaches medical students in the areas of cultural competency training, population health, integration with behavioral health, and rural health.

G. GRAHAM KALTON is senior vice president and chairman of the board at Westat and a research professor in the Joint Program of Survey Methodology (JPSM) at the University of Maryland. Prior to working at Westat, he was a research scientist in the Survey Research Center of the University of Michigan, where he also served a term as chairman of the Department of Biostatistics. Before that, he was on the faculty of the University of Southampton and the London School of Economics and Political Science. Dr. Kalton is a recognized leader in survey methodology and statistical sampling.

JANICE C. PROBST is professor in the Department of Health Services Policy and Management and director of the South Carolina Rural Health Research Center at the Arnold School of Public Health of the University of South Carolina. In 2000, she contributed to the establishment of the South Carolina Rural Health Research Center, which represents a multi-year effort examining health disparities among poor and minority rural populations. Dr. Probst has extensive experience in health services research with an emphasis on rural and vulnerable populations.

LANCE A. WALLER is Rollins professor and chair of the Department of Biostatistics and Bioinformatics in the Rollins School of Public Health at Emory University. His research involves the development and application of statistical methods for spatially referenced data including

applications in environmental justice, neurology, epidemiology, disease surveillance, conservation biology, and disease ecology. Dr. Waller is interested in both the statistical methodology and the environmental and epidemiologic models involved in the analysis of this type of data

Presenter Biographies

WELCOME AND INTRODUCTIONS

BRIAN HARRIS-KOJETIN is director of the Committee on National Statistics. He comes from OMB where he served as senior statistician in the Statistical and Science Policy Office. He chaired the Federal Committee on Statistical Methodology and was the lead at OMB on issues related to standards for statistical surveys, survey nonresponse, survey respondent incentives, and the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA). Prior to joining OMB in 2001, he was the senior project leader of Research Standards and Practices at the Arbitron Company. He also previously served as a research psychologist in the Office of Survey Methods Research in the Bureau of Labor Statistics.

ROBERT T. CROYLE, PHD, is director of the Division of Cancer Control and Population Sciences at the National Cancer Institute (NCI). The division covers a wide range of scientific domains and disciplines, including cancer epidemiology, behavioral science, surveillance, survivorship, and health services research. Before coming to NCI, he was professor of psychology and a member of the Huntsman Cancer Institute at the University of Utah in Salt Lake City. His research has examined how individuals process, evaluate, and respond to cancer risk information.

SESSION 1: What do we Mean by Small Populations? How to Decide when a Small Population is Important or Meaningfully Different Enough to Study? Why did we Structure the Workshop this Way?

HOWARD K. KOH is the Harvey V. Fineberg professor of the practice of public health leadership at the Harvard School of Public Health and the Harvard Kennedy School. He has previously served as the 14th Assistant Secretary for Health for the U.S. Department of Health and Human Services (2009-2014) after being nominated by President Barack Obama, and as Commissioner of Public Health for the Commonwealth of Massachusetts (1997-2003) after being appointed by Governor William Weld.

SCARLETT LIN GOMEZ is professor in the Department of Epidemiology and Biostatistics and a member of the Helen Diller Family Comprehensive Cancer Center, at the University of California, San Francisco. She is also a research scientist at the Cancer Prevention Institute of California, where she is the director of the Greater Bay Area Cancer Registry, a participant in the NCI SEER (Surveillance, Epidemiology, End Results) program and the California Cancer Registry. Her research focuses primarily on cancer health disparities and aims to understand the multilevel drivers of those disparities.

LISA SIGNORELLO is the acting director and acting chief of the Cancer Prevention Fellowship Program (CPFP) Branch in the National Cancer Institute's Division of Cancer Prevention. Dr. Signorello came to the NCI after having held academic positions at the Harvard School of Public

Health, Harvard Medical School, and Vanderbilt University, as well as having had significant private sector research experience. Her research is broadly focused on issues related to the macro- and individual-level factors that give rise to socioeconomic and racial disparities in cancer incidence and survival.

SESSION 2: Challenges in Using Available Data for Small Population Health Research

KELLY J. DEAVERS is a senior fellow in the Health Care department at NORC at the University of Chicago. She is a widely recognized expert in health services and policy research with particular expertise in alternative payment models, delivery system reforms, and their impacts on access, cost, and quality. She has conducted evaluations of bundled and episode-based payment initiatives and published on topics such as accountable care organizations, medical homes, health information technology, care coordination, and quality improvement. Devers is also a nationally recognized expert in mixed methods and qualitative research and evaluation, including a new method called qualitative comparative analysis (QCA).

CHRIS FOWLER is assistant professor of geography at Penn State University. His research interests are in urban and economic geography, demographics, poverty, planning and economic development policies, spatial statistics, and complex economics systems. His current work focuses on methods for representing neighborhood change in complex, multiscalar contexts and developing a line of research that explores the increasing neighborhood-scale diversity in U.S. cities. He is particularly interested in blending complex, interview-based research on neighborhoods with innovative quantitative methods of spatial analysis.

ELLEN CROMLEY was professor of geography at the University of Connecticut, adjunct professor in the Department of Community Medicine and Health Care in the University's School of Medicine, and guest professor in the Department of Occupational and Environmental Medicine at Lund University in Sweden. Her research interests include geographical patterns of health and disease, location of health services and geographical factors affecting their utilization, and mapping and spatial analysis of health data.

SESSION 3: Techniques Used in Survey Research to Identify and Find Small Populations for Health Research

MARC ELLIOTT is a senior principal researcher at RAND and holds its distinguished chair in Statistics. His areas of interest include health disparities, Medicare, vulnerable populations, healthcare experiences, profiling of health care institutions, survey sampling, experimental design, casual inference, and case-mix adjustment in U.S. and U.K. applications. He has developed Bayesian methods of estimating race/ethnicity and associated disparities using surname and address information.

SUNGHEE LEE is an associate research scientist at the Survey Methodology Program. Before joining the Survey Methodology Program, she served as Survey Methodologist for California Health Interview Survey and an Adjunct Assistant Professor in Biostatistics at UCLA. Her

research interest includes sampling and measurement issues in data collection with linguistic and racial minorities as well as hard-to-reach populations and cross-cultural survey methodology.

PATRICK SULLIVAN is a professor in the Department of Epidemiology, Rollins School of Public Health, Emory University. He has worked on HIV testing programs with migrant farm workers, with the inclusion of Hispanic participants in online sexual health surveys, and has investigated methods to increase participation of African American and Latino MSM in his research. He has also worked with MSM prevention and vaccine studies in Peru and Brazil.

KRISTA GILE is associate professor of mathematics and statistics at the University of Massachusetts. Her research focuses on developing statistical methodology for social and behavioral science research, particularly related to making inference from partially-observed social network structures. Most of her current work is focused on understanding the strengths and limitations of data sampled with link-tracing designs such as snowball sampling, contact tracing, and respondent-driven sampling.

SESSION 4: New and Emerging Designs for Intervention Studies

AMY KILBOURNE is professor of psychiatry at the University of Michigan (UM) Medical School and director of the VA Quality Enhancement Research Initiative (QUERI). Dr. Kilbourne's goal is to improve outcomes for persons with mental disorders through research that accelerates the implementation of effective practices into real-world settings. She is a national expert in implementation science, mental health services, and academic-community research partnerships.

CHRISTINE LU is an associate professor in the Department of Population Medicine at Harvard Medical School and Harvard Pilgrim Health Care Institute and she co-directs the PRecisiOn Medicine Translational Research (PROMoTeR) Center. She is a pharmacist, health policy scientist and pharmacoepidemiologist. Her program of research focuses on the policy, legal, ethical, economic and societal issues of precision medicine. She has contributed substantially to evaluations of health policies in developing and developed countries using large, longitudinal administrative healthcare data and rigorous quasi-experimental research methods.

DIANE KORNGIEBEL is an assistant professor in biomedical health informatics and an adjunct assistant professor in bioethics and humanities at the University of Washington, School of Medicine, in Seattle, where she works at the intersection of bioethics, informatics, and the delivery of healthcare innovation. She brings her expertise in mixed methods research, bioethics, and user-centered design to developing innovative, people-informed interventions—using collaborative approaches—that improve health while addressing issues of accessibility, acceptability, inclusivity, and equity.

PATRICK H. TOLAN is the Charles S. Robb Professor of Education at the University of Virginia in the Curry School of Education and in the Department of Psychiatry and Neurobehavioral

Sciences in the School of Medicine. Over the past 35 years he has organized and lead multiple longitudinal and randomized control studies focused on prevention of problem-behavior and academic and social failure and promoting resilience and effective functioning among youth in high risk communities and/or at critical developmental transitions. His studies have provided insights and innovation in how multiple systems of influence converge to affect developmental course and provide opportunities for promoting positive outcome, understanding of various forms of violence and their interrelation, how families can manage stress, what schools can do to improve child social and emotional development, and how robust scientific methods can be integrated into community based efforts and collaborations.

SESSION 5: Recruitment, Retention, and Collection of Data with a Focus on Small or Hard to Reach Populations

VETTA SANDERS-THOMPSON is professor at Washington University in St. Louis in the George Warren Brown School of Social Work and the Public Health Program. Dr. Thompson's research focuses on racial and ethnic disparities in health and well-being, particularly among African Americans. Cultural competence and community engagement practices inform her research efforts.

F. DOUGLAS SCUTCHFIELD is the initial incumbent in the Peter P. Bosomworth Professorship in Health Services Research and Policy at the University of Kentucky. He was born and raised in Appalachia Kentucky. He holds the MD degree from the University of Kentucky and completed residency training at UK and The Centers for Disease Prevention and Control. Prior to his academic career he practiced in rural Appalachia. He holds fellowships in both the American College of Preventive Medicine and the American Academy of Family Practice. He was one of the founders of the College of Community Health Science at the University of Alabama and founded the Graduate School of Public Health at San Diego State University; he founded the school, now college, of public health at the University of Kentucky. His current research focuses on community health, public health organization and delivery, quality of care issues and democracy in health care decision making.

KATHI MOONEY is a distinguished professor and holds an endowed chair in the College of Nursing at the University of Utah. She is the co-leader of the Cancer Control and Population Sciences Program at the Huntsman Cancer Institute. Her program of research is focused on patient-reported outcomes, the improvement of cancer symptom outcomes and cancer family caregiver research. She has demonstrated the efficacy of an automated telehealth system in improving both patient and family caregiver outcomes.

TRACY L. ONEGA is associate professor of epidemiology in the Geisel School of Medicine Dartmouth University. Her major interests in cancer control center on: access to cancer care, including screening, treatment, and surveillance; how where care is received influences treatment and outcomes; and how early intervention affects patients' health and health care experiences. She has a special interest in how cancer care resources are allocated across populations, and how variations thereof impact cancer patients. Dr. Onega's research program

is largely built around her expertise in using registry and claims data to address these lines of inquiry.

SESSION 6: Analysis Techniques for Small Population Research

RICK H. HOYLE is professor of psychology and neuroscience and director of undergraduate studies at Duke University. His research interests include the foundations of self-esteem, which includes an interest in the interplay of self-evaluations across different domains (e.g., appearance, social life), processes of self-attention and self-regulation as they are implicated in the maintenance of self-esteem, and the influence of social acceptance and rejection on self-esteem processes. He is also interested in the role of personality in problem behavior, with particular interest in how prevention interventions can be designed to capitalize on the link between certain personality dimensions (e.g., sensation seeking) and problem behaviors (e.g., use of illicit drugs, sexual risk taking). Finally, he is interested in strategic applications of structural equation modeling and related techniques for the purpose of modeling complex processes that unfold over time, with a particular focus on measurement and design issues relevant for models that include mediated and moderated effects.

THOMAS A. LOUIS is emeritus professor of biostatistics at the Johns Hopkins Bloomberg School of Public Health. His research interests include: Bayesian methods; clinical and field studies; health services research, environmental risk assessment, genomics, and survey methods. He is an elected member of the International Statistical Institute, a fellow of the American Statistical Association, the American Association for the Advancement of Science, and the Institute of Mathematical Statistics; also a National Associate of the National Research Council.

KATHERINE K. McLAUGHLIN is an assistant professor in the Department of Statistics at Oregon State University. Her recent research was a rational-choice preferential recruitment model for respondent-driven sampling. This work involved collaborations with members of the Hard-to-Reach Population Methods Research Group and the World Health Organization to develop new statistical methodology geared toward improved estimation for hidden populations, including those at high risk for HIV/AIDS. Her research interests include survey sampling methodology, social network analysis, network sampling, and social science applications of statistics.

B

TAB B

Improving Health Research for Small Populations: A Workshop

Materials in this tab:

"What is a Small Population?" January 11, 2018.
A background document prepared for the workshop by staff of the Committee on National Statistics, National Academy of Sciences, Engineering and Medicine with substantial input from sponsors, steering committee members and presenters.

"Small is Essential: Importance of Subpopulation Research in Cancer Control." Editorial, *American Journal of Public Health*, April 23, 2015. By Srinivasan, S., Moser, R. Willis, G., Riley, W., Alexander, M., Berrigan, D., and Kobrin, S.

January 11, 2018

What is a Small Population?[1]

Srinivasan and colleagues[2] (2015, p.1) provided their definition of a small population as one for which "the size, dispersion, or accessibility of the population of interest makes it difficult to obtain adequate sample sizes to test specific research questions." They go on to note that "it is critical to ensure that all segments of the US population benefit from [health] research and from the latest technologic advances…"

Unfortunately, populations for which it is difficult to obtain adequate sample size are also likely to be expensive to study because dispersion and accessibility increase logistical costs. Hence, even if it is technically feasible to study a small population, it may not be easy to obtain funding for the study. This argues for increased efforts to document the needs, potential benefits, and methods for enhancing the efficiency of the study.

In other situations there may be no good sample frames because there is no agreed to definition of the population or a way to identify it. These, so called "hidden populations" are "small" by the above definition but raise more fundamental questions in health research and require additional data source work even to measure health disparities. Devers et al. (2013) provides some examples of this challenge and potential options for addressing it.

Much of the interest in studying health disparities for "small" populations was stimulated by the Department of Health and Human Service's Healthy People project[3] in 2010. The project, which set a vision and strategy for improved health outcomes by 2020, listed as one of their goals a desire to "achieve health equity, eliminate disparities, and improve the health of all groups." The elimination of these disparities and inequity is to be assessed across the following categories: race/ethnicity; gender; socioeconomic status; disability status; lesbian, gay bisexual, and transgender status; and geography. Several of these categories—in particular, some races, LGBT status, and some geographies—can be indicative of small populations.

The National Cancer Institute (NCI)[4] and Health Resources and Services Administration (HRSA)[5] both provide a wealth of information about measuring diversity of health outcomes and about measuring diversity in access to health care. Diversity is usually established by comparisons. A diversity measure of a small population may be compared to that of the U.S. population, or comparisons may be made among the diversity measures for its subpopulations. For example a small population of interest might be people whose work exposes them to a particular hazard,

[1]This is a living document prepared as background for the workshop by staff of the Committee on National Statistics, National Academy of Sciences, Engineering and Medicine with substantial input from sponsors, steering committee members and presenters.

[2]Authors are co-sponsors of this workshop.

[3]See, https://www.healthypeople.gov/sites/default/files/HP2020_brochure_with_LHI_508_FNL.pdf (December 2017).

[4]See, https://www.cancer.gov/about-nci/organization/crchd/about-health-disparities.

[5]See, https://bhw.hrsa.gov/shortage-designation/muap.

such as miners in a particular area or type of mine who experience poor health outcomes. Small populations can also occur from combinations of characteristics, e.g., members of American Indian and Alaska Native tribal groups who live in small distinct communities; immigrants who are "undocumented" by country of origin; individuals at risk for HIV by category (men who have sex with men, sex workers, illicit drug users.) As such, the range of possibilities is large and contingent on a researcher's specification of a research question and other covariates for control or study.

Small populations and the inference challenges associated with small sample size are inevitable due to variation in incidence of disease, prevalence of health-related behaviors, and heterogeneity in population characteristics. Promoting and strengthening research with small populations is of particular importance because substantial health disparities may arise from the combination of disparities in many small and distinct demographic groups. Lack of evidence concerning etiology of outcomes and most effective treatments for such groups may perpetuate disparities.

Different Kinds of Health Research

Commonly used approaches in public health -- surveillance/epidemiological studies and intervention studies – involve different types of inferences. Public health *surveillance* and *epidemiological* studies are generally accomplished through observational studies of the health status and health needs of population groups, either using existing data sources or designing surveys to collect needed information from a target population. These studies are strictly observational, with no attempt by the researcher to affect the outcome. On the other hand, intervention studies examine the effect of a treatment, behavior modification or treatment delivery option on an outcome. Examining the effects of an *intervention* requires a carefully designed study that may be referred to as an intervention, prevention, behavioral study, implementation study or clinical trial.

Surveillance/Epidemiological Study

The goal of observational studies for *surveillance* is typically descriptive: to estimate the percentage of some target population or subpopulation within a geography, or the percentage of that population with health disparities or certain health outcomes. *Epidemiological* observational studies are analytical, seeking to estimate associations, risk factors, odds ratios, or relative risks. They may include cohort studies, case control studies, or cross-sectional analysis. Typically, available survey data, administrative records, registries, electronic health records, and other data may be used for surveillance studies or as a guide to selecting qualified individuals to survey/enroll in epidemiological studies.

For small populations, these studies may be challenged to find available data. The question of finding people for rare population research has been well addressed by the survey research community. In this type of survey research, "small" might depend on how the population is perceived in relation to either a larger group, such as the rest of the US population or

partitioned into subgroups by features such as race/ethnicity, geography or socioeconomic status. In his introductory chapter to the monograph on hard-to-survey populations, Tourangeau noted that "problems [in sampling] arise when a target population represents a small fraction of the frame population." Tourangeau et al. (2014, p. 4).

Recently there has been considerable interest in conducting surveillance on so-called "hidden" populations—those that are not easily identifiable from administrative records and household/self-report surveys because individuals therein are reluctant to self-identify. In statistical terms, for these populations there is no sampling frame. Examples include the homeless, migrant workers, immigrants and various gender preference minorities. While survey research traditionally relies on probability sampling from a frame to make sure estimates that are derived are unbiased and generalizable to the target population, that method cannot be used to sample from hidden populations because information about the identify of individuals in the group is either not available or cannot be reliably ascertained from survey respondent's reports. There are non-probability sampling approaches for reaching them, such as respondent-driven sampling, venue-based sampling, and on-line sampling; however, these techniques raise questions about potential bias and lack of generalizability if used to estimate a population size or disparity level. Lack of sample frames is a problem that requires more fundamental methods and data source work to be done, even to understand the basic issues about the population and its health.

In addition to issues with size, dispersion, and accessibility, small populations may also be hard-to-reach because their unwillingness to participate in research studies, or because of negative histories with social institutions and with past research. Tourangeau et al. (2014) listed many examples of the challenges with surveying these groups and possible methods that may be used. In general, this monograph provides a useful taxonomy for distinguishing and surveying hard-to-survey populations.

Intervention Study

For the purpose of this workshop "intervention study" is meant to define any study with a goal to establish a causal effect of a treatment applied to individuals. The randomized controlled trial (RCT) is the gold standard in health intervention research. Other examples of intervention studies are prevention, behavioral, and implementation studies with different approaches to randomization. Clinical trials to establish drug efficacy are the most well-known examples of intervention studies and frequently use randomized controlled trials (RCT). The challenge, especially with small populations or small samples, include a number of logistical and ethical issues that can arise, along with inefficiencies in how the RCT makes use of information that can result in low power and low external validity. An RCT is also impractical in some real world settings such as dissemination and implementation studies. Recent alternative designs (stepped wedge, interrupted time series, regression discontinuity, and dynamic waitlist) make use of optimization strategies to more efficiently use available information to maximize power with modest sample size. All of these new approaches have their strengths and weaknesses that should be carefully considered in any given situation.

IOM (2001) lists the following situations that might warrant a *small* clinical trial: rare diseases, unique study populations, individually tailored therapies, environments that are isolated, emergency situations, and public health urgency. The same situations might result in small samples for any type of intervention trial. IOM (2001) also provides a summary of statistical issues, designs and analysis approaches that might be useful for small clinical trials and provides the following recommendations for researchers designing such studies: define the research question; tailor the design; clarify methods of reporting of results; perform corroborative statistical analysis; and exercise caution in interpretation. IOM (2001, p 10) also recommended more federal funding of research on alternative designs for small sample studies.

The need to address methodological challenges concerning small populations is predicated on determining whether or not a "small" population is meaningfully different and should be studied. Answers to this question arise from concerns of populations, funders and researchers. These may differ. For the researchers and funders, answers to the "meaningfully different" question may arise from the significant amount of prior analysis including surveillance studies, epidemiological studies, laboratory studies, etc. that occur prior to funding. IOM (2010, pp. 2, 3, and 5-9) proposes that the L.E.A.D framework (for Locate Evidence, Evaluate Evidence, Assemble Evidence, and Inform Decisions) be followed when designing a study. It goes on to specify that first the researcher must identify the question to be answered by the study. Then the following steps should be followed: (1) Locate all the types of evidence that could be useful in answering the question. (2) Evaluate the quality of the evidence, especially its level of certainty (internal validity) and generalizability. (3) Develop a transparent and comprehensive summary of the evidence related to why an action should be taken, what that action should be, and how it should be taken. If evidence is limited, examine the potential for blending it with theory, professional experience, and local wisdom. (4) Use the summary to inform the decision-making process. Explicitly or implicitly, plans and proposals for research on small populations must address the meaningfully different criteria and consider the different needs of the relevant stakeholders. Lack of clarity concerning this issue may well be the source of some of the frustration felt over funding patterns addressing health in small populations.

If evidence is limited other options might be more initial data source work, accumulation of data over multiple small studies or efforts to understand mechanism in biological studies. Accumulation of data over multiple small studies might be challenging because slow accumulations of results require data harmonization and may be subject to secular change. Identification of appropriate biomarkers or intermediate endpoints may allow studies with larger or more easily obtainable outcomes.

In specifying the target population of the study, researchers may consider whether the research question would support the combination of the "small population" with others to make the study more manageable. A population may not be small if it can be combined with similar groups from other geographic areas. For example, developing interventions for Nicaraguan immigrants in the US might prove difficult because of the small population size. However, if relevant social, psychological and biological research suggested that the intervention approach was appropriate for Central American immigrants generally then the scope of the study might

be expanded. Even in this case, whether or not to combine groups is a difficult question. There are huge differences in the environments in different Central American countries and immigrants in different parts of the U.S. experience health care differently. Proposals to group demographic subgroups in order to strengthen etiological or intervention studies require considerable discussion and sensitivity, especially where there has been history of injustice and health disparities.

In some instances, however, researchers may be even more challenged when subgroups of the small target population are ethno-culturally distinct. This cultural distinctiveness can require adapted or culturally grounded interventions for the subgroups, requiring small sample intervention research to test if the new intervention is effective by subgroup. This is typically the case for subpopulations, for example, for an American Indian, Alaska Native, and Native Hawaiian group, or for a rural Scotch-Irish Appalachian population.

A further challenge involves contextual variables such as toxic exposures from a single factory or busy roadway or neighborhood characteristics such as lead levels in homes or neighborhood poverty. Because such exposures may occur over small areas, they may either define a small population of interest in itself (e.g., residents near a specific factory or mine) or they may result in confounding or other statistical issues for studies of small populations.

However small populations are defined, intervention studies with these groups will likely necessitate small sample research. Study designs are often underpowered due to their sample sizes. As noted by Fok et al. (2015),

> "It is therefore tempting to define "small" merely in terms of statistical power. However, lack of power may result from weak effects as much as from sample size. A sample size that is adequate for a medication study with strong effects may be insufficient for a psychosocial prevention trial with more modest effect sizes."

The goals of this work shop include clarifying the overall taxonomy of small population challenges and articulating opportunities and gaps in efforts to address them.

References

Devers, K., Gray, B., Ramos, C., Shah, A., Blavin, F., and Waidmann, T. (2013). The Feasibility of Using Electronic Health Records (EHRs) and other Electronic Health Data for Research on Small Populations", A report by the Urban Institute, delivered to the Office of the Assistant Secretary for Planning and Evaluation at HHS. https://www.urban.org/research/publication/feasibility-using-electronic-health-data-research-small-populations.

Fok, C.C., Henry, D. and Allen, J. (2015). "Maybe small is too small a term: Introduction to advancing small sample prevention science." Prevention Sciences. Oct; 16 (7):943-949. Springer.

Institute of Medicine (2001), "Small Clinical Trials: Issues and Challenges. Committee on Strategies for Small Number Participant Clinical Research Trials. Evans, C. H. and Ildstad, S. T. (editors). National Academy Press.

Institute of Medicine (2010), "Bridging the Evidence Gap in Obesity Prevention: A Framework to Inform Decision Making." Washington D.C. The National Academy Press.

Tourangeau, R., Edwards, B., Johnson, T.P., Wolter, K.M, and Bates, N. (editors) (2014). "Hard-to-Survey Populations." Cambridge University Press.

Srinivasan, S., Moser, R. Willis, G., Riley, W., Alexander, M., Berrigan, D., and Kobrin S. (2015). "Small is Essential: Importance of Subpopulation Research in Cancer Control." Editorial, American Journal of Public Health, April 23, 2015.

EDITORIAL

Small Is Essential: Importance of Subpopulation Research in Cancer Control

The ability to harness the benefits of "big data" has had a revolutionary impact on science, with its focus on the volume and variety of data sources, and application of both traditional and innovative analytic methods appropriate for large, aggregated data sets. We are concerned, however, about the opposite: "small data," for which the size, dispersion, or accessibility of the population of interest makes it difficult to obtain adequate sample sizes to test specific research questions. Examples include racial or ethnic subpopulations (e.g., Honduran Latin Americans), populations occurring in specific geographic areas (e.g., reservations), and populations that have relatively rare characteristics (e.g., transgender persons). A great challenge is determining when a small group is of practical or theoretical interest (Figure 1). We define "practical and theoretical interest" broadly to include issues involving social justice, biological or geographic factors, and disease burden.[1] Ultimately, it is critical to ensure that all segments of the US population benefit from this research and from the latest technologic advances in cancer care services and delivery.

INCLUDING UNDERREPRESENTED GROUPS IN RESEARCH

An example of the potential negative ramifications of not including underrepresented groups in research—or inappropriately aggregating them across groups—comes from the study of racial and ethnic health disparities and issues of equity in the United States. Intervention research often does not include a wide range of racial/ethnic subgroups; so it is not feasible to test whether an intervention created specifically for the majority group is also efficacious for the subgroups. Likewise, the ability to test whether an intervention can be altered for a particular subgroup is also often not possible. Epidemiological and surveillance research usually involves the inclusion of "minority or underserved populations" in addition to White or non-Hispanic White (NHW) groups. While this has allowed for a better understanding of these smaller populations and provides some progress toward addressing health inequities, there remain pockets of communities that are severely underrepresented within the broader "minority and underserved populations."[2–6]

As a further example, although Asian Americans as a whole have high incomes and good health outcomes overall when compared with NHWs, Hispanics, African Americans, and American Indian/Alaska Natives, this generalized statistic masks the fact that subgroups of Asian Americans, such as the Cambodians and Hmong, lag severely behind other Asian Americans.[3,4,7,8] Even within the NHW population there are communities that have long been disadvantaged (such as those living in Appalachian states), with low levels of income, literacy, and health outcomes.[9–11] These subgroups have generally been omitted or excluded from the research process because of challenges with identification and recruitment. Through this commentary, we hope to encourage research in subpopulations; we recommend both the development of new methods and the innovative use of existing methodological and analytic strategies across both intervention and epidemiological research.

ALTERNATIVE STUDY DESIGNS AND ANALYSIS PROCEDURES

There is growing recognition that to implement interventions in small populations, it may be necessary to consider alternative study designs, such as the use of single-case designs attributing propensity scores, and randomized group designs. In 2013, many of the studies submitted to the Division of Cancer Control and Population Sciences at the National Cancer Institute (NCI) on subpopulation research that did not score well in peer review received comments that the randomized clinical trial design was not appropriate because the sample size was insufficient to detect changes in the effect of the intervention. This criticism raises the question of whether these studies would be better suited for alternatives to the standard randomized control trial design, such as single-case, within-subject controls, and a variety of quasi-experimental designs.

One solution for testing interventions in small samples is to focus on within-rather than between-group designs. Because a within-group design uses the sample as its own control, there is

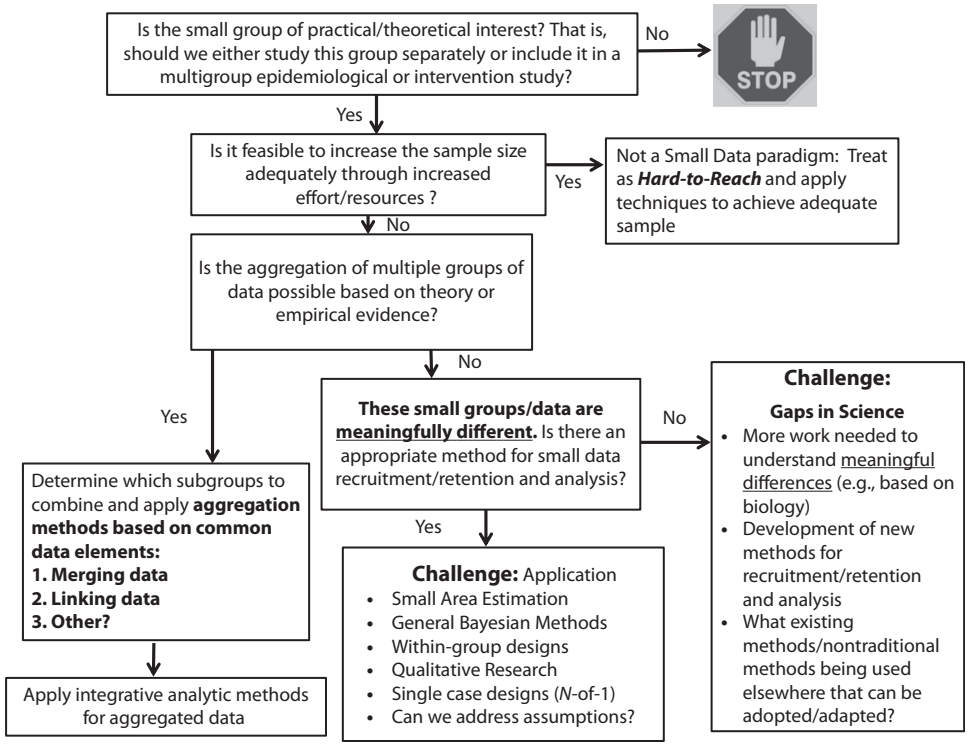


FIGURE 1—Research with small data: identifying challenges.

no need for a separate control group, reducing by up to half the sample size required for accurate statistical comparisons. Among group designs, there are a number of quasi-experimental approaches that could be considered, including interrupted time series[12–15] and stepped wedge designs, the latter being particularly useful for studies in which there are distinct and dispersed cohorts or communities in which the intervention can be rolled out in a staggered manner.[16,17] Single-case studies involving a series of *N*-of-1 trials could be used to test intervention adaptations in an iterative manner, and Bayesian estimates can be produced from this series of trials to evaluate the potential generalizability of the findings to the subpopulation.[18] Within-subject designs require more intensive longitudinal data than typically obtained through between-subject designs, but the advent of technologies for capturing temporally dense data, such as ecological momentary assessment and passive sensor technologies, makes these approaches more viable. Such data could also be used in conjunction with multilevel analyses of behavior across different spatial areas. This kind of study design can be statistically powerful, even with modest numbers of samples per geographic unit.[19]

For epidemiological research, innovative *recruitment* methods may be very useful. For example, respondent-driven sampling[20,21] has been successfully employed to identify and recruit groups for studies in which there is no existing sample frame, such as drug addicts or ethnic subgroups. Innovative *analytic* approaches, such as integrative data analysis,[22] could be employed where independent data sets are combined together and analyzed as a whole to produce adequate representation and sample sizes. Integrative data analysis can also be used to combine data across multiple iterations of the same national survey where any one sample does not constitute an adequate sample size.

ADDRESSING THE CHALLENGE OF SMALL DATA

The National Institutes of Health (NIH)—and by extension the NCI—has an obligation to conduct research to improve the health of all Americans, not just the health of the majority population or those who are easy to identify. We therefore recommend the development and the use of methodological and analytic procedures to allow subpopulations to be meaningfully included in research. Figure 1 illustrates a model for determining when a "small" group is of research interest. However, it is also clear that other entities need to be involved in identifying populations of interest and in developing initiatives to address these groups, not just those who are responsible for grant funding decisions. For example, at the NIH, training for peer reviewers in study sections may be needed to ensure that they are knowledgeable about these innovative methods so that sound, rigorous scientific applications that employ them are understood and scored appropriately.

In addressing the above issues, NCI is planning a workshop to address three areas related to small populations:

(1) identification, recruitment, and retention strategies;
(2) epidemiological design and analytic approaches for small samples; and
(3) intervention design and analytic approaches for subpopulations.

Based on the products of this workshop and responses to this editorial, the NCI will explore next steps to strengthen subpopulation research. ■

Shobha Srinivasan, PhD
Richard P. Moser, PhD
Gordon Willis, PhD
William Riley, PhD
Mark Alexander, MSc
David Berrigan, PhD, MPH
Sarah Kobrin, PhD, MPH

About the Authors

All of the authors are with the Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD.

Correspondence should be sent to Shobha Srinivasan, Health Disparities Research Coordinator National Cancer Institute, Division of Cancer Control and Population Sciences, 9609 Medical Center Drive, Room 4E432, MSC 9764, Rockville, Maryland 20850 (e-mail: ss688k@nih.gov). Reprints can be ordered at http://www.ajph.org by clicking the "Reprints" link.

This editorial was accepted August 12, 2015.

doi:10.2105/AJPH.2014.302267

Contributors

All authors contributed equally to this editorial.

Acknowledgments

The authors would like to thank and acknowledge the guidance and support of Robert T. Croyle, PhD. The authors would also like to thank Kathleene Ulanday, MPH, for assistance with the citations.

References

1. Waller LA, Gotway CA. *Applied Spatial Statistics for Public Health Data.* Hoboken, NJ: John Wiley & Sons; 2004.

2. Yancey AK, Ortega AN, Kumanyika SK. Effective recruitment and retention of minority research participants. *Annu Rev Public Health.* 2006;27: 1–28.

3. Islam NS, Khan S, Kwon S, Jang D, Ro M, Trinh-Shevrin C. Methodological issues in the collection, analysis, and reporting of granular data in Asian American populations: historical challenges and potential solutions. *J Health Care Poor Underserved.* 2010;21 (4):1354–1381.

4. Ghosh C. Healthy People 2010 and Asian Americans/Pacific Islanders: defining a baseline of information. *Am J Public Health.* 2003;93(12):2093–2098.

5. Institute of Medicine. *The Health of Lesbian, Gay, Bisexual, and Transgender Aging: Biographical Approaches for Inclusive Care and Support.* London, UK: Jessica Kingsley; 2012.

6. Fredriksen-Goldsen KI, Kim HJ, Barkan SE, Muraco A, Hoy-Ellis CP. Health disparities among lesbian, gay, and bisexual older adults: results from a population-based study. *Am J Public Health.* 2013;103(10): 1802–1809.

7. Chen MS Jr, Hawks BL. A debunking of the myth of healthy Asian Americans and Pacific Islanders. *Am J Health Promot.* 1995;9(4):261–268.

8. National Heart, Lung, and Blood Institute. Addressing Cardiovascular Health in Asian American and Pacific Islanders: a Background Report. 2000. Available at: http://www.nhlbi.nih.gov/health/prof/heart/other/aapibkgd/aapibkgd.pdf. Accessed June 25, 2014.

9. Behringer B, Friedell GH. Appalachia: where place matters in health. *Prev Chronic Dis.* 2006;3:A113.

10. Appalachian Regional Commission. Underlying Socioeconomic Factors Influencing Health Disparities in the Appalachian Region. Final Report. 2008. Available at: http://www.arc.gov/assets/research_reports/SocioeconomicFactorsInfluencingHealthDisparitiesinAppalachianRegion5.pdf. Accessed June 25, 2014.

11. Vanderpool RC, Gainor SJ, Conn ME, Spencer C, Allen AR, Kennedy S. Adapting and implementing evidence-based cancer education interventions in rural Appalachia: real-world experiences and challenges. *Rural Remote Health.* 2011;11(4):1807.

12. Cook TD, Campbell DT. *Quasi-Experimentation: Design & Analysis Issues for Field Settings.* Boston, MA: Houghton Mifflin Company; 1979.

13. Hartmann DP, Gottman JM, Jones RR, Gardner W, Kazdin AE, Vaught RS. Interrupted time-series analysis and its application to behavioral data. *J Appl Behav Anal.* 1980;13(4): 543–559.

14. Gillings D, Makuc D, Siegel E. Analysis of interrupted time series mortality trends: an example to evaluate regionalized perinatal care. *Am J Public Health.* 1981;71(1):38–46.

15. Biglan A, Ary D, Wagenaar AC. The value of interrupted time-series experiments for community intervention research. *Prev Sci.* 2000;1(1): 31–49.

16. Mdege ND, Man MS, Taylor Nee Brown CA, Torgerson DJ. Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation. *J Clin Epidemiol.* 2011;64(9): 936–948.

17. Woertman W, de Hoop E, Moerbeek M, Zuidema SU, Gerritsen DL, Teerenstra S. Stepped wedge designs could reduce the required sample size in cluster randomized trials. *J Clin Epidemiol.* 2013;66(7):752–758.

18. Zucker DR, Ruthazer R, Schmid CH. Individual (N-of-1) trials can be combined to give population comparative treatment effect estimates: methodologic considerations. *J Clin Epidemiol.* 2010;63(12): 1312–1323.

19. Raudenbush SW, Bryk AS. *Hierarchical Linear Models.* 2nd ed. Thousand Oaks, CA: Sage Publications; 2002.

20. Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations. *Soc Probl.* 1997; 44(2):174–199.

21. Salganik MJ, Heckathorn DD. Sampling and estimation in hidden populations using respondent-driven sampling. *Sociol Methodol.* 2004;34: 193–239.

22. Curran PJ, Hussong AM. Integrative data analysis: the simultaneous analysis of multiple data sets. *Psychol Methods.* 2009;14(2):81–100.

C

TAB C

SESSION 1: What do we Mean by Small Populations? How to Decide when a Small Population is Important or Meaningfully Different Enough to Study? Why did we Structure the Workshop this Way?

Presentations in this tab:

Data Issues in Studying Small Populations: Challenges, Opportunities, and a Case Study
Scarlett Lin Gomez, University of California, San Francisco

Fielding Studies in Underrepresented Populations: Challenges and Considerations
Lisa Signorello, Division of Cancer Prevention, National Cancer Institute

Data Issues in Studying Small Populations

Challenges, Opportunities, and a Case Study

Scarlett Lin Gomez, MPH, PhD

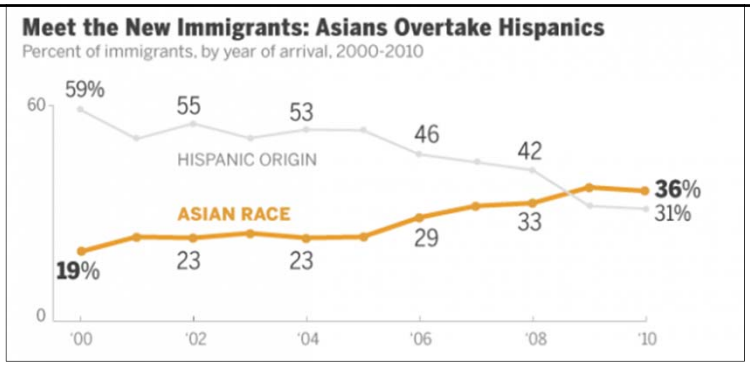
Department of Epidemiology & Biostatistics,
University of California, San Francisco

UCSF University of California San Francisco Department of Epidemiology & Biostatistics

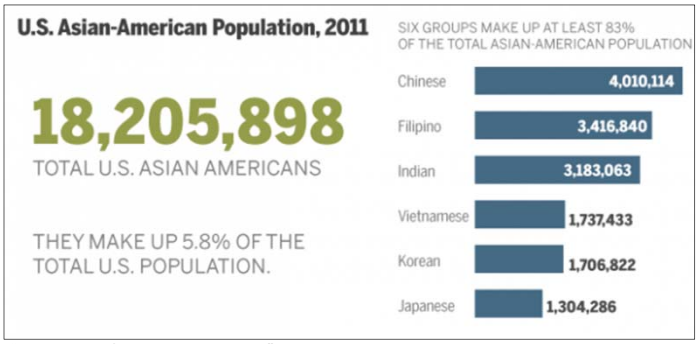
CANCER PREVENTION INSTITUTE OF CALIFORNIA
Preventing Cancer. Promoting Life.

Case study: cancer statistics in Asian Americans

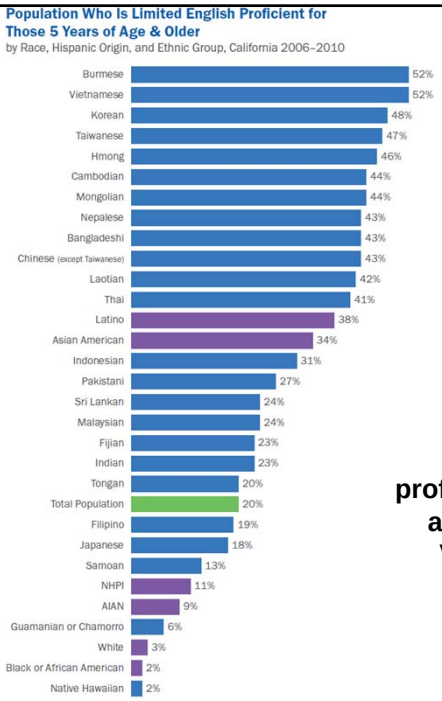
CANCER PREVENTION INSTITUTE OF CALIFORNIA
Preventing Cancer. Promoting Life.



A rapidly growing population



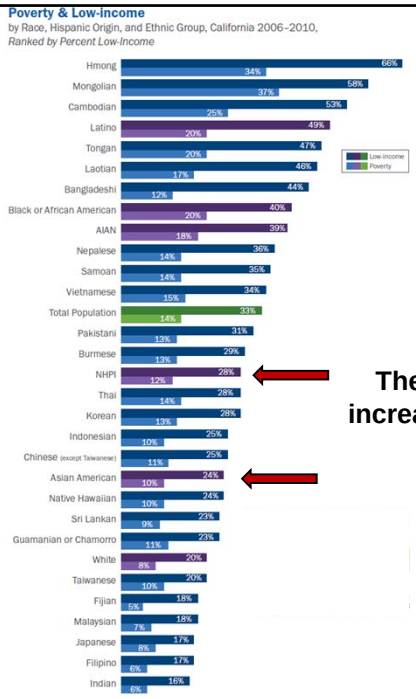
From "The Rise of Asian Americans." Pew Report. 2012



Tremendous heterogeneity & hidden socioeconomic disparities

% limited English proficient range from 52% among Burmese and Vietnamese to 18% among Japanese.

From "A Community of Contrasts. Asian Americans, Native Hawaiians and Pacific Islanders in California." Asian American Center for Advancing Justice. 2013.



The number of poor AAs increased 50% between 2007 and 2011.

From "A Community of Contrasts. Asian Americans, Native Hawaiians and Pacific Islanders in California." Asian American Center for Advancing Justice. 2013.

Breast cancer incidence patterns in Asian American women

Breast cancer incidence rates*, California, 1988-2004

Race/Ethnicity	Rate (95% CI)
N-H White	146.1 (145.5-146.7)
Asian **	82.7 (81.6-83.8)

* Rates, per 100,000, adjusted to the US 2000 standard
** Asian = Chinese + Japanese + Filipina + Korean + Vietnamese + South Asian

Breast cancer incidence rates*, California, 1988-2004

Race/Ethnicity	Rate (95% CI)
N-H White	146.1 (145.5-146.7)
Asian **	82.7 (81.6-83.8)
Chinese	**73.5 (71.6-75.4)**
Japanese	**102.5 (99.3-105.9)**
Filipina	**100.4 (98.1-102.8)**
Korean	**46.3 (43.8-49.0)**
South Asian	**77.0 (72.1-82.1)**
Vietnamese	**59.9 (56.7-63.1)**

* Rates, per 100,000, adjusted to the US 2000 standard
** Asian = Chinese + Japanese + Filipina + Korean + Vietnamese + South Asian
From: Gomez et al. Am J Public Health 2010

Breast cancer incidence rates*, California, 1988-2004

Race/ ethnicity	Combined (US+foreign born)	US-born	Foreign-born	Rate ratio (95% CI) (US/foreign)
N-H White	146.1	-	-	-
Asian**	82.7	120.6	76.3	1.6 (1.5-1.6)
Chinese	73.5	122.1	66.3	1.8 (1.7-2.0)
Japanese	102.5	106.1	103.1	1.0 (1.0-1.1)
Filipina	100.4	129.5	98.2	1.3 (1.2-1.4)

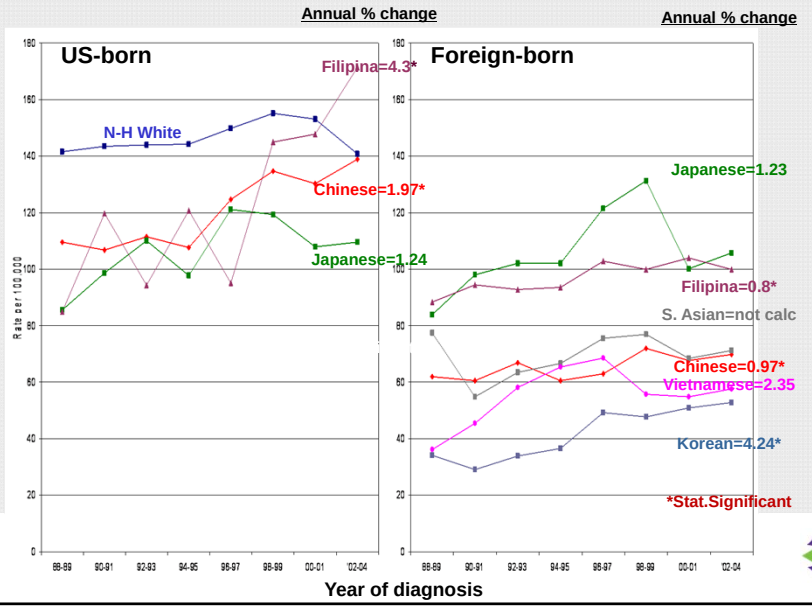
* Rates, per 100,000, adjusted to the US 2000 standard
** Asian = Chinese + Japanese + Filipina + Korean + Vietnamese + South Asian
From: Gomez et al. Am J Public Health 2010

Breast cancer incidence rates* by age, California, 1988-2004

Race/ethnicity	≤44 yrs	45-54 yrs	≥55 yrs
N-H White	27.1	240.7	449.2
US-born			
Chinese	**39.8**	**276.9**	275.6
Japanese	23.9	205.8	294.2
Filipina	**43.1**	**334.3**	263.8
Foreign-born			
Chinese	18.9	161.2	167.9
Japanese	24.8	196.0	283.6
Filipina	25.9	215.1	245.0

Rates, per 100,000, adjusted to the US 2000 standard
From: Gomez et al. Am J Public Health 2010

Breast cancer trends by nativity, 1990-2004, California



ARTICLE

Cancer Incidence Trends Among Asian American Populations in the United States, 1990 to 2008

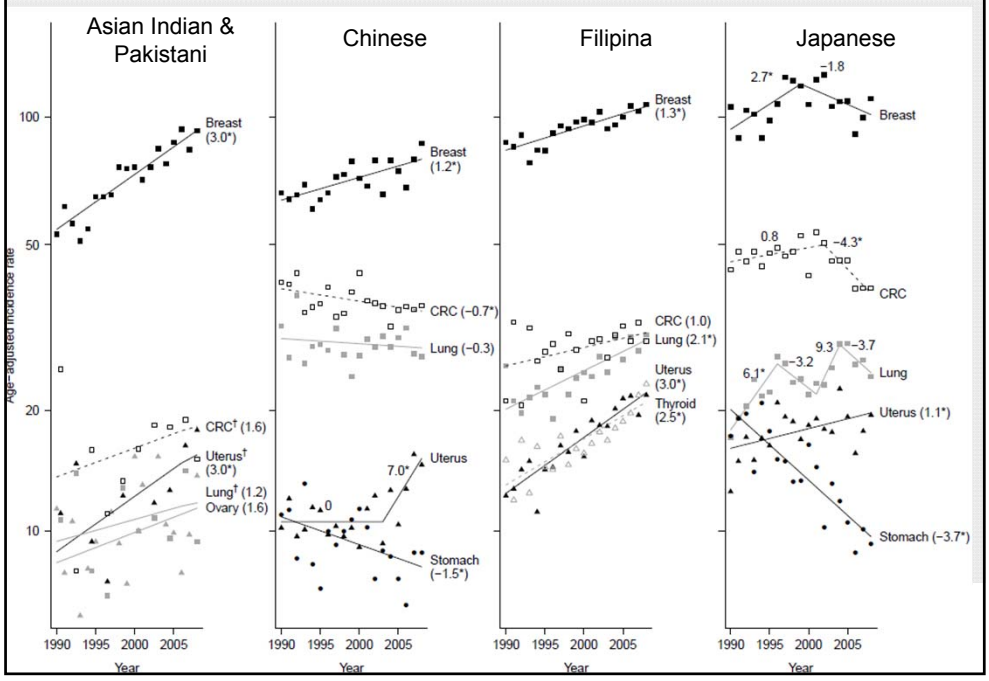
Scarlett Lin Gomez, Anne-Michelle Noone, Daphne Y. Lichtensztajn, Steve Scoppa, James T. Gibson, Lihua Liu, Cyllene Morris, Sandy Kwong, Kari Fish, Lynne R. Wilkens, Marc T. Goodman, Dennis Deapen, Barry A. Miller

Manuscript received September 19, 2012; revised April 17, 2013; accepted April 18, 2013.

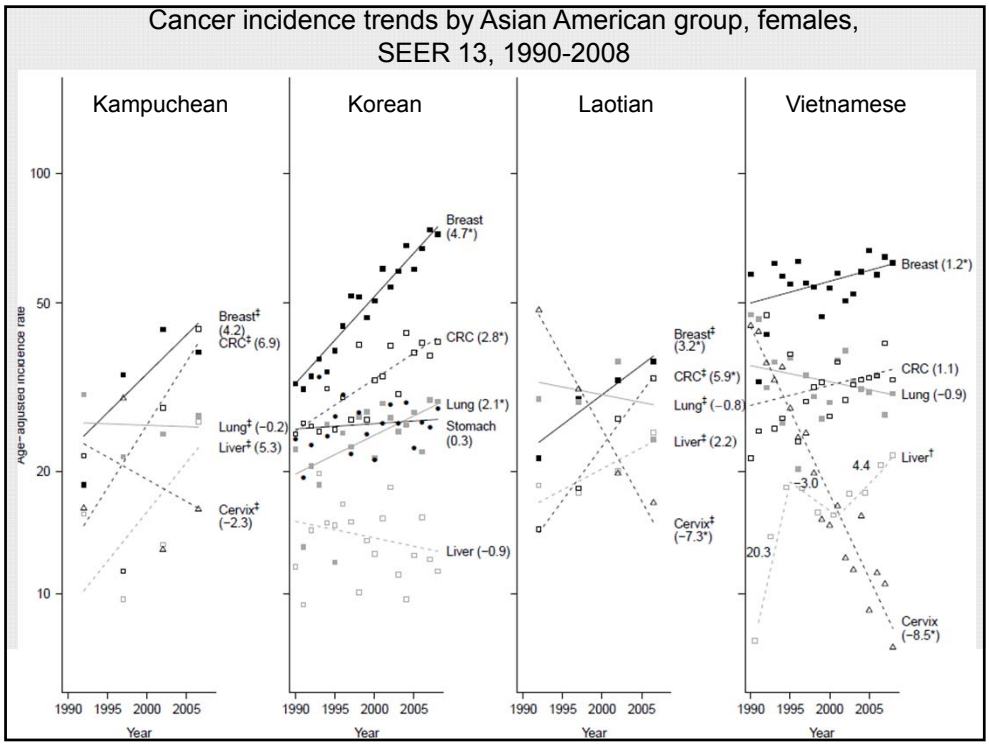
Correspondence to: Scarlett Lin Gomez, PhD, Cancer Prevention Institute of California, 2201 Walnut Ave, Ste 300, Fremont, CA 94538 (scarlett@cpic.org).

Journal of the National Cancer Institute 2013

Cancer incidence trends by Asian American group, females, SEER 13, 1990-2008



Cancer incidence trends by Asian American group, females, SEER 13, 1990-2008



Published OnlineFirst September 30, 2016; DOI: 10.1158/1055-9965.EPI-16-0167

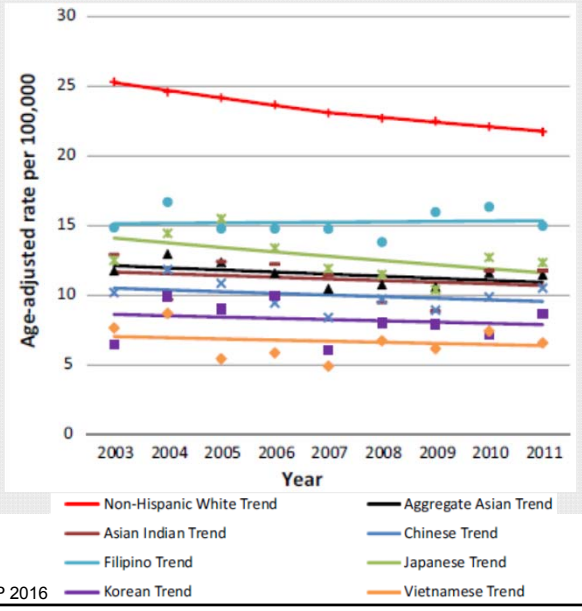
Research Article

Cancer Epidemiology, Biomarkers & Prevention

The Burden of Cancer in Asian Americans: A Report of National Mortality Trends by Asian Ethnicity

Caroline A. Thompson[1,2], Scarlett Lin Gomez[3,4,5], Katherine G. Hastings[6], Kristopher Kapphahn[7], Peter Yu[8], Salma Shariff-Marco[3,4,5], Ami S. Bhatt[9,10], Heather A. Wakelee[5,11], Manali I. Patel[11,12], Mark R. Cullen[6,13], and Latha P. Palaniappan[6]

Breast cancer mortality, U.S., 2003-2011



Thompson et al., CEBP 2016

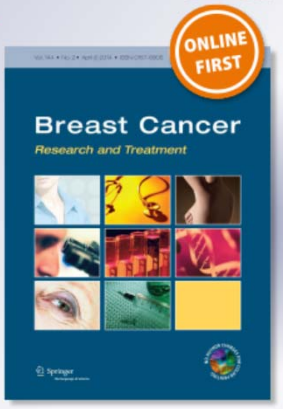
Breast cancer in Asian Americans in California, 1988–2013: increasing incidence trends and recent data on breast cancer subtypes

Scarlett Lin Gomez, Julie Von Behren, Meg McKinley, Christina A. Clarke, Salma Shariff-Marco, Iona Cheng, Peggy Reynolds, et al.

Breast Cancer Research and Treatment

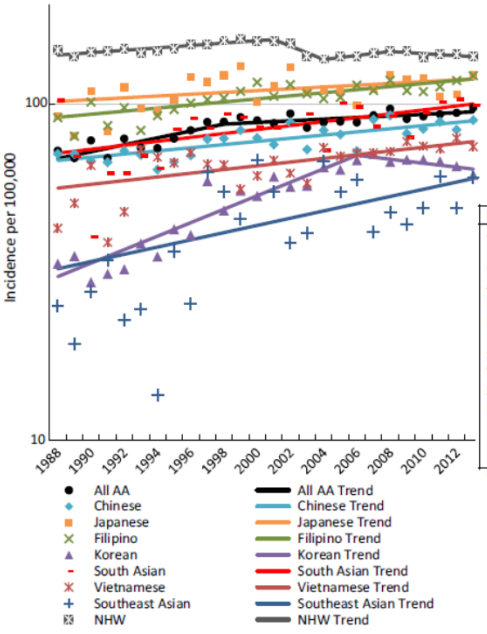
ISSN 0167-6806

Breast Cancer Res Treat
DOI 10.1007/s10549-017-4229-1



Springer

Incidence rates over time of invasive breast cancer among Asian American ethnic groups and non-Hispanic Whites, California, 1988-2013



Ethnic Group	Year Segment	APC (95% CI)*
All Asian Americans	1988-1998	**2.4 (1.2, 3.6)**
	1998-2013	**0.6 (0.1, 1.0)**
Chinese	1988-2013	**1.1 (0.7, 1.5)**
Japanese	1988-2013	0.6 (0.0, 1.2)
Filipino	1988-2013	**1.1 (0.7, 1.4)**
Korean	1988-2006	**4.7 (3.8, 5.7)**
	2006-2013	-1.4 (-3.5, 0.9)
South Asian	1988-2013	**1.4 (0.7, 2.1)**
Vietnamese	1988-2013	**1.3 (0.7, 1.9)**
Southeast Asian	1988-2013	**2.5 (0.8, 4.2)**
NH White	1988-2001	**0.8 (0.5, 1.1)**
	2001-2004	-4.5 (-9.4, 0.7)
	2004-2008	1.4 (-1.4, 4.2)
	2008-2013	-0.7 (-1.9, 0.5)

*Bolded text indicates a significant APC

Incidence rates adjusted to 2000 U.S. Standard population

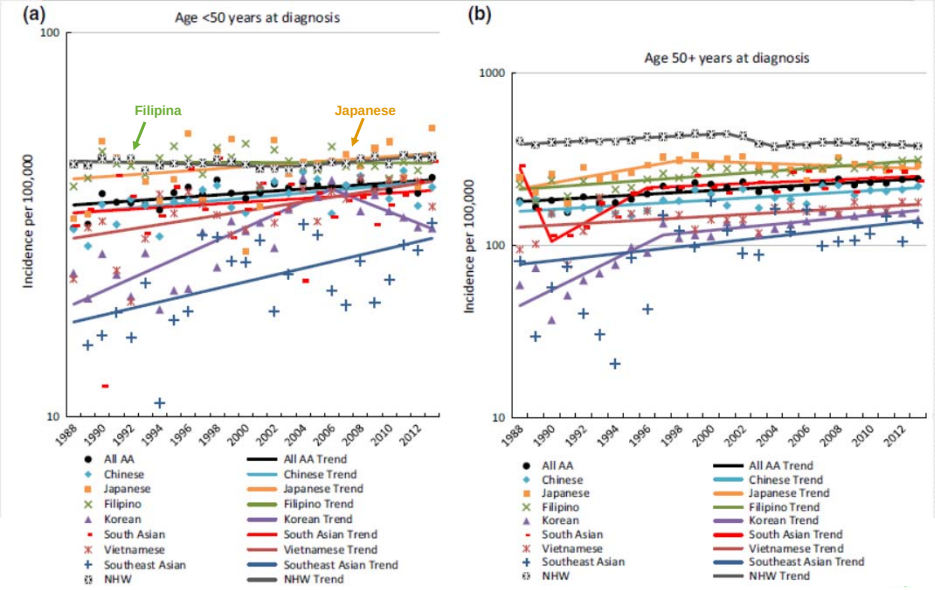
Gomez et al. BCRT 2017

Higher incidence of breast cancer in young Asian American women?

- 3 studies showed higher incidence rates in young Asian American women (~age <50) relative to non-Hispanic White women*
  - More pronounced in US-born Asian American women (Gomez et al. AJPH 2010)
- Recent international analysis shows trend may be due to cohort (Sung et al. JNCI 2015) or period (Wang et al. Int J Env Res Pub Health 2015) effects of increasing breast cancer rates among Asian populations worldwide, not age-specific effect
  - due to changing risk factors

*Liu et al. Int J Cancer 2012; Reynolds et al. Eth & Dis 2011; Gomez et al. AJPH 2010

Higher incidence of breast cancer in young Asian American women (cont)?



*Gomez et al. BCRT 2017

Disparities among distinct ethnic groups

- More HER2Neu+ tumors?
  - Higher proportional prevalence of HER2Neu+ tumors (Telli et al. BCRT 2010)
  - Compared to non-Hispanic white women, Filipinas and older Vietnamese women had higher incidence rates of some HER2+ subtypes (Gomez et al., BCRT 2017)
- Increasing rates of distant stage disease among Filipinas (2.1% per year) (Gomez et al., BCRT 2017)

Conclusions

Disaggregated data by ethnicity, nativity, and age shows that:

→Vastly differing patterns in incidence, mortality, and incidence and mortality trends across sub-populations.

→Burden of breast cancer is <u>not low</u> among Asians!

"I was diagnosed with breast cancer 5 years ago. When the doctor told me that I had breast cancer I was in shock because I thought this is a white women/old people disease. Later, I was even more surprised to find out that many of the Asian women I knew had breast cancer, but nobody talked about it."

(personal communication from a breast cancer survivor)

Cancer research in Asian Americans, Native Hawaiians, Pacific Islanders:
Challenges & opportunities

Gaps in research

- NCI portfolio review showed virtually no studies of cancer etiology focused on this population (Nguyen, Srinivasan, et al., CEBP 2014)
- Lack of representation in current NCI-funded Cancer Epidemiology Cohorts and other cohorts
  - Multiethnic Cohort (MEC) includes only Japanese Americans and Native Hawaiians (with sufficient numbers for ethnic-specific analyses)
  - Many cohorts in Asia, but none in the U.S.

Challenges

- Small numbers in many ethnic groups
  - How granular can we go?
- Ethnicity information (often) not captured in health surveys, registries, hospital data
- Lack of standardization in data collection
- Other relevant data not captured, e.g., nativity, immigration factors, language, SES

Size + heterogeneity = Opportunities for accelerating cancer discoveries

- Heterogeneity (risk factors, disease risk) within population provides potential opportunities for identifying novel risk factors
- Research into what determines favorable prognosis despite poor prognosis tumor biology
- Potential of migrant studies, longitudinal studies, intergenerational studies


CEBP FOCUS

Cancer Research in Asian American, Native Hawaiian, and Pacific Islander Populations: Accelerating Cancer Knowledge by Acknowledging and Leveraging Heterogeneity

Scarlett Lin Gomez[1,2], Sally L. Glaser[1,2], Pamela L. Horn-Ross[1,2], Iona Cheng[1,2], Thu Quach[1,2], Christina A. Clarke[1,2], Peggy Reynolds[1,2], Salma Shariff-Marco[1,2], Juan Yang[1], Marion M. Lee[3], William A. Satariano[4], and Ann W. Hsing[1,2]

"My 1991 diagnosis was only obtained after I sought a second opinion, following a surgical oncologist's "refusal" to biopsy a very prominent and palpable breast lump. The reasons he refused to perform the biopsy was because I was *"too young to have breast cancer"*, had *"no family history of cancer"*, and ***"besides, Asian women don't get breast cancer"***. I believe the latter statement was made because of his familiarity with NCI SEER race/ethnic cancer data for "API" populations, which -- as you are well aware -- were then and continue to be reported in the aggregate. I have that surgeon to thank for turning me into a fierce cancer advocate*" (personal communication from Susan Shinagawa, cancer survivor)*

Thank you!

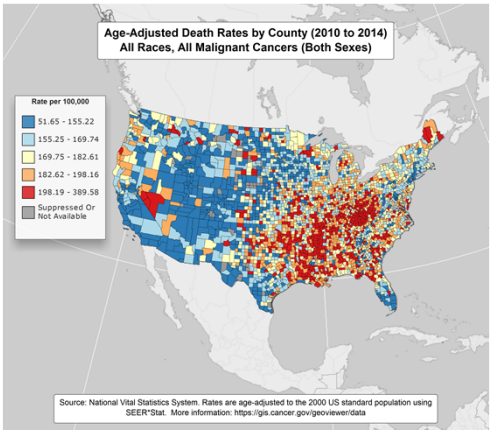
Scarlett.gomez@ucsf.edu

Fielding studies in underrepresented populations: challenges and considerations

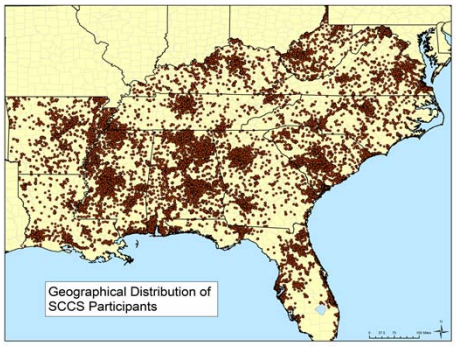
Lisa B. Signorello, ScD
Senior Biomedical Scientist
Acting Director, Cancer Prevention Fellowship Program
Division of Cancer Prevention, NCI/NIH

NIH NATIONAL CANCER INSTITUTE

The views expressed in this presentation reflect those of the author and do not necessarily reflect the official views of the National Cancer Institute, the National Institutes of Health, the U.S. Department of Health and Human Services, or the federal government.



Southern Community Cohort Study (SCCS)
R01 CA092447 PI: Blot, Zheng
~85,000 participants in a longitudinal study



Signorello et al., J Nat Med Assoc, 2005

US subgroup populations

African American | Rural | Poor

Proportion of US population → 13% | 19% | 13%

Health differences

Not well represented in population-based cancer research

Access challenges

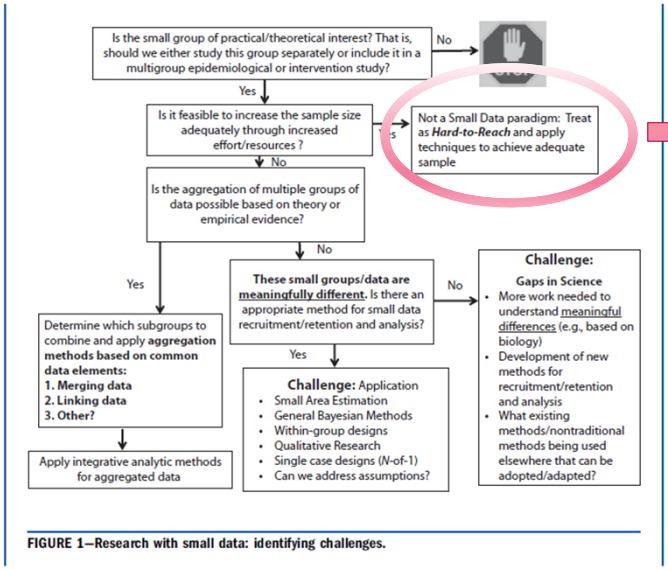


FIGURE 1—Research with small data: identifying challenges.

Commonalities

Srinivasan et al., AJPH, 2015

Some commonalities: Hard-to-Reach and Small Groups

Gaps in knowledge

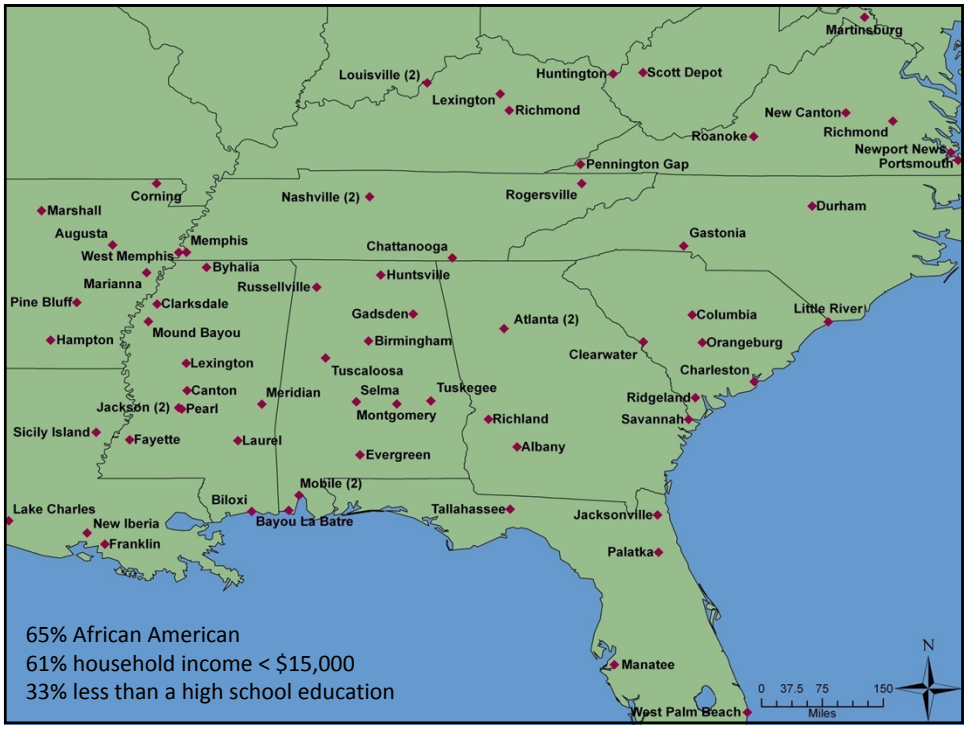
Untapped or overtapped populations

Study design / methodology:

1. Identify and access population
2. Recruitment
3. Data collection
4. Retention

Intensified efforts
Planning
Developmental groundwork
Time and resources

New strategies?



Community-based, in-person recruitment

Community engagement

Trust building

Making it win:win


CHC Engagement
Medical Director
CEO
Community hire
Director of Nursing


INFORMED CONSENT FORM


ThermoSafe
Insulated Shipper
ThermoSafe
Insulated Shipper
Neal Guthrie


Sheila
Theresa
Foster

Montgomery Primary Health Care Center, Montgomery, AL

St. Matthews Family Health Center, St. Matthews, SC

Data Collection (what and how)

Primary data collection, and the need for tailoring.

Exposures → Unique

- Understanding the context
- Focus groups, other foundational work
- Develop and validate questions or instruments
- Develop new geospatial indices
- Develop new biomarkers

Exposures → Not unique

- Language
- Cultural slang / colloquialisms
- Literacy
- Need validation in the specific population
- Questions or instruments need adaptation

Refining the SCCS Food Frequency Questionnaire (FFQ)

Using data files from the 24-hour dietary recalls conducted within NHANES III (1988–1994), CSFII (1994–1996; day 1 recalls), NHANES 1999–2000, NHANES 2001–2002, and NHANES 2003–2004 (day 1 recalls)

SCCS FFQ Item "Broccoli, cabbage, brussels sprouts, or cauliflower"

White females

1. **Broccoli, raw (15.2%)**
2. **Cauliflower, raw (10.3%)**
3. **Broccoli, cooked, from fresh, no fat added (9.5%)**

Black females

1. **Cabbage, green, cooked, fat added (21.4%)**
2. **Cabbage, green, cooked, fat added in cooking (18.0%)**
3. **Broccoli, cooked, NS as to form, fat not added in cooking (8.3%)**

Signorello et al., AJE, 2009

Retention

Hard to reach can mean hard to reach again

- Collect more contact information
- Offer avenues “in” to update information or collect follow-up data
- Expand the options of where you look

- New technologies?

D

TAB D

SESSION 2: Challenges in Using Available Data for Small Population Health Research

Presentations in this tab:

The Feasibility of Using Electronic Health Records and Electronic Health Data for Research on Small Populations
Kelly Devers, NORC

Using Geospatial Methods with Demographic Data to Identify Populations
Chris Fowler, Pennsylvania State University

Using Geospatial Methods with Other Health and Environmental Data to Identify Populations
Ellen Cromley, Consultant

The Feasibility of Using Electronic Health Records (EHRs) and Other Electronic Health Data for Research on Small Populations

Kelly J. Devers, Ph.D.

January 18, 2018

NORC at the UNIVERSITY of CHICAGO

Outline

- Importance of Studying Small Populations
- Challenges in Studying Small Populations
- Growing Availability of Electronic Health Record (EHR) and other Electronic Health Data
- Potential Uses of EHR and Other Electronic Health Data
- Future Research
  - Conditions for Greater Use of EHRs and Electronic Health Data
  - Potential Next Steps

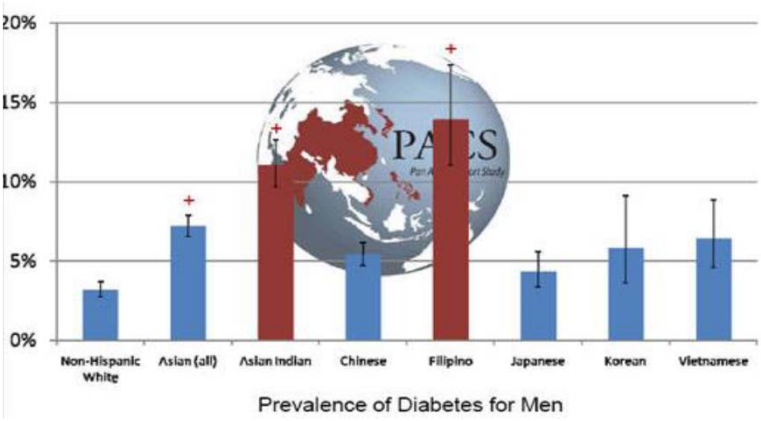
NORC at the UNIVERSITY of CHICAGO

Asian-American Sub-Populations

- Challenges exist in obtaining adequate sample sizes to conduct analysis on Asian Americans overall and for sub-populations
- There also is a lack of consistent race/ethnicity categories used in data collection
- Instances where sub-population analysis has been possible reveal major difference in health

NORC at the UNIVERSITY of CHICAGO

Pan-Asian Cohort Study: Preliminary Findings



Source: Pan Asian Cohort Study. "Preliminary Findings for Diabetes Prevalence." Palo Alto Medical Foundation. Accessed March 1, 2013. http://www.pamf.org/pacs/men.jpg.

NORC at the UNIVERSITY of CHICAGO

Lesbian, Gay, Bisexual, and Transgender Populations

- Many of the health issues and research challenges facing this population are related to stigma
  - Historically caused hesitation in collecting data on LGBT status and has prevented this population from identifying themselves
- Historic lack of standard definitions by which to identify this population through surveys
  - Questions regarding behavior, attraction, and identity all result in different responses and each has important implications for health
- Available research shows differences in needs and disparities in care and outcomes

5

NORC at the UNIVERSITY of CHICAGO

Adolescents with Autism Spectrum Disorders

- Much research has concentrated on diagnosis of these disorders
- But, little is known about health and health care during the transition to adulthood for individuals with ASDs, a critical time for their future well-being
- The cross-sectional nature of most surveys, as well as inconsistency in how disability is measured among children and adults, make it impossible to follow this population over time in most existing survey data

6

NORC at the UNIVERSITY of CHICAGO

Rural Populations

- Geographic isolation and low population density has limited both economic opportunities and access to health services for rural populations
- Rural populations face significant challenges such as the health care needs of aging populations and unique environment health issues
- Variations in how to define the boundaries of rural areas complicates the study of this population
  - E.g., Definitions may not align with county boundaries, the smallest geographic unit used in most surveys

NORC at the UNIVERSITY of CHICAGO

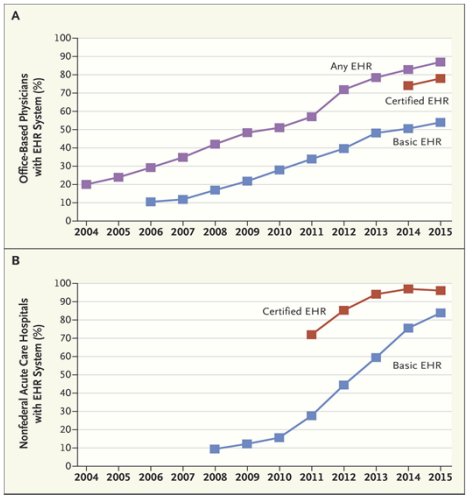
Limitations of National Surveys for Small Populations

Population	General Problem: Small n relative to frame	General Problem: Lack of approaches to increase sample	Frame Problem:* Telephone number frame	Frame Problem:* Area frame samples	Data Collection Problem: Unit nonresponse	Data Collection Problem: Item nonresponse	Data Collection Problem: Instrumen-tation
Asian Americans	X	X		X	X		X
LGBT	X	X				X	X
Adolescents on the autism spectrum	X	X			X	X	X
Rural populations	X	X	X	X	X		X

* These frame problems refer to specific challenges to constructing sampling frames based on telephone numbers or geographic areas. See the "Limitations in Survey Data" section for more information on general problems obtaining an adequate frame for small sample size groups relative to the rest of the population.

NORC at the UNIVERSITY of CHICAGO

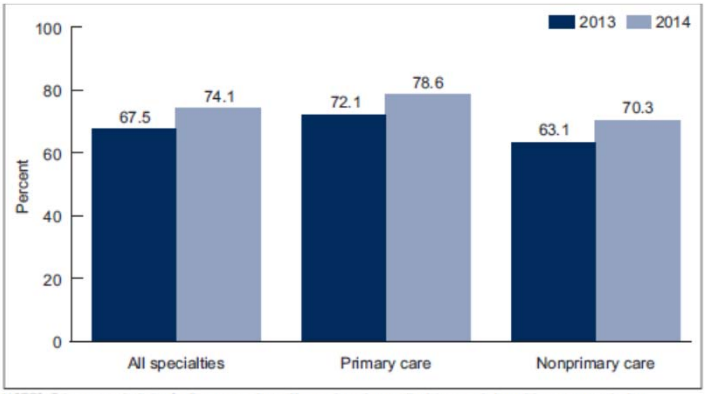
Percent of Office-based Physicians (Panel A) and Acute Care Hospitals (Panel B) with EHR Systems



Source: Washington, V. et al., "The HITECH Era and the Path Forward," N Engl J Med 2017; 377:904-906, September 7, 2017

NORC at the UNIVERSITY of CHICAGO

Office-based Physicians with a Certified Electronic Health Record System, by Physician Specialty: United States, 2013-2014

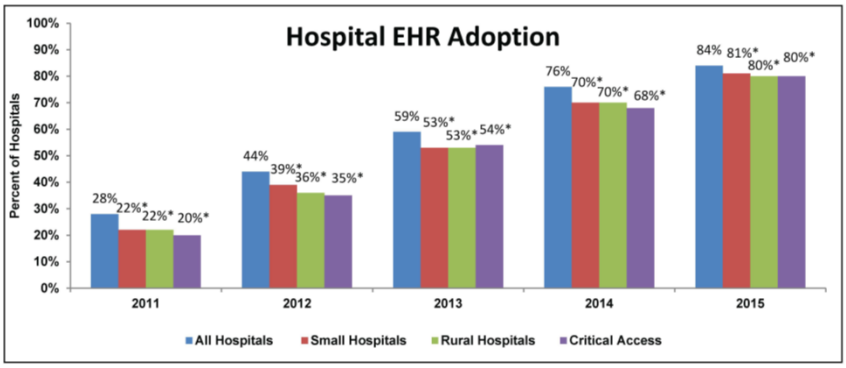


NOTES: Primary care includes family or general practitioners, internists, pediatricians, and obstetricians or gynecologists. Having a certified electronic health record system was defined by physicians answering "yes" to "Does your current system meet meaningful use criteria as defined by the Department of Health and Human Services?" Estimates are based on nonfederal, office-based physicians and exclude radiologists, anesthesiologists, and pathologists. All percentage differences by year and specialty are statistically significant ($p < 0.05$).
SOURCE: CDC/NCHS, National Electronic Health Records Survey, 2013–2014.

NORC at the UNIVERSITY of CHICAGO

At least 8 out of 10 small, rural, and critical access hospitals adopted a Basic EHR

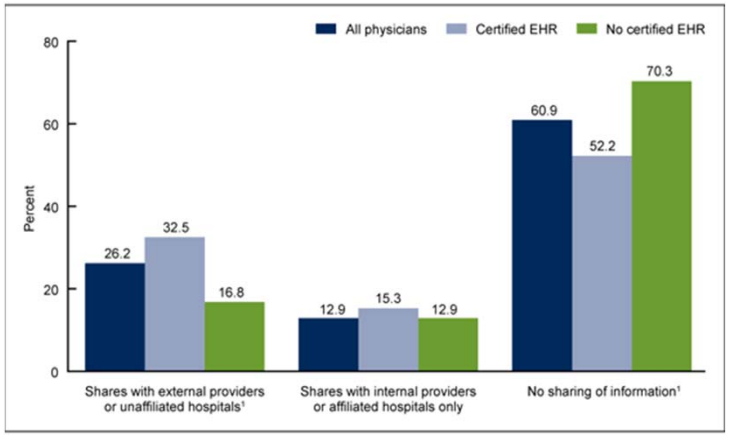
Percent of non-federal acute care hospitals with adoption of at least a Basic EHR system by hospital type



Source: ONC/American Hospital Association (AHA), AHA Annual Survey Information Technology Supplement.
Source: Henry, J. et al. " Adoption of Electronic Health Record Systems among U.S. Non-Federal Acute Care Hospitals: 2008-2015," ONC Data Brief, No. 35, May 2016

NORC at the UNIVERSITY of CHICAGO

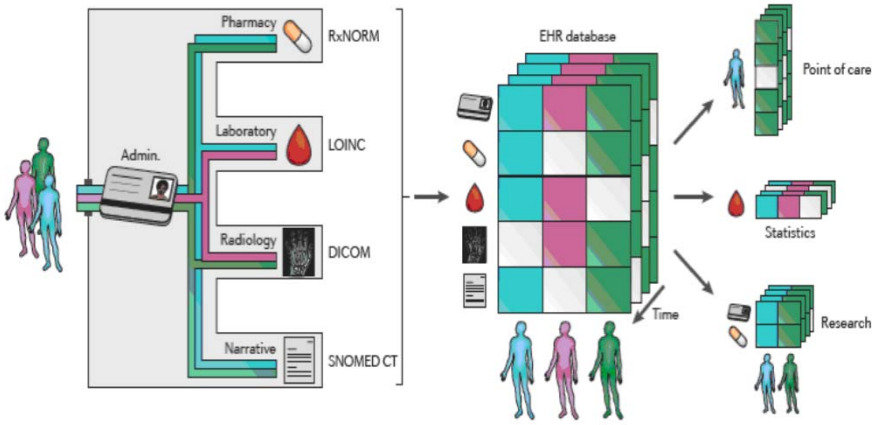
Office-based physicians with EHR systems who shared patient health information electronically with other providers: United States, 2014



SOURCE: CDC/NCHS, National Electronic Health Records Survey, 2014
Jamoom, EW, Yang, N, and E. Hing, "Adoption of certified electronic health record system and electronic information sharing in physician offices: United States, 2013 and 2014. NCHS data brief, no 236, Hyattsville, MD: National Center for Health Statistics, 2016

NORC at the UNIVERSITY of CHICAGO

EHRs and Other Electronic Health Data are Potentially Rich and Powerful Resources to Identify and Study Small Populations



Source: Jensen PB, Jensen LJ, and Brunak S. Mining electronic health records: towards better research applications and clinical care. Nature Reviews, June 2012 (13): 395-403.

NORC at the UNIVERSITY of CHICAGO

Characteristics of EHRs and Other Electronic Health Data That Make Them Useful for Research

- Potential to reach larger samples of individuals, perhaps in some case approaching the majority of the population or sub-populations of interest
- Many types of data including:
  - Claims and administrative data
  - Clinically rich, detailed information
  - Patient reported data
- Ability to identify sub-populations in novel ways
  - E.g. Natural language processing
- Potential to link with other data sources (e.g., surveys)
- Potential longitudinality of some data sets

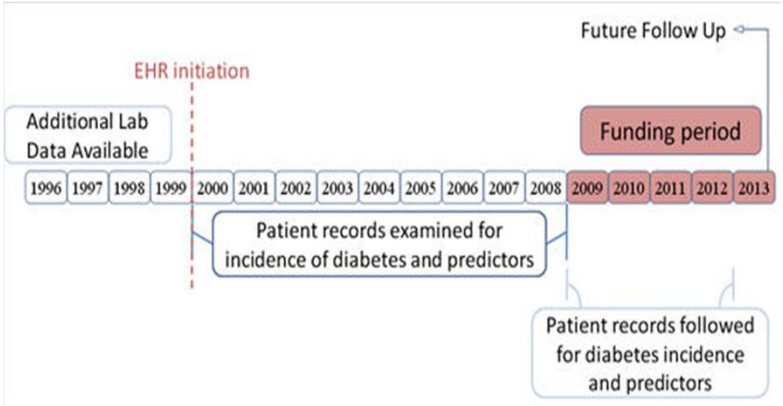
NORC at the UNIVERSITY of CHICAGO

Examples of EHRs and Electronic Health Data to Study Illustrative Populations

- Asian Americans
  - Pan Asian Cohort Study, on which the earlier diabetes results are based, is an EHR based study
  - Kaiser Permanent Northwest collects information about primary language spoken at home as well as need for translation services, and has standardized this variable across health plans so someone could easily look up language sub-groups, such as patients who speak Tagalog
  - At University of Vermont, refugee and immigrant patients have been identified through billing data where interpreters were used

NORC at the UNIVERSITY of CHICAGO

Pan Asian Cohort Study Design and Methods: Virtual EHR Cohort of Asian and White Patients Age 35 and Older



Source: Palo Alto Medical Foundation, Sutter Health, Pan Asian Cohort Study
http://www.pamf.org/pacs/design.html

NORC at the UNIVERSITY of CHICAGO

Examples of EHRs and Electronic Health Data to Study Illustrative Populations

- Lesbian, Gay, Bisexual, and Transgender Populations
  - Vanderbilt University Medical Center found that the time between when patients were first seen and when their LGBT status appeared in their medical records averaged 30 months
  - Now using national language processing (NLP) of unstructured EHR data to identify and analyze information about sexual orientation, gender identity, and sexual behavior
  - Both Vanderbilt and UC-Davis health systems are collecting information about patient's sexual orientation through EHR patient portals as well
  - Stage 3 Meaningful Use certified EHRs are required to add gender identity, sexual orientation capabilities

NORC at the UNIVERSITY of CHICAGO

Examples of EHRs and Electronic Health Data to Study Illustrative Populations

- Adolescents with Autism Spectrum Disorder
  - Kaiser Permanente in Northern California has developed a list of valid autism diagnoses based on ICD codes and who made the diagnosis
  - The EHR sub-network of the Pediatric Research in Office Settings network, known as ePROS, led by the American Academy of Pediatrics
- Rural Populations
  - Kaiser Permanente Northwest studied rural Hispanic patients whose primary language is Spanish, among whom drug seeking behavior has been a particular problem
  - Intermountain Health has studied rural residents with three or more chronic conditions
  - Oregon Community Health Information Network (OHIN), a network of nearly all federally qualified health centers (FQHCs) in the state of Oregon, is also studying drug seeking behavior by those who attempt to obtain opiate-containing drug products from multiple FQHCs and also harness the system for other studies of rural and racial/ethnic sub-populations

NORC at the UNIVERSITY of CHICAGO

Technical Conditions Required for Research Using EHRs and Other Electronic Health Data

- Data extraction and formatting
- Processing free-text data
- Missing data and data quality
- Restricted data
- Legacy systems and longitudinal data
- Expertise

NORC at the UNIVERSITY of CHICAGO

Privacy and Security Conditions Required for Research Using EHRs and Other Electronic Health Data

- Legal landscape
  - HIPPA and the Common Rule
- Opportunities for patients to make meaningful choices
- De-identified data
- Data governance
  - Ownership, control, and regulation

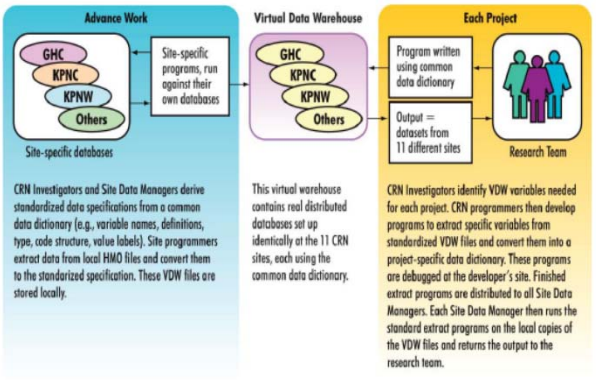
NORC at the UNIVERSITY of CHICAGO

Organizational Conditions Required for Research Combining Multiple Data Sources

- Using EHR and other electronic health data from multiple organizations
- Interoperability of EHR systems
- Research networks
- Regional health information exchanges
- Linking EHR and other electronic health data with other data sources
  - Patient registries
  - Genetic data
  - Other Data Sources, including surveys and claims

NORC at the UNIVERSITY of CHICAGO

Example: The Cancer Research Network (CRN) Virtual Data Warehouse



Source: Hornbrook et al. "Building a Virtual Cancer Research Organization." Journal of the National Cancer Institute Monographs. 2005 (35), 12-25

NORC at the UNIVERSITY of CHICAGO

Potential for Future Research on Small Populations

- Data validation
- New tools and/or methods
- Descriptive studies
- Outcomes research
- Stakeholder engagement and collaboration
- Legal framework and other policy issues

NORC at the UNIVERSITY of CHICAGO

Summary and Conclusion

Many of the conditions required for harnessing the power of EHRs and other electronic data for research on the health and health care needs of the American people, and key small populations, are present or closer to being realized.

While some significant barriers remain, innovative solutions and promising approaches are being developed in the public and private sectors.

We have identified possible suggestions and next steps for moving the field forward.

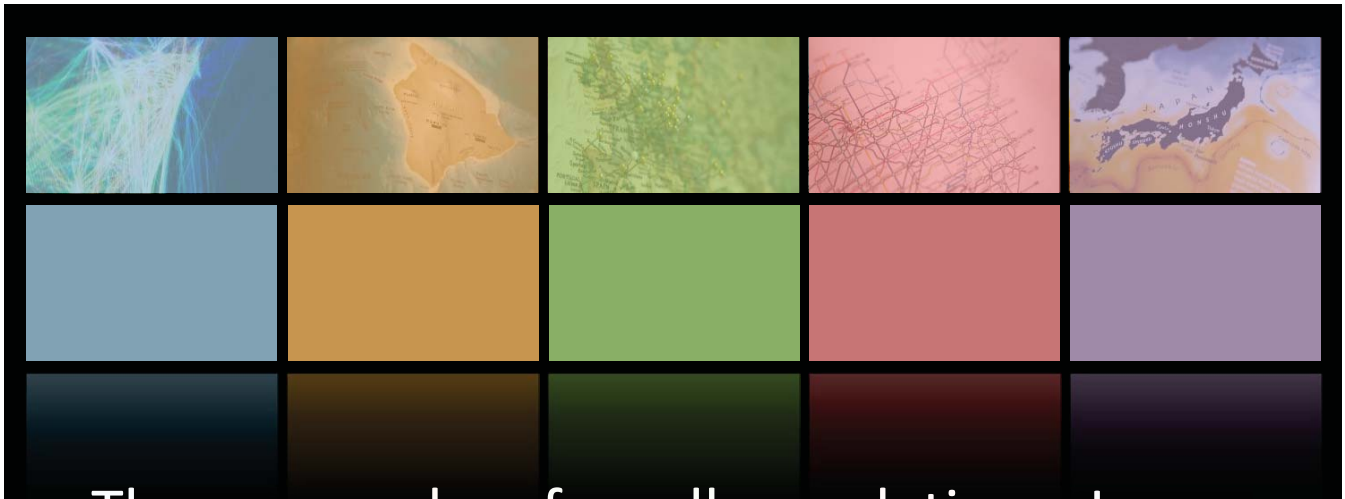
NORC at the UNIVERSITY of CHICAGO

Thank You!

NORC at the UNIVERSITY of CHICAGO

For Further Information Contact:
Kelly J. Devers, Ph.D., M.A.
Senior Fellow
Devers-Kelly@norc.org
301-634-9523

insight for informed decisions™



The geography of small populations: Issues in defining an appropriate geographic context

Christopher S. Fowler
Assistant Professor of Geography and Demography
csfowler@psu.edu

PennState

Contextual variables are a key way that geography is incorporated into health research

Exploring the role of the built and social neighborhood environment in moderating stress and health.

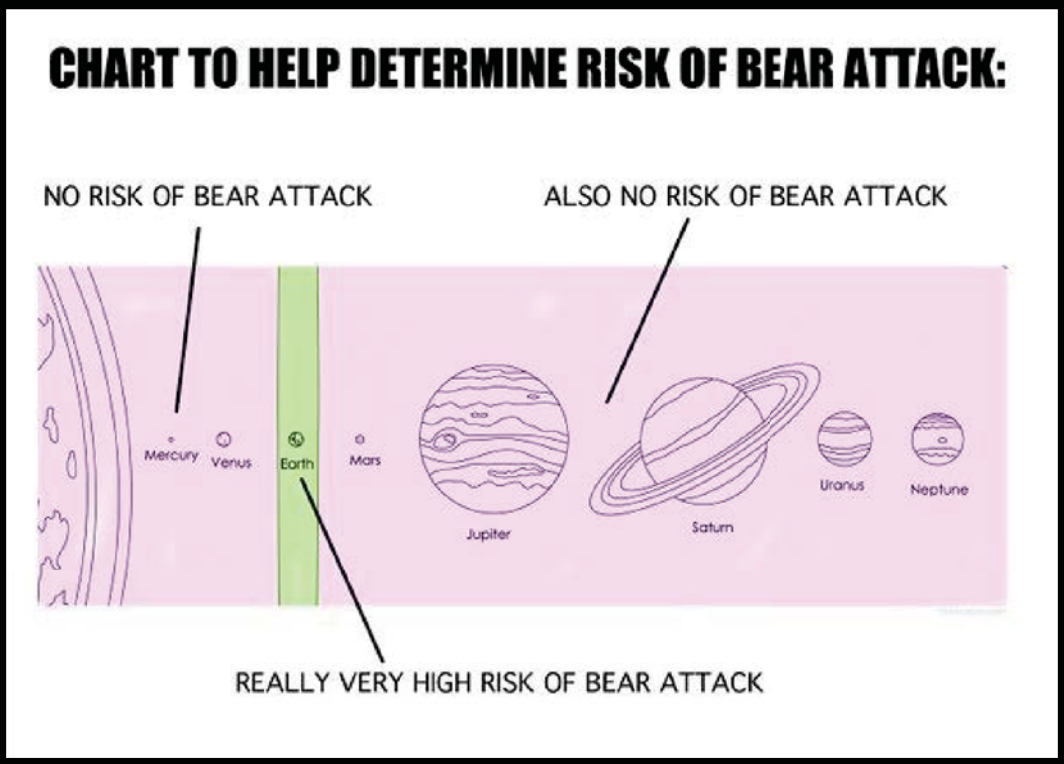
Insomnia and urban neighbourhood contexts – are associations modified by individual social characteristics and change of residence? Results from a population-based study using residential histories

The Impact of Neighborhood Social and Built Environment Factors across the Cancer Continuum: Current Research, Methodologic Considerations, and Future Directions

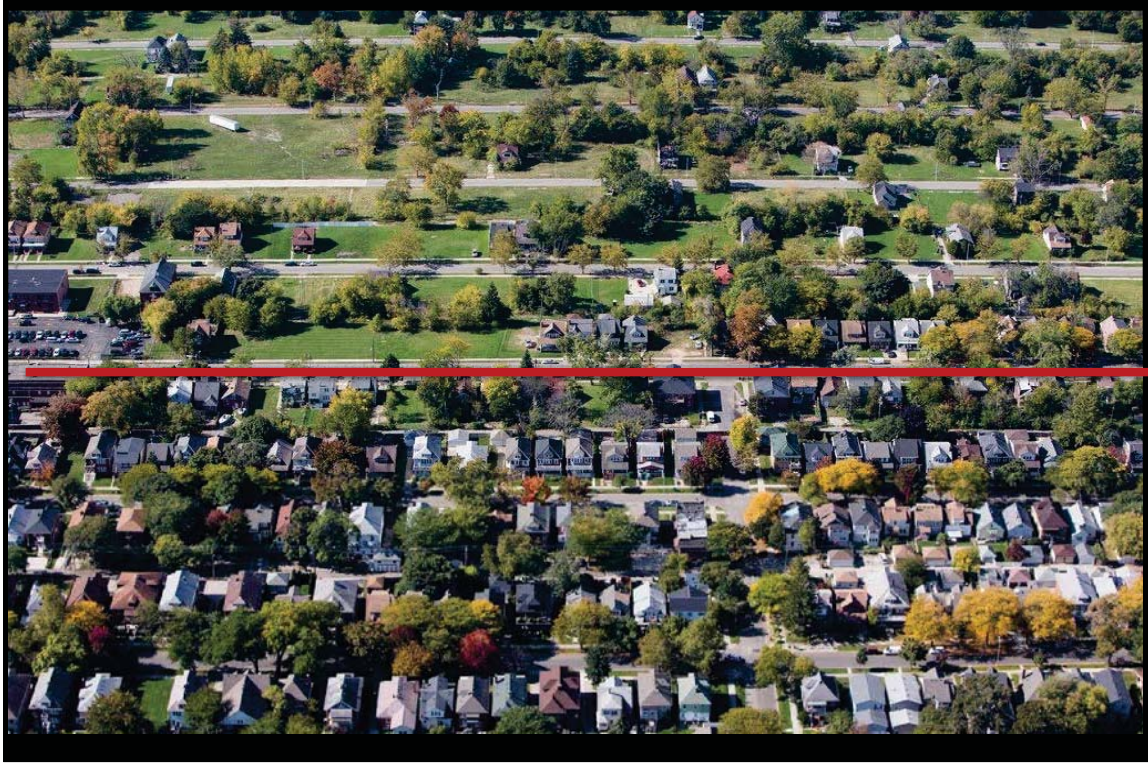
Neighborhood Socioeconomic Status and Substance Use by U.S. Adults

Beyond the Supermarket Solution: Linking Food Deserts, Neighborhood Context, and Everyday Mobility

With contextual measures, it is important to get the SCALE right



Contextual measures assume that BOUNDARIES are meaningful

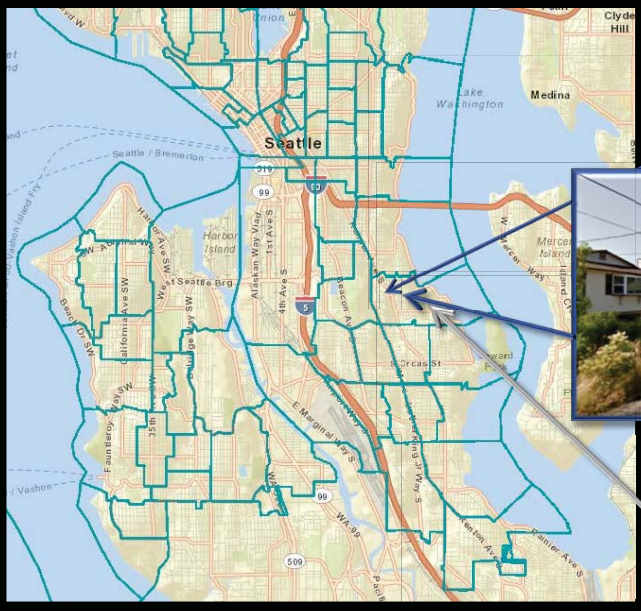


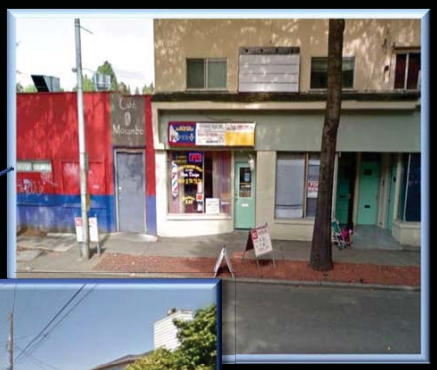
Detroit School District

Grosse Pointe Park School District

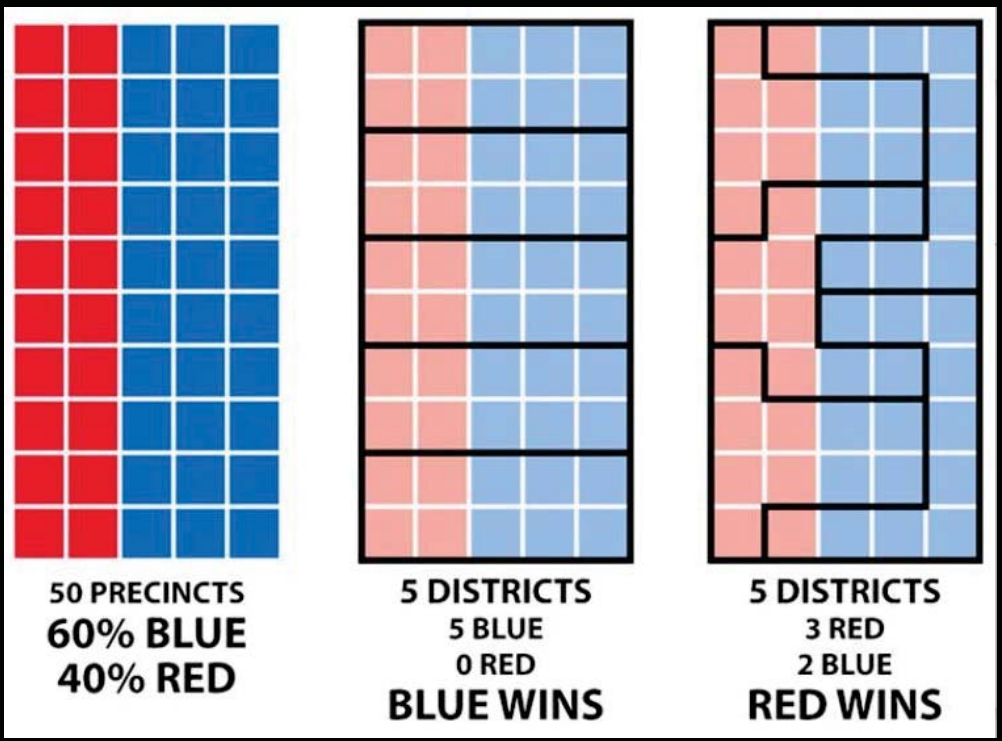
This is not always a reasonable assumption

Census Tracts in Seattle, WA





How we define contextual observations can condition outcomes



Outline

- Motivation
- Why contextual variables may be appropriate for 'small populations'
- Visualizing the effects of scale and boundary choices on contextual variables
- Addressing uncertainty in contextual variables

Contextual variables may be useful when direct access to a population is not possible

Individual test scores or blood lead level may be ideal

…but a lot can be learned from a home address or other locational information available in administrative data

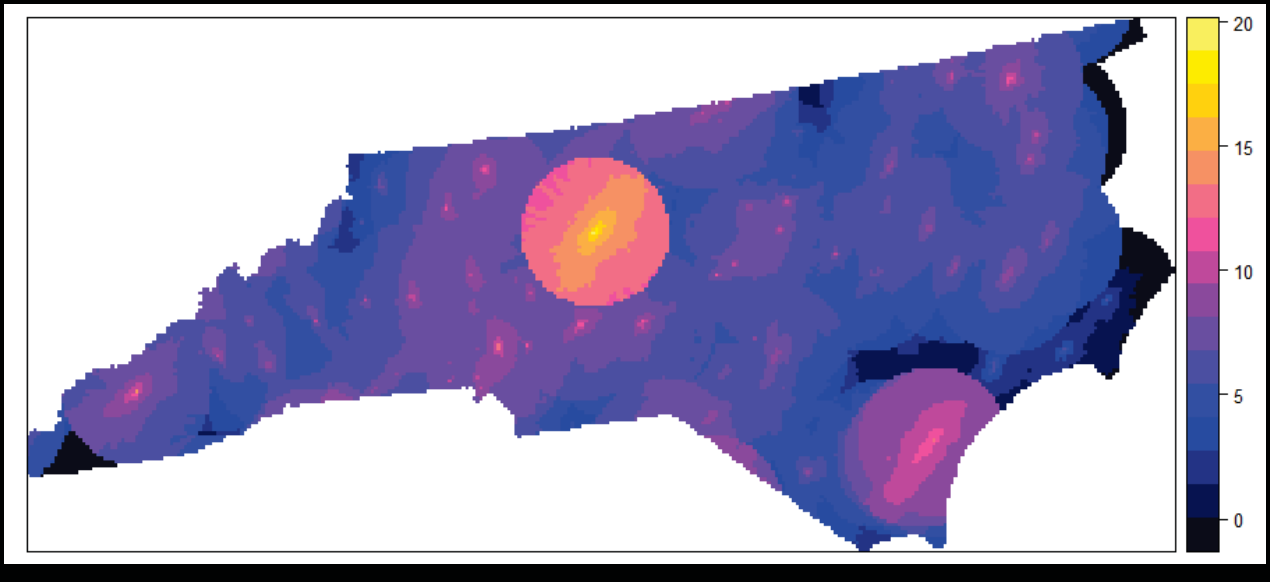
Example Contextual Variables:

- Demographic characteristics (Census)
- Environmental Toxicity (EPA)
- Educational Context (SABINS, Census)
- Crime (NCHS)
- Economy (BLS)

Example: Environmental Toxicity for poor kids in rural places

Airborne chemical toxicity in NC for 2007

800 m^2 grid cells



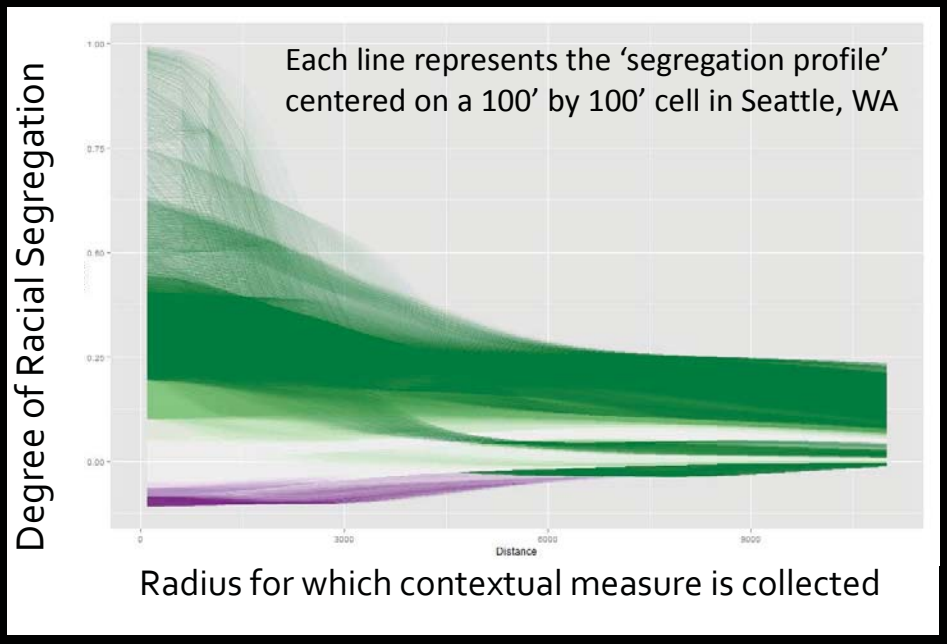
Outline

- Motivation
- Why contextual variables may be appropriate for 'small populations'
- Visualizing the effects of scale and boundary choices on contextual variables
- Addressing uncertainty in contextual variables

Contextual variables need to match the process they expect to evaluate

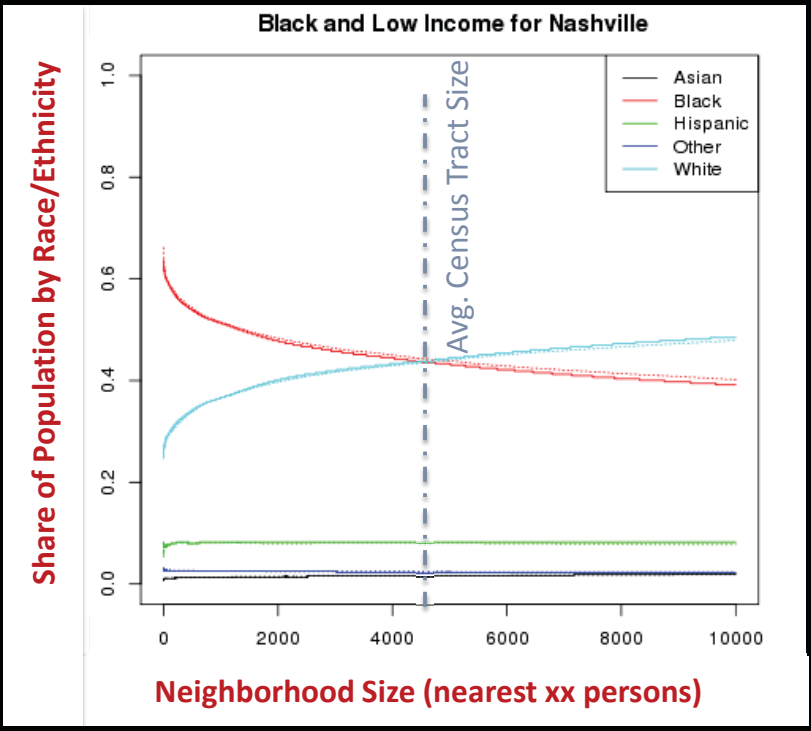
- At small scales (small populations) variability is higher
- At too large scales there is regression to the mean

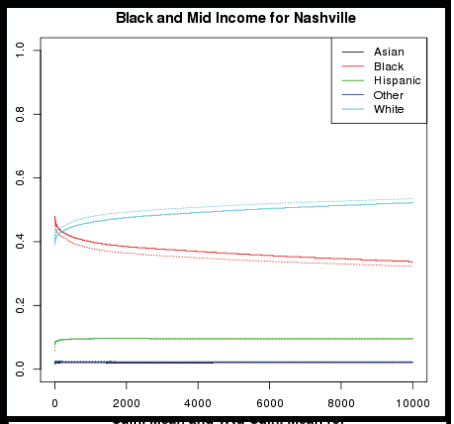
Demographic measures tend to have higher intensity and higher variability at small scales.

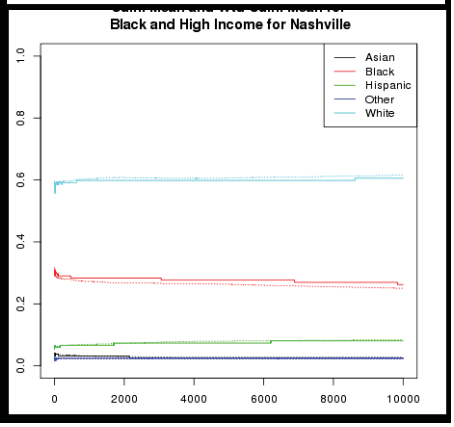


Fowler, Christopher S. (2015) "Segregation as a multiscalar phenomenon and its implications for neighborhood-scale research: the case of South Seattle 1990-2010" *Urban Geography.* ***37*** (1), 1-25.

Context changes a lot by the time we get to units the size of Census tracts



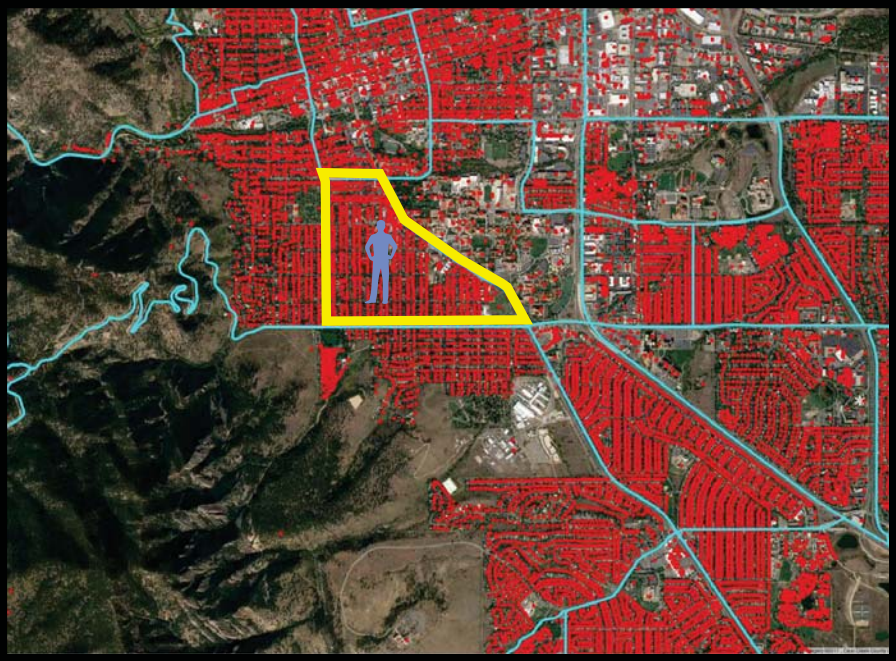




Boundaries vary in quality

- Some research questions have clearly defined geographic boundaries
  - Variation in property tax rates ➔ municipalities, school districts, counties, states, etc.
- Others do not…
  - Rural vs. urban differentials ➔ need to draw the line between "urban" and "rural"
- Other issues
  - Edge effects ➔ equally valid for the household at the center and the one along the border?
  - Scale ➔ Does the size match the social phenomenon being studied?

Boundaries: Often the solution is as simple as mapping the boundaries and determining their suitability visually.



Outline

- Motivation
- Why contextual variables may be appropriate for 'small populations'
- Visualizing the effects of scale and boundary choices on contextual variables
- Addressing uncertainty in contextual variables

Addressing uncertainty in the use of contextual variables

- Problem: How do we know if we have the right scale or the right boundaries?

- Scale Solution: Run the analysis using different scales
  - Blocks, Block Groups, Tracts for Demographic Data
  - e.g. Root, E. D. (2012). Moving neighborhoods and health research forward: using geographic methods to examine the role of spatial scale in neighborhood effects on health. *Annals of the Association of American Geographers*, *102*(5), 986-995

Boundary Solution: How much do boundaries matter for the statistics being calculated

- The tract has 1000 people in it.
- For each of those 1000 people calculate the context based on their 1000 NEAREST NEIGHBORS

Standard Deviation of Individual Context

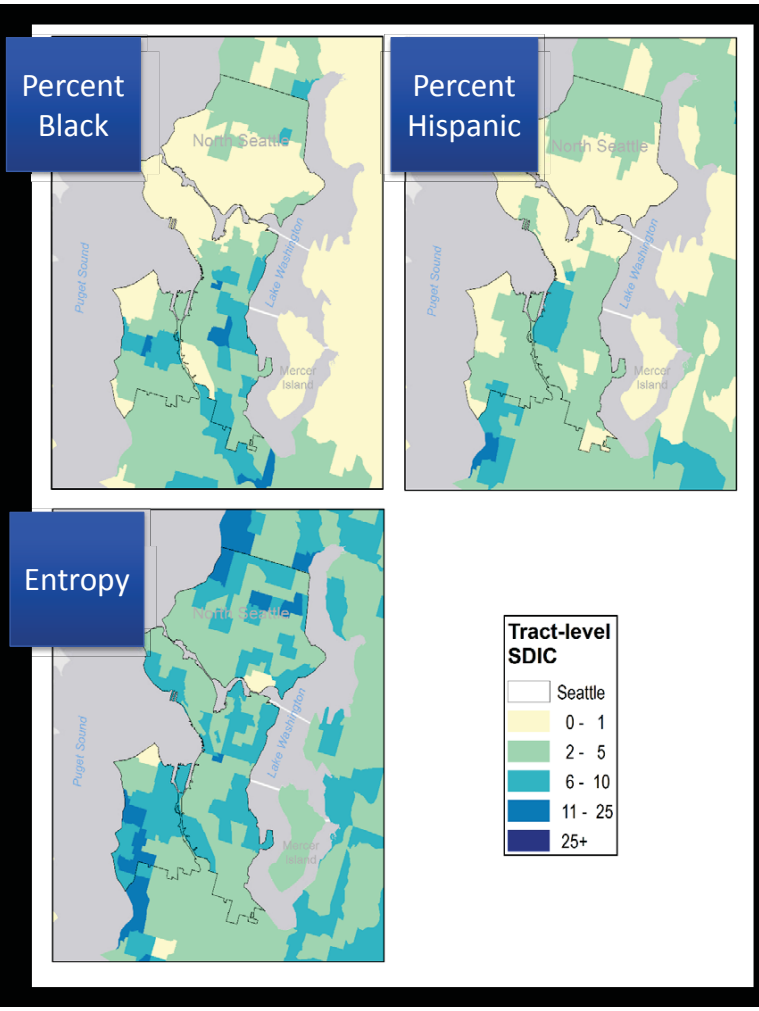
The degree to which individual experience varies within a geographic unit

$$SDIC = \sqrt{\frac{\sum_{i \in c}(x_i - x_c)^2}{k}}$$

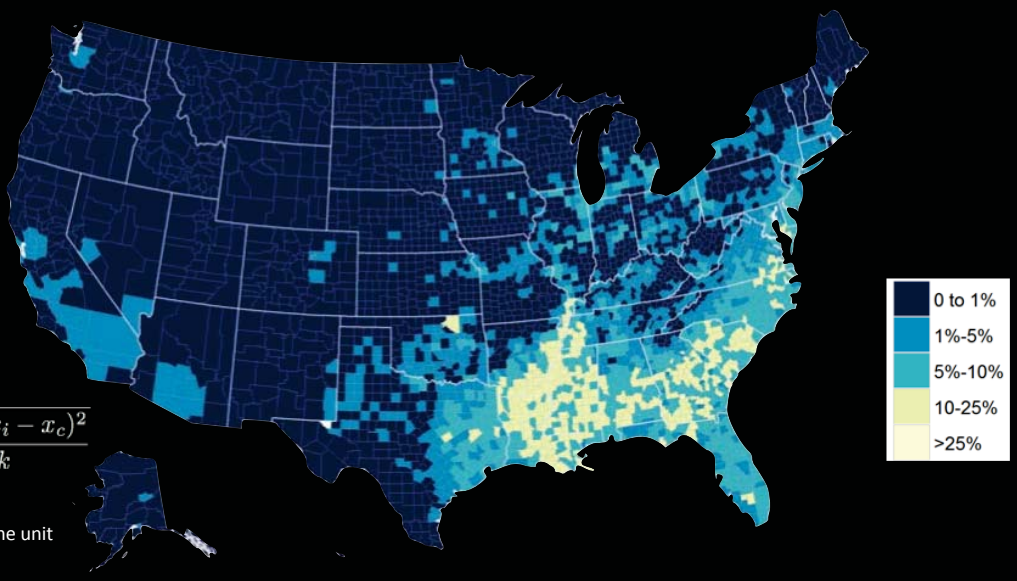
c = contextual unit (like tract)
k= number of people in the unit

Tract-level variation in SDIC for City of Seattle and surrounding area.

'Worst Case' Scenario: Egocentric measures of Percent Black at 0% and 100% within the same tract



The uneven geography of context: County average of tract-level SDIC for Pct. Black



$$SDIC = \sqrt{\frac{\sum_{i \in c}(x_i - x_c)^2}{k}}$$

c = contextual unit (tract)
k= number of people in the unit

Acknowledgements

NIH ([R24-HD041025NSF-Census Research Network: Spatial Science Node. NSF #1132008

David Folch , Nathan Frey, Nicholas Nagle, Seth Spielman

"Early Life Stress and the Environmental Origins of Disease: A Population-based Prospective Longitudinal Study of Children in Rural Poverty." NIH #1UG3OD023332-01

David Folch, Levon Mikaelian, Clancy Blair

Using Geospatial Methods with Other Health and Environmental Data to Identify Populations

Ellen K. Cromley, PhD
Consultant, Health Geographer

ellen.cromley@gmail.com

Purpose and Outline

To illustrate the use of geospatial methods to identify populations from data sources other than electronic health records or the U.S. Census

- The spatial view of data and spatial sampling
- Locating populations--
  - From their places of residence
  - From administrative records
  - From their social networks
  - From their activity locations and activity spaces
  - Taking into account environmental exposure
- Where can we go from here?

Views of Data: Tabular

Tabular View of Two Databases (The Same)

NAME	GRADE	NAME	GRADE
Andy	B	Andy	B
Bob	B	Bob	B
Carmela	C	Carmela	C
Dave	A	Dave	A
Ed	A	Ed	A
Felicia	C	Felicia	C
Gordon	A	Gordon	A
Hank	B	Hank	B
Inez	C	Inez	C

Views of Data: Statistical

Statistical View of Two Databases (The Same)

Grade	Frequency	Grade	Frequency
A	3	A	3
B	3	B	3
C	3	C	3

Views of Data: Spatial

Spatial View of Two Databases (Not the Same)

Teacher		
B	C	A
C	A	B
A	B	C

Teacher		
A	A	A
B	B	B
C	C	C

- The locations of the observations are part of the data record
- The location data are necessary to support spatial data analysis

Implications for Sampling

- Samples in geographic space
  Such samples select observations from a population that is itself geographically distributed

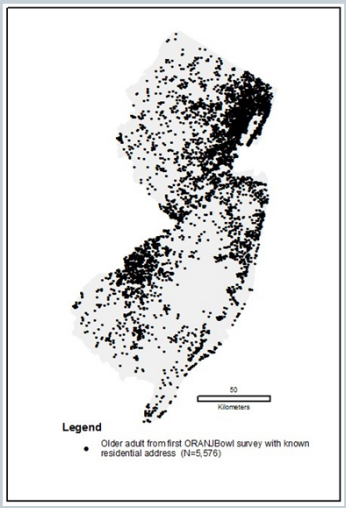
  Every sample of people or other entities located on the earth's surface is implicitly a spatial sample

  A random sample of all people is not a random sample of all places unless people are uniformly distributed

- Geospatial technologies make the spatial basis of evidence explicit so that the sample captures the spatial characteristics of the population

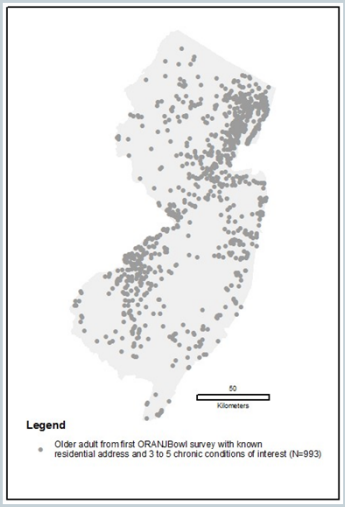
Studies Based on Residential Locations

- Locating older adults with multiple chronic conditions
- ORANJBowl telephone survey rachelpruchno.net/OB.html
- Residential locations of participants from first survey geocoded to census block centroids
- Combinations of three of the following: arthritis, diabetes, heart disease, hypertension, pulmonary disease

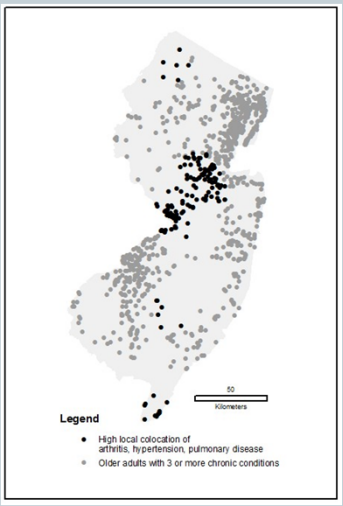


Locating People with Multiple Conditions

- Identified older adults with 3-5 chronic conditions
- Calculated local colocation quotients as a local measure of spatial association of conditions among older adults
- Mapped and tested the significance of any observed areas of spatial association



Spatial Patterns of Multiple Chronic Conditions



- Statewide proportion of older adults with 3-5 chronic conditions including arthritis-hypertension-pulmonary disease was 38 %
- Proportion of highlighted 155 older adults with 3-5 chronic conditions including arthritis-hypertension-pulmonary disease was 50 %

Studies of People in Group Quarters

- People who do not live in housing units reside in group quarters (Census Bureau definition)
- Types
  Institutional—prisons, nursing homes, inpatient mental health facilities
  Non-institutional--college dormitories, military barracks, group homes, shelters
- Size and distribution vary widely in space

Long Term Care Facilities in Massachusetts

- Data from MassGIS
  www.mass.gov/orgs/massgis-bureau-of-geographic-information
- "Global" view of capacity
  N = 740 facilities
  Total beds = 65,272

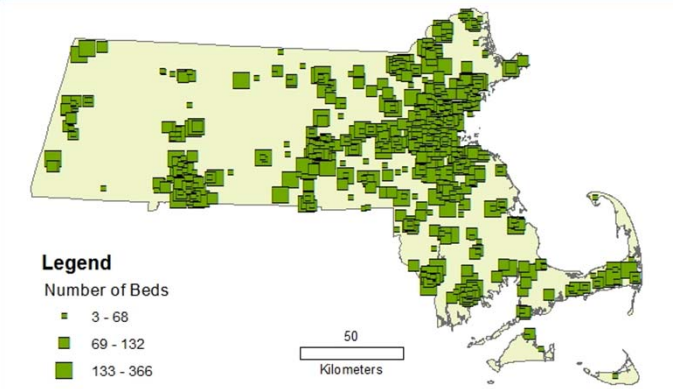
  Mean = 88.2
  Median = 83
  Min = 3, Max = 366
  Standard deviation = 52.4

Spatial Distribution of Facilities by Size

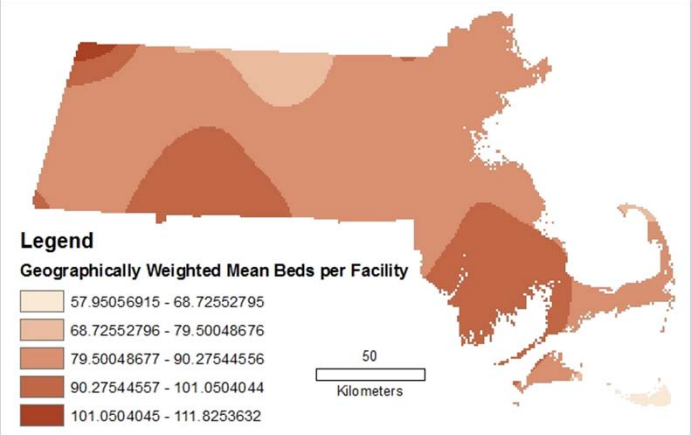
- "Local" view of capacity

Is facility size important in sampling?

Is capacity the same in every region?



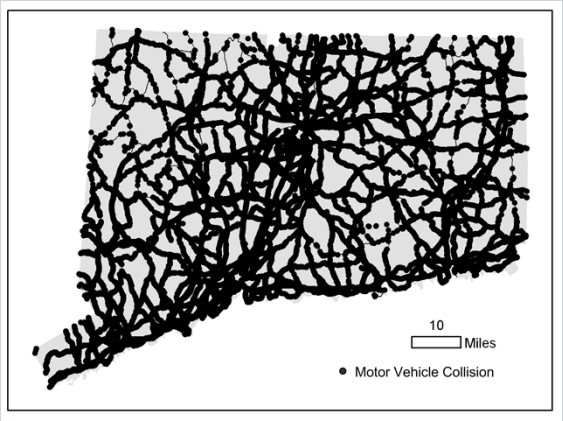
Geographically Weighted Mean Capacities



Studies Based on Administrative Records

- Definition (Statistics Canada | Statistique Canada)
  Data collected for the purpose of carrying out various non-statistical programs
- Examples
  Vital records (births and deaths)
  Registries (immunization records, tumor registries)
  Health and social services records
  Public safety reports (collisions, shootings, fires)
- Administrative record linkage
- Data privacy and confidentiality

Connecticut CODES Project



- Police accident reports from state DOT
- Map shows collisions on federal and state roads 1995-1996 (N=124,053)
- Where to target an intervention to reduce fixed-object collisions?

Identifying Collisions in Local Areas

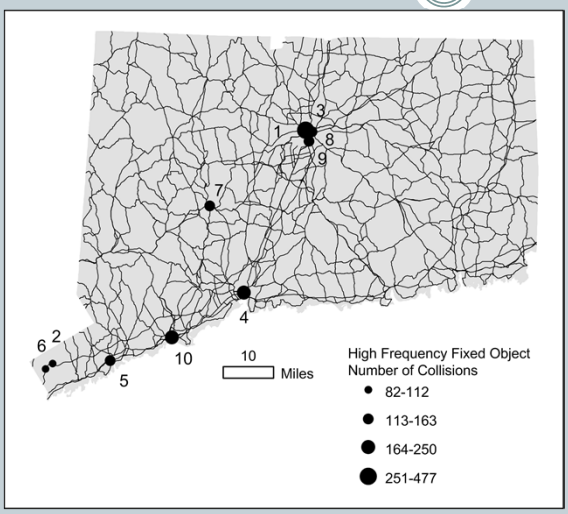
- **Geospatial methods**

Used a box-shaped kernel based on stopping distance to group collisions along road segments

Identified places with high numbers of collisions by type of collision

Eliminated overlapping "places"

Fixed Object Collision Places



- 10 places with highest frequency of fixed object collisions
- Geographically distinct

Local Proportions

Place ID	Fixed Object	Rain or Snow	Dry Road	Daylight	Age 25-44	Male	Too Fast for Conditions
1	**0.29 z=5.23**	**0.46 z=11.19**	**0.46 z=-8.86**	0.71 z=.65	**0.52 z=2.82**	0.60 z=-.09	**0.27 z=9.75**
2	**0.91 z=16.48**	**0.66 z=8.87**	**0.10 z=-10.55**	0.76 z=-.86	0.51 z=1.16	0.58 z=1.43	**0.62 z=13.61**
3	**0.55 z=9.75**	**.39 z=3.82**	**0.48 z=-3.99**	0.66 z=-.86	0.51 z=1.25	0.68 z=-.61	**0.62 z=9.87**
4	**0.29 z=3.48**	0.25 z=.38	0.65 z=-.18	0.67 z=-1.02	0.49 z=1.10	0.58 z=-.73	**0.26 z=5.81**
5	**0.40 z=6.65**	**0.40 z=4.61**	**0.48 z=-4.73**	**0.52 z=-4.78**	**0.54 z=2.24**	0.61 z=.14	**0.29 z=6.36**
6	**0.70 z=11.61**	**0.69 z=9.52**	**0.20 z=-8.56**	0.78 z=1.71	0.54 z=1.58	0.54 z=-1.18	**0.57 z=12.16**
7	**0.37 z=5.48**	**0.38 z=3.95**	**0.57 z=-2.14**	0.73 z=.84	0.48 z=.69	0.62 z=.26	**0.31 z=6.93**
8	**0.35 z=4.82**	**0.39 z=4.42**	**0.49 z=-4.28**	0.71 z=.37	0.52 z=1.52	**0.52 z=-2.32**	**0.29 z=6.27**
9	**0.44 z=6.78**	**0.39 z=3.83**	**0.54 z=-2.62**	0.64 z=-1.45	0.51 z=1.17	00.65 z=.90	**0.27 z=5.03**
10	0.22 z=1.09	0.29 z=1.75	0.62 z=-1.01	0.67 z=-.89	0.51 z=1.58	0.66 z=1.81	**0.19 z=2.84**
State	0.19	0.24	0.65	0.70	0.46	0.61	0.12

Local Odds Ratios

Place ID	Rain or Snow	Dry Road	Daylight	Age 25-44	Male	Too Fast for Conditions
1	**2.79 z=4.91**	**0.29 z=-5.58**	**0.45 z=-3.76**	0.73 z=-1.56	0.87 z=-.69	**11.09 z=10.08**
2	2.83 z=1.30	0.22 z=-1.63	1.27 z=.27	0.00 z=.00	1.70 z=.66	4.71 z=1.78
3	**4.80 z=3.75**	**0.17 z=-4.46**	**0.36 z=-2.52**	1.34 z=.80	0.56 z=-1.54	**17.50 z=5.65**
4	**2.89 z=3.21**	**0.19 z=-5.16**	**0.30 z=-3.82**	1.25 z=.73	0.63 z=-1.53	**8.62 z=6.14**
5	**3.06 z=3.33**	**.28 z=-3.74**	**.48 z=-2.27**	1.32 z=.87	1.03 z=.09	**3.11 z=3.17**
6	1.05 z=.09	0.54 z=-1.10	0.23 z=-1.86	1.14 z=.27	1.14 z=.27	**2.64 z=1.98**
7	**5.53 z=4.65**	**0.16 z=-4.87**	1.03 z=.08	1.24 z=.65	**.36 z=-2.90**	**16.55 z=6.41**
8	**3.04 z=3.21**	**0.32 z=-3.27**	**0.35 z=-.292**	1.20 z=.55	1.20 z=.55	**10.50 z=5.81**
9	**2.33 z=2.25**	**0.29 z=-3.28**	**0.46 z=-2.02**	0.83 z=-.52	1.02 z=.06	**3.32 z=2.85**
10	0.94 z=-.19	0.78 z=-.81	0.32 z=-3.59	0.80 z=-.72	0.92 z=-.25	**3.86 z=3.81**
State Odds Ratio	1.76	0.50	0.40	1.13	1.10	6.38

Error in Administrative Records

- Challenges
  Errors due to inclusion criteria
  Errors due to coding of thematic attributes
  Errors in coding spatial attributes

- Responses to challenges
  Talk to the people who collect and code the data
  Engage with agencies to improve data quality
  "Adopt robust methods of analysis"–
  Waldo Tobler, Analytic Cartographer,
  Member of the National Academy of Sciences

Studies Based on Social Networks

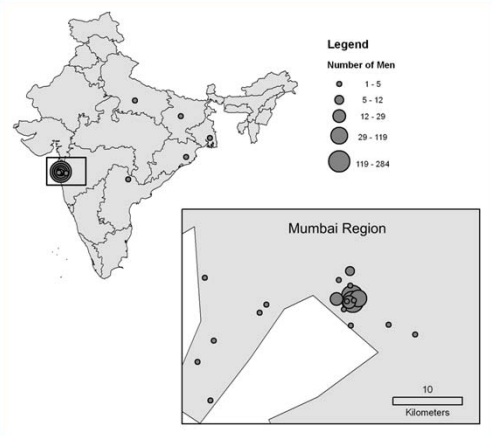
- Social networks function in geographic space and they can be mapped
- In respondent-drive sampling, examine the distribution of locations for recruiting "seed" participants
- Include mechanisms for monitoring the spatial locations of network members recruited into the research study

Studies Based on Activity Sites

- Places outside the home where people engage in activities can be used for locating populations
  - Health service delivery sites
  - Workplaces and schools
  - Bars
  - ....and many more
- Geospatial methods can be used to map these venues to show their spatial distribution
- A critical consideration in venue-based research is understanding who is seen at these venues

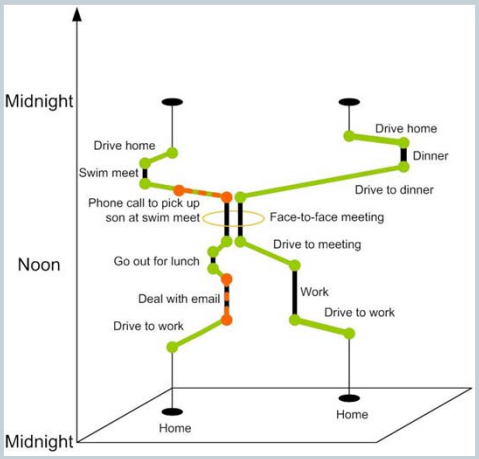
The ASHRA Project

- Study of alcohol and sexual risk in three low-income communities in Mumbai, India
- 640 drank with friends at places in one of the three study communities
- 111 drank in other widely-dispersed locations in the Mumbai region and elsewhere



Locating People Who Share Activity Spaces

- Home base, activity sites, routes of travel
- Various ways we can locate populations for health research are represented in this graphic
- With geo-enabled devices, the information can be collected continuously in real time



Grouping People Based on Activity Sites

- Make pair-wise comparisons of study participants
- Calculate an index of association based on co-location in space or in time and space

$$I = \frac{2 \times (\text{Number of activity sites in common})}{\text{Number of sites Person 1} + \text{Number of sites Person 2}}$$

- Analyze resulting similarity measures by using scaling techniques to group people with similar activity patterns in space or in time and space

Exposure Studies

- Some epidemiological study designs require information on exposure (cohort and case-control studies)
- Dominant view of what geospatial methods contribute to health research
  - Map air quality made by taking samples of space
- Geospatial technology can be used to assess exposure without reference to such "maps"
  - Personal exposure monitoring
  - Monitoring environmental conditions at residences, other activity sites, or travel routes and assigning measurements to study subjects
- Contextual analyses can be "geographic" but not "spatial"

Where Can We Go from Here?

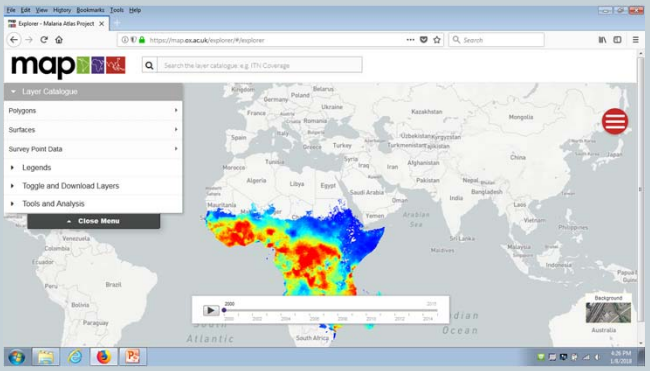
- The key impact of geospatial technologies has been to enable us to study large areas at high levels of detail (maintaining individual locations and attributes)
- Building spatial data commons
- Adopting a spatial analytic framework for health science explicitly addressing spatially varying processes

Building Data Commons

- The Malaria Atlas Project aims to disseminate free, accurate and up-to-date information on malaria and associated topics, organized on a geographical basis (map.ox.ac.uk/)

Open access policies

Subscribes to GATHER



Putting it All Together: Matlab

- Uses geospatial technology to determine
  - How cholera vaccine efficacy varies spatially in the study area
  - What ecological socio-environmental variables are related to cholera vaccine efficacy (which variables are effect modifiers?)
  - How protective efficacy varies with access to treatment facilities (is access a spatial confounder?)
  - Whether cholera incidence in the placebo group is related to vaccine coverage rates (is herd immunity important?)
- Currently integrating social networks into a spatial analytic framework for vaccine trial evaluation
- Cholera vaccine efficacy as a spatially-varying process

Questions?

Using Geospatial Methods with Other Health and Environmental Data to Identify Populations
Ellen Cromley

References

Cromley, EK, (2007), Risk Factors Contributing to Motor Vehicle Collisions in an Environment of Uncertainty. *Stochastic Environmental Research and Risk Assessment (SERRA)*, 21;5:473-486. doi: 10.1007/s00477-007-0130-5

Cromley EK, et al. (2010). Spatial Dimensions of Research on Alcohol and Sexual Risk: A Case Example from a Mumbai Study. *AIDS & Behavior*, 14;S1:S104-S112.

Cromley EK, et al., (2015), Colocation of Older Adults with Successful Aging Based on Objective and Subjective Measures. *Journal of Applied Geography*, 56;1:13-20. doi: 10.1016/j.apgeog.2014.10.003

Cromley EK, et al., (2016), Spatial associations of multiple chronic conditions among older adults, *Journal of Applied Gerontology*, 1-25. doi: 10.1177/0733464716672044.

Fotheringham, A. Stewart, Chris Brunsdon, and Martin Charlton (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester, England: John Wiley & Sons Ltd. (Software for performing GWR analyses is available at gwr.maynoothuniversity.ie/gwr4-software/.)

Hägerstrand, T, (1970), What about People in Regional Science? *Papers in Regional Science*, 24;1:6-21

Jutte DP, et al., (2011), Administrative Records Linkage as aTool for Public Health Research, *Annual Review of Public Health*, 32:91-108. doi:10.1146/annurev-publhealth-031210-100700

Magnani, R., et al. (2005), Review of sampling hard-to-reach and hidden populations for HIV surveillance. AIDS, 19;suppl2:S67- S72

Olsen SJ, et al., (2003), Transmission of the Severe Acute Respiratory Syndrome on Aircraft, *New England Journal of Medicine*, 349;25:2416-2422;

Pierce, JR, et al., (2006), Place-of-residence errors on death certificates for two contiguous U.S. counties, *Population Health Metrics*, 4:6. doi: 10.1186/1478-7954-4-6

Radil, SM, et al., (2010), Spatializing Social Networks: Using Social Network Analysis to Investigate Geographies of Gang Rivalry, Territoriality, and Violence in Los Angeles, *Annals of the Association of American Geographers*, 100;2:307-326. doi: 0.1080/00045600903550428

Stevens GA, et al., (2016). Guidelines for Accurate and Transparent Health Estimates Reporting: the GATHER statement. *PLOS Medicine*, 13;8:e1002116. doi: 10/1371/journal.pmed.1002056

Troped PJ, et al., (2014), Relationships between the Built Environment and Walking and Weight Status among Older Women in three U.S. States, *Journal of Aging and Physical Activity*, 22;1:114-25. doi: 10.1123/japa.2012-0137.]

Wang, et al., (2012), A Review of Spatial Sampling, *Spatial Statistics*, 2;1-14. doi: 10.1016/ j.spasta.2012.08.001

Links

In December, 2017, *The Guardian* published a very interesting piece titled “Bussed Out” on programs to bus homeless people from one city to another, complete with animated maps (www.theguardian.com/us-news/ng-interactive/2017/dec/20/bussed-outamerica-moves-homeless-people-country-study).

spatial health web site at the University of North Carolina (spatialhealth.web.unc.edu/projects/presentprojects/incorporating-geographic-context-into-randomized-controlled-trials-case-studieson-the-rtss-malaria-and-the-oral-cholera-vaccines/)

E

TAB E

SESSION 3: Techniques Used in Survey Research to Identify and Find Small Populations for Health Research

Presentations in this tab:

Two Applications of Respondent Driven Sampling: Ethnic Minorities and Illicit Substance Users
Sunghee Lee, University of Michigan

Venue-Based and On-line Sampling
Patrick Sullivan, Emory University

Two Applications of Respondent Driven Sampling: Ethnic Minorities and Illicit Substance Users

Sunghee Lee, Ai Rene Ong, Michael Elliott
University of Michigan

Workshop on Improving Health Research for Small Populations
National Academy of Sciences, Engineering and Medicine
January 18, 2018

Introduction

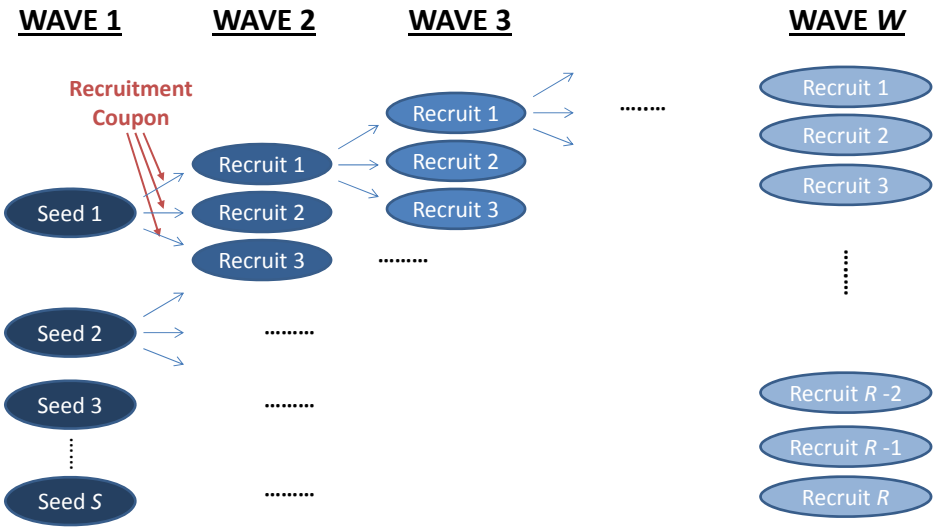
Respondent Driven Sampling – 1

- Growing interest in studying hard-to-reach, rare, elusive, hidden populations
  - HIV at-risk population: MSMs, Sex workers, IDUs
  - LGBT populations
  - Recent immigrants
- No clear and practical solution with probability sampling
  - High screening costs
  - Hesitant to be identified

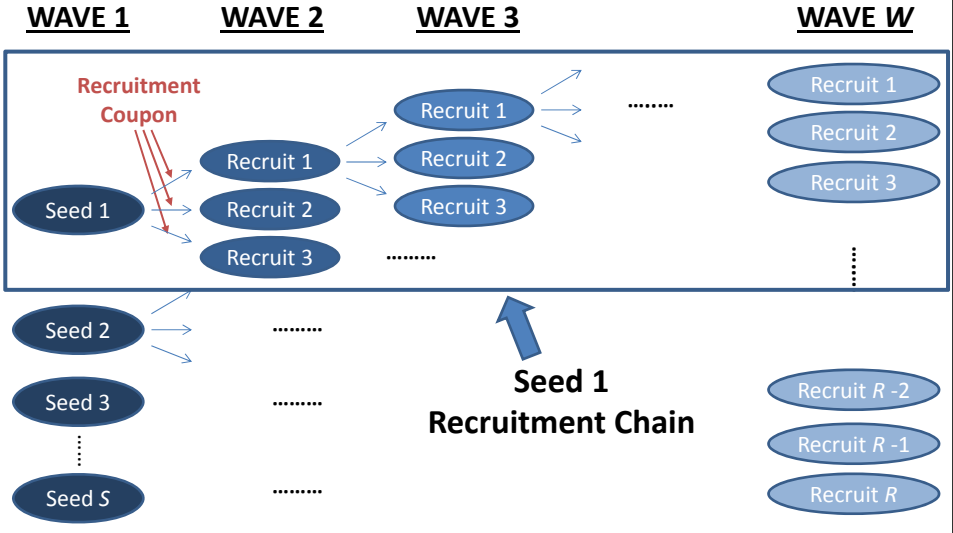
Respondent Driven Sampling – 2

- Proposed by Heckathorn (1997, 2002)
- Popular usage in public health
- Exploits social networks among rare population members for sampling purposes
  - Sampled members also play a role of a recruiter
  - Incentivized recruitment from own network through coupons and this continues in waves/chains
  - Recruitment assumed to be random within each individual's network and to follow memory-less Markov chain and reach equilibrium
    - Under these assumptions, unbiased estimators can be obtained after equilibrium using weights equal to the number of nodes for a subject's recruiter.

Respondent Driven Sampling – 3



Respondent Driven Sampling – 4



Network Sampling vs. RDS

Similar:

- Rely on social networks

Different:

- Network specification
  - NS: biological siblings, immediate family members
  - RDS: jazz musicians
- Who selects the sample
  - NS: researchers
  - RDS: study participants

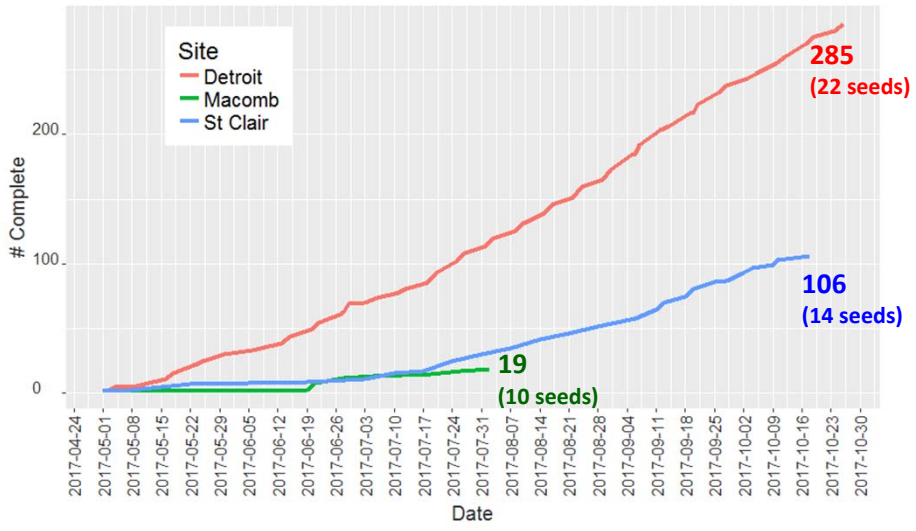
Application 1:
Project PATH (Positive Attitudes Towards Health)

Funded by the National Science Foundation (GRANT NUMBER SES-1461470)

PATH Data Collection

- Injection drug users in Southeast Michigan
- Phone screener
  → In-person screener + Main interview + ~3 Coupons
  → In-person follow-up interview
- Data collection sites
  - Detroit: Urban; Tues, Thur @ Detroit Center
  - Macomb: Suburban; Weds @ County PH Depart
  - St. Clair: Rural; Mon (+Weds) @ County PH Depart
  - 4 interviewers rotating between sites
- Field Period: 5/1/2017 – 10/31/2017

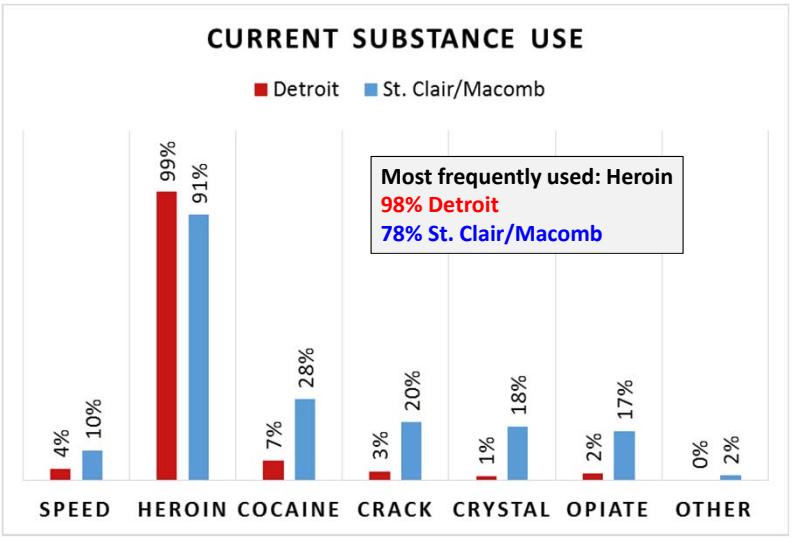
PATH Data Collection Progress



Demographics

	Detroit	St Clair/ Macomb
Age (avg)	56 yrs	40 yrs
Age: <30 years old	2%	32%
Male	68%	53%
Non-Hispanic White	11%	73%
Non-Hispanic Black	81%	16%
Education: <High School	32%	18%
Employed	8%	18%
Ever homeless past 12 mos	40%	56%

Substance Use



Application 2:
Health and Life Study of Koreans (HLSK)

Funded by the National Science Foundation (GRANT NUMBER SES-1461470)

HLSK

- Targets foreign-born Korean American adults in
  - Los Angeles County
  - State of Michigan
- Web-RDS survey
  http://sites.lsa.umich.edu/korean-healthlife-study/
  - Unique number required for participation
  - Incentive payment through checks
- Target n=800 (currently ~600)
- Benchmarks from American Community Survey

HLSK Formative Research

- 3 rounds of focus group discussions
  - ~30 participants; 2 rounds in Korean and 1 in English
  - Discussion focused on
    - Web surveys
      → URL, Web site contents, etc.
    - Concept of RDS
    - Coupons
      → Up to 2 coupons
      → "Expire" in 2 weeks
    - Level of incentives
      → $20 for main, $5 for follow-up, $0 for recruitment

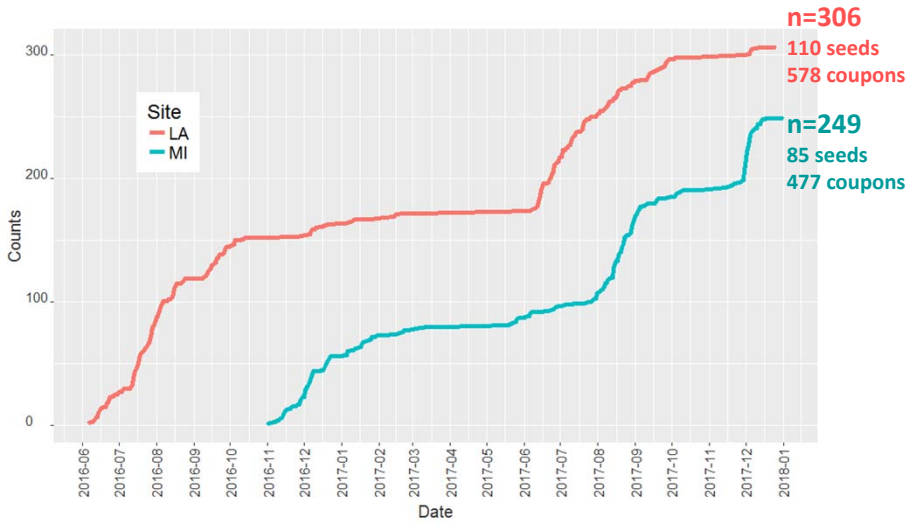
HLSK Data Collection

- Started with 12 seeds in LA in June 2016
- MI added in November 2016

- LA seeds (initially)
  - Recruited through referral
  - Balanced on gender, age, dominant language
  - In-person introduction about the study

→ It became clear the protocols would not work
  - Provide recruitment incentives
  - Add more seeds

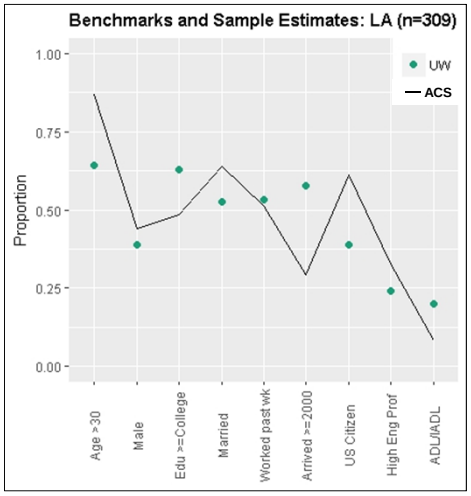
HLSK Data Collection Progress

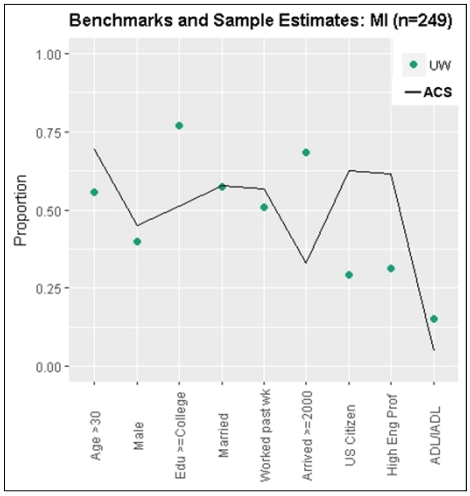


HLSK vs. ACS – 1

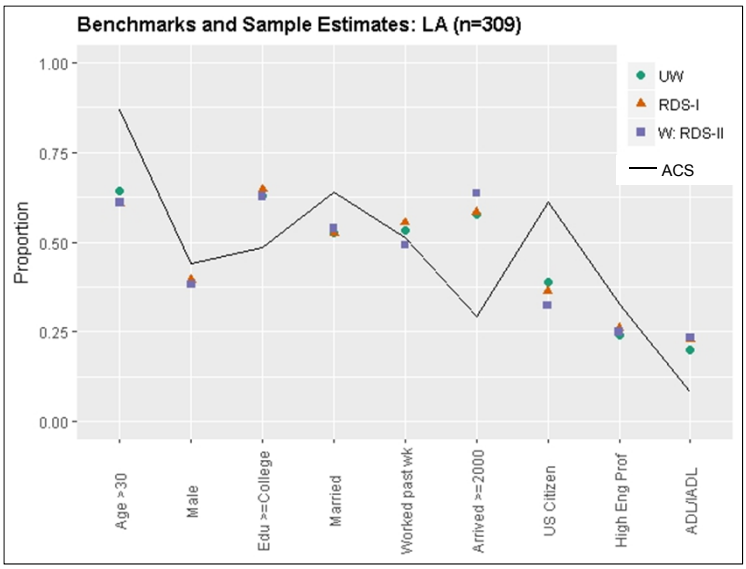
- American Community Survey 2011-2015 data
- HLSK sample estimates
  - Unweighted (UW)
  - RDS-I
  - Weighted: RDS-II
  - Weighted: Post-stratification (PS) by age, sex, educ
  - Weighted: RDS-II + PS

HLSK vs. ACS – 2

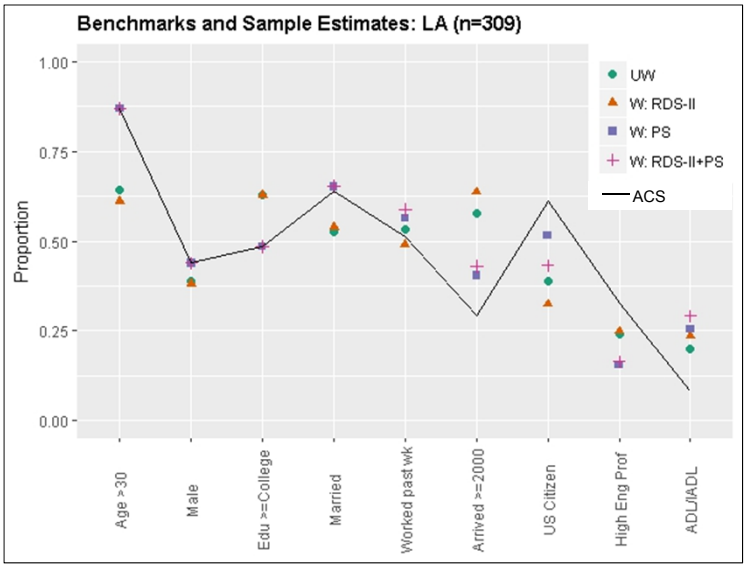




HLSK vs. ACS – 3



HLSK vs. ACS – 4



Summary

What did we learn?

- Non-cooperation is an issue for generating long chains (memorylessness unlikely)
- Had to improvise to make RDS "work"
- Sample size (hence, chain length) is a random variable affected by many (mostly unknown) factors
- Inferences limited
- YET, difficult-to sample groups can be recruited
  - E.g., highly-educated young recent immigrants

Where should we go?

- Non-cooperation is critical for
  - meeting theoretical assumptions (hence, inferences)
  - study design
  - replications of the same study
- Yet to be addressed in the literature and accounted for in inferences

Thank you

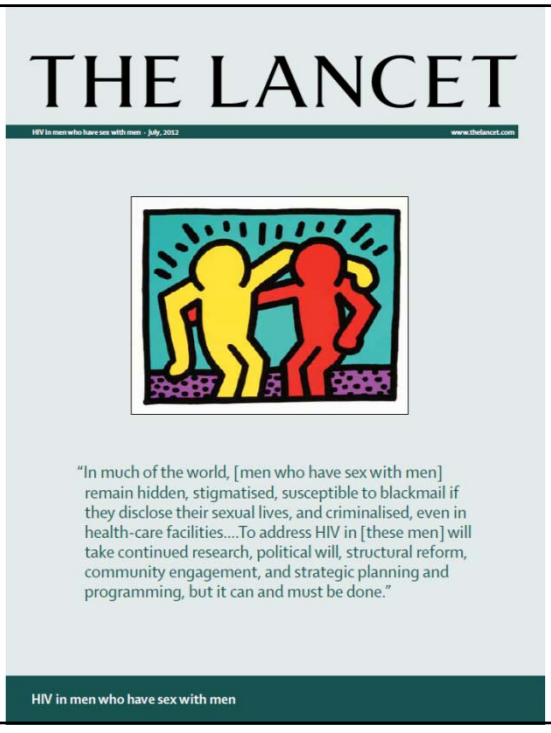
sungheel@umich.edu

Venue-Based and On-line Sampling

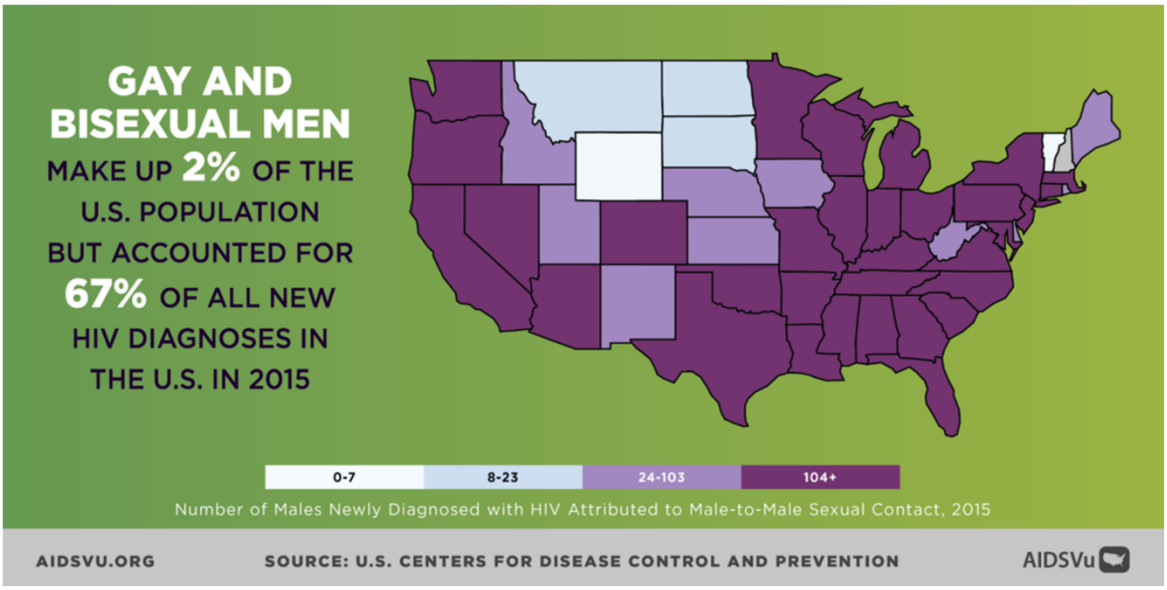
Patrick S. Sullivan, DVM, PhD
Department of Epidemiology

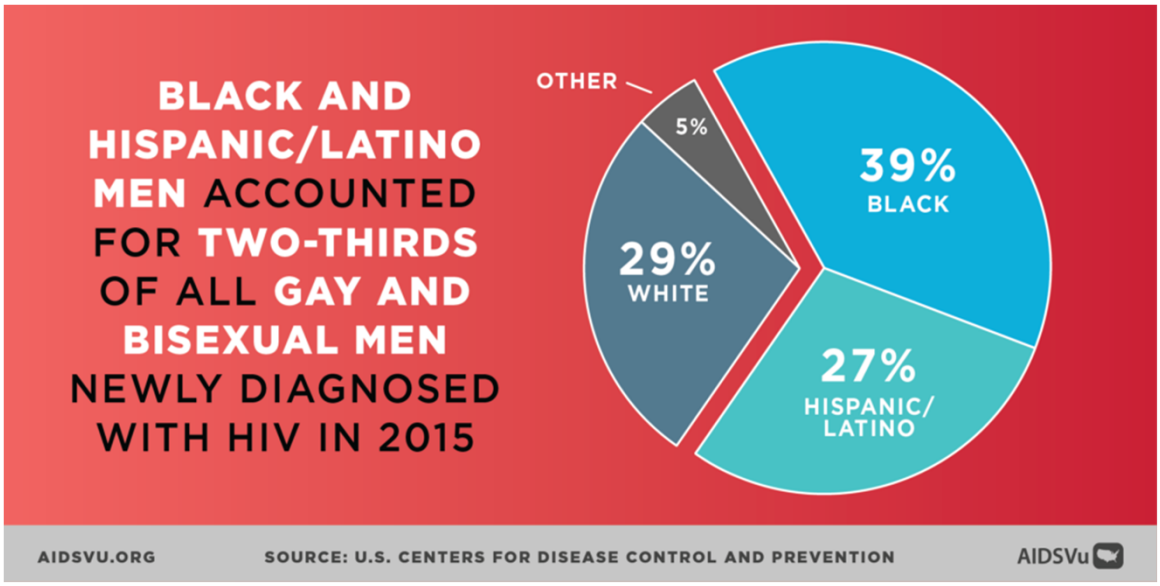
January 18, 2018

EMORY UNIVERSITY
Center for AIDS Research

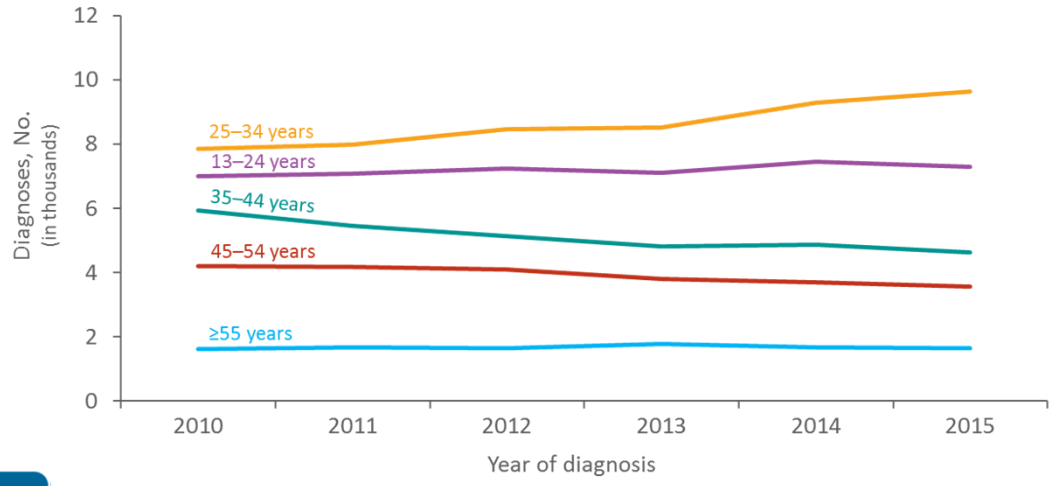


Source: Beyrer, Mayer, Sanchez, Sullivan, Guest eds. Lancet 2012.


GAY AND BISEXUAL MEN
MAKE UP 2% OF THE U.S. POPULATION BUT ACCOUNTED FOR 67% OF ALL NEW HIV DIAGNOSES IN THE U.S. IN 2015
0-7
8-23
24-103
104+
Number of Males Newly Diagnosed with HIV Attributed to Male-to-Male Sexual Contact, 2015
AIDSVU.ORG
SOURCE: U.S. CENTERS FOR DISEASE CONTROL AND PREVENTION
AIDSVu


BLACK AND HISPANIC/LATINO MEN ACCOUNTED FOR TWO-THIRDS OF ALL GAY AND BISEXUAL MEN NEWLY DIAGNOSED WITH HIV IN 2015
OTHER
5%
39%
BLACK
29%
WHITE
27%
HISPANIC/
LATINO
AIDSVU.ORG
SOURCE: U.S. CENTERS FOR DISEASE CONTROL AND PREVENTION
AIDSVu

Diagnoses of HIV Infection among Men Who Have Sex with Men, by Age at Diagnosis, 2010–2015—United States and 6 Dependent Areas



Note: Data have been statistically adjusted to account for missing transmission category. Data on men who have sex with men do not include men with HIV infection attributed to male-to-male sexual contact *and* injection drug use.

How to sample MSM for HIV prevention research?

- Venue-based sampling
- Online sampling – General social media
- Virtual venues (sex-seeking apps)

Surveillance of HIV Risk and Prevention Behaviors of Men Who Have Sex with Men—A National Application of Venue-Based, Time-Space Sampling

Duncan A. MacKellar, MA, MPH[a]
Kathleen M. Gallagher, DSc, MPH[a]
Teresa Finlayson, MPH[a]
Travis Sanchez, DVM, MPH[a]
Amy Lansky, PhD[a]
Patrick S. Sullivan, DVM, PhD[a]

SYNOPSIS

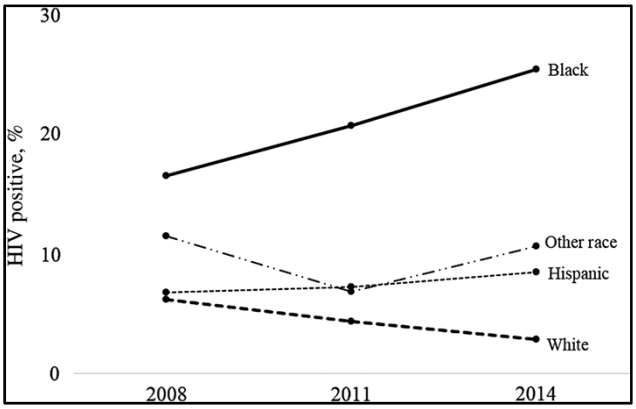
In collaboration with the Centers for Disease Control and Prevention, participating state and local health departments, universities, and community-based organizations applied venue-based, time-space sampling methods for the first wave of National HIV Behavioral Surveillance of men who have sex with men (NHBS-MSM). Conducted in 17 metropolitan areas in the United States and Puerto Rico from November 2003 through April 2005, NHBS-MSM methods included: (1) formative research to learn the venues, times, and methods to recruit MSM; (2) monthly sampling frames of eligible venues and day-time

Source: Public Health Reports 122 1 suppl (2007): 39-47

Venue Based Sampling

- Formative work – venue enumeration
- Observations
- Development of venue-day-time periods
- Construction of sampling frame of VDTs
- Development of sampling calendar
- Selection of sample
- Within venues: systematic, flow-based sampling

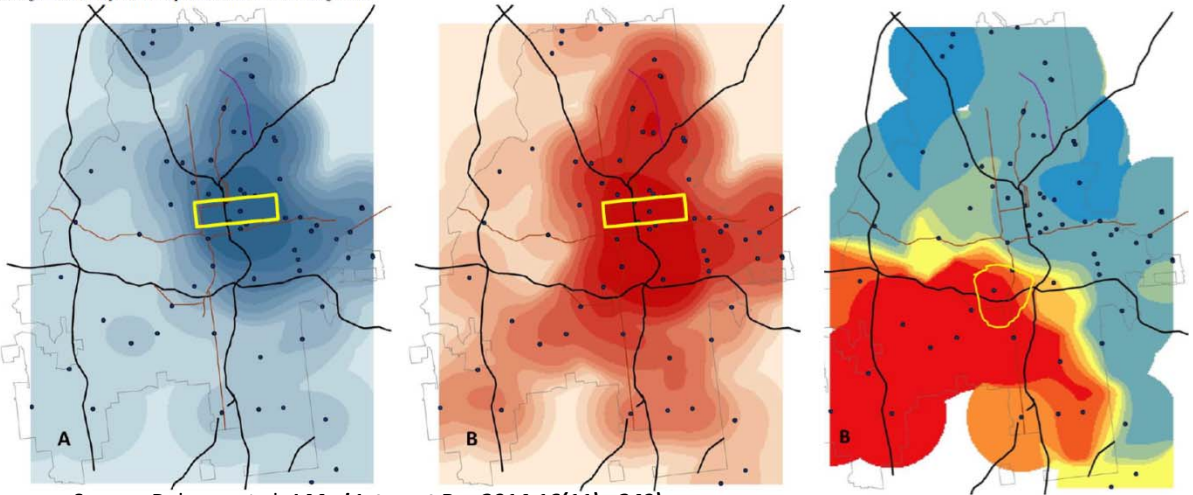
Human immunodeficiency virus (HIV) prevalence among 18–24-year-old men who have sex with men (MSM) who were interviewed in 2008, 2011, and 2014, by year of interview, National HIV Behavioral Surveillance, 20 US cities.



From: Age-Specific Race and Ethnicity Disparities in HIV Infection and Awareness Among Men Who Have Sex With Men—20 US Cities, 2008–2014
J Infect Dis. 2015;213(5):776-783. doi:10.1093/infdis/jiv500

VBS: Place matters

Figure 6. Estimated density of white (A) and black (B) social network application users in Atlanta (gray outline), showing major highways (black lines) and roads (dark red lines) and highlighting the "Midtown" area of Atlanta (yellow rectangle); kernel densities estimated from sample data standardized to 1-mile circular radii and smoothed to 2 miles using a Gaussian smoother that concentrates the majority of the density at the sample point and averages over all adjacent data points within the smoothing radius.



Source: Delaney et al, ***J Med Internet Res 2014;16(11):e249)***

JOURNAL OF MEDICAL INTERNET RESEARCH Sullivan et al

Original Paper

Bias in Online Recruitment and Retention of Racial and Ethnic Minority Men Who Have Sex With Men

Patrick S Sullivan[1], DVM PhD; Christine M Khosropour[1], MPH; Nicole Luisi[1], MPH; Matthew Amsden[2], MBA; Tom Coggia[2], BSID MFA; Gina M Wingood[3], ScD MPH; Ralph J DiClemente[3], PhD MSc

[1]Rollins School of Public Health, Department of Epidemiology, Emory University, Atlanta, GA, United States

[2]Cyclogram, West Hollywood, CA, United States

[3]Rollins School of Public Health, Department of Behavioral Sciences and Health Education, Emory University, Atlanta, GA, United States

Online studies undersample Black and Hispanic MSM

Table 1. Selected Internet-based HIV prevention studies of men who have sex with men depicting population prevalence from recruitment location, enrolled study population prevalence, and corresponding prevalence ratio of black and Hispanic men

		Black Men			Hispanic Men		
Internet Study	Location	Population Prevalence (%)	Enrolled Prevalence (%)	Prevalence Ratio	Population Prevalence (%)	Enrolled Prevalence (%)	Prevalence Ratio
Grosskopf et al, 2010 [12]	New York City	25.1	17.9	0.71	27.4	13.5	0.49
Chiasson et al, 2009 [11]	United States	12.4	6.3	0.51	15.1	14.2	0.94
Rosser et al, 2009[a] [13]	United States	12.4	16.4	1.3	15.1	25.1	1.7
Berg et al, 2007 [14]	United States	12.4	2.5	0.20	15.1	1.7	0.11
Mackellar et al, 2007[b] [15]	6 US cities	25.3	8.6	0.34	30.2	18.8	0.62
Chiasson et al, 2007 [16]	United States and Canada	11.3	4.6	0.41	15.1	7.1	0.57
Bull et al, 2004[c] [17]	United States	12.4	6.6	0.53	15.1	10.9	0.72
Hirshfield et al, 2004 [18]	United States	12.4	2.0	0.16	15.1	5.5	0.36

Source: Sullivan et al, J Med Int Res 2011; 13(2):e28

Figure 1. Shown are six banner advertisements displaying white (left), black (middle), and Asian (right) models used to recruit potential participants from MySpace.com for an online HIV behavioral risk study in the United States in 2009

Source: Sullivan et al, J Med Int Res 2011; 13(2):e28

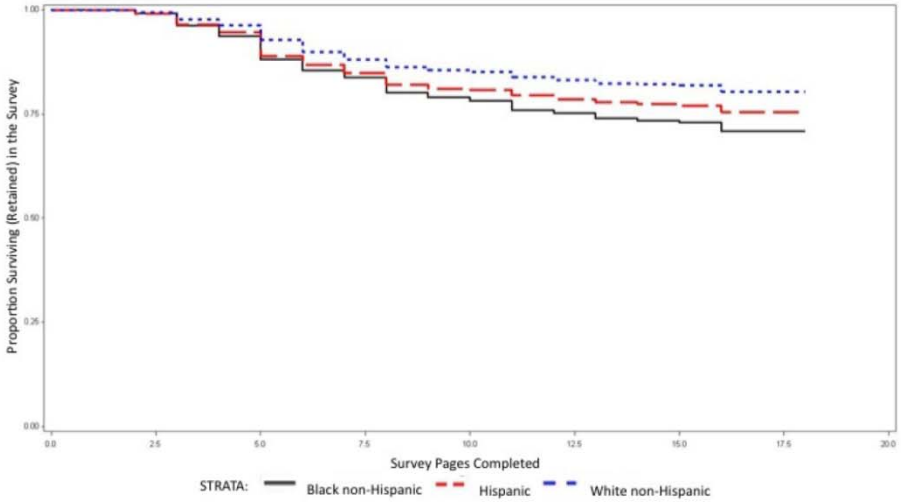
Table 2. Odds of clicking on study banner advertisements by MySpace.com users controlling for self-reported education, sexual identity, and race of model in advertisements and stratified by race of the MySpace.com user in the United States in 2009

Characteristic	White Men Adjusted OR (95% CI)	Black Men Adjusted OR (95% CI)	Hispanic Men Adjusted OR (95% CI)	Other Men Adjusted OR (95% CI)
Education				
< High School (referent)				
> High School	0.99 (0.95 - 1.04)	*1.20 (1.14 - 1.26)* [a]	*1.05 (1.01 - 1.10)*	*1.10 (1.04 - 1.16)*
Identity				
Unsure (referent)				
Gay	*2.10 (1.98 - 2.24)*	*1.62 (1.53 - 1.71)*	*1.45 (1.38 - 1.52)*	*3.07 (2.88 - 3.28)*
Bisexual	*1.63 (1.53 - 1.74)*	*1.78 (1.67 - 1.89)*	*1.58 (1.49 - 1.67)*	*2.83 (2.63 - 3.04)*
Race of model				
White (referent)				
Black	*0.74 (0.70 - 0.79)*	*1.83 (1.72 - 1.95)*	1.05 (0.99 - 1.11)	0.95 (0.89 - 1.00)
Asian	*1.56 (1.47 - 1.64)*	*1.46 (1.37 - 1.56)*	*1.70 (1.62 - 1.79)*	*1.61 (1.52 - 1.69)*

[a]Results presented in italics denote significance at $P < .05$.

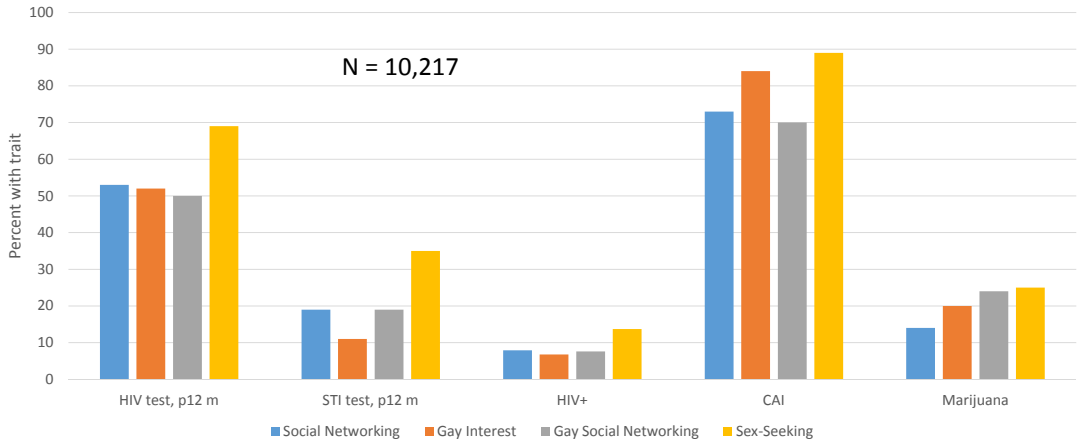
Source: Sullivan et al, J Med Int Res 2011; 13(2):e28

Figure 3. Retention in an online behavioral risk survey among participants reporting only male partners in the past 12 months, by race/ethnicity of the participants in the United States in 2009



Source: Sullivan et al, J Med Int Res 2011; 13(2):e28

Characteristics of MSM recruited through general social networking, general gay interest, gay social networking, and sex-seeking apps, 2016



Source: Zlotprzynska et al, ***JMIR Public Health Surveill 2017;3(1):e13***

Original Paper

The Comparability of Men Who Have Sex With Men Recruited From Venue-Time-Space Sampling and Facebook: A Cohort Study

Alfonso C Hernandez-Romieu[1], MBBS, MPH; Patrick S Sullivan[1], DVM, PhD; Travis H Sanchez[1], MPH, DVM; Colleen F Kelley[1,2], MD, MPH; John L Peterson[3], PhD; Carlos del Rio[2,4], MD; Laura F Salazar[5], PhD; Paula M Frew[2,6], MPH, PhD; Eli S Rosenberg[1], PhD

[1]Rollins School of Public Health, Department of Epidemiology, Emory University, Atlanta, GA, United States

- How to men recruited through Facebook, versus those recruited through VBS, differ in terms of STI and HIV prevalence, retention, and risk behaviors?

Source: Hernandez-Romieu et al, J Med Int Res 2014: 3(3):e37

Risk and testing behaviors among Facebook versus VBS recruited, Atlanta, 2011-2014

FB < VBS	FB = VBS	FB > VBS
• # male partners • Condomless sex partners	• HIV+ • Rectal STI • Syphilis • Main partners • Casual partners • HIV testing • Retention	--

Source: Hernandez-Romieu et al, J Med Int Res 2014: 3(3):e37

Summary

- MSM constitute the major risk group in the US HIV epidemic
- Black MSM, Hispanic MSM and young MSM are disproportionately impacted by HIV
- Historically MSM have been recruited through venues associated with risk (bars, clubs)
- Venue-based sampling is a systematic approach to sampling MSM
- Online sampling can access Black and Hispanic MSM, but are generally underrecruited.
- Race-concordant ads may increase recruitment efficiency for online recruitment for Black MSM
- Black MSM are more prone to loss to follow up within surveys and in prospective studies
- Online-recruited and venue-recruited samples of MSM can be combined

Acknowledgements

- Eli Rosenberg
- Travis Sanchez
- Aaron Siegler
- Christine Khosropour
- Kevin Delaney
- Alfonso Hernandez-Romieu
- PRISM Health Staff
- Research Participants

Supported by

NIAID
NIMH
NICHD
NIDA
CDC
Emory CFAR
The MAC AIDS Fund
Gilead Sciences

F

TAB F

SESSION 4: New and Emerging Designs for Intervention Studies

Presentations in this tab:

Designs for Dissemination and Implementation Research for Small Populations
Amy M. Kilbourne, University of Michigan

Addressing the Challenges of Research with Small Populations
Diane Korngiebel, University of Washington

Designs for Dissemination and Implementation Research for Small Populations

Amy M. Kilbourne, PhD, MPH
Department of Psychiatry, University of Michigan
VA Quality Enhancement Research Initiative (QUERI)
amykilbo@umich.edu

QUERI
VA Quality Enhancement Research Initiative

Outline

Implementation and dissemination science overview

Intervention study designs for implementation research, e.g.,
- Hybrid designs
- Stepped-wedge
- Sequential Multiple Assignment Trial (SMART) designs

Setting the Stage

- ***Dissemination research*** is the scientific study of targeted distribution of information and intervention materials to a specific public health or clinical practice audience. The intent is to understand how best to spread and sustain knowledge and the associated evidence-based health interventions.

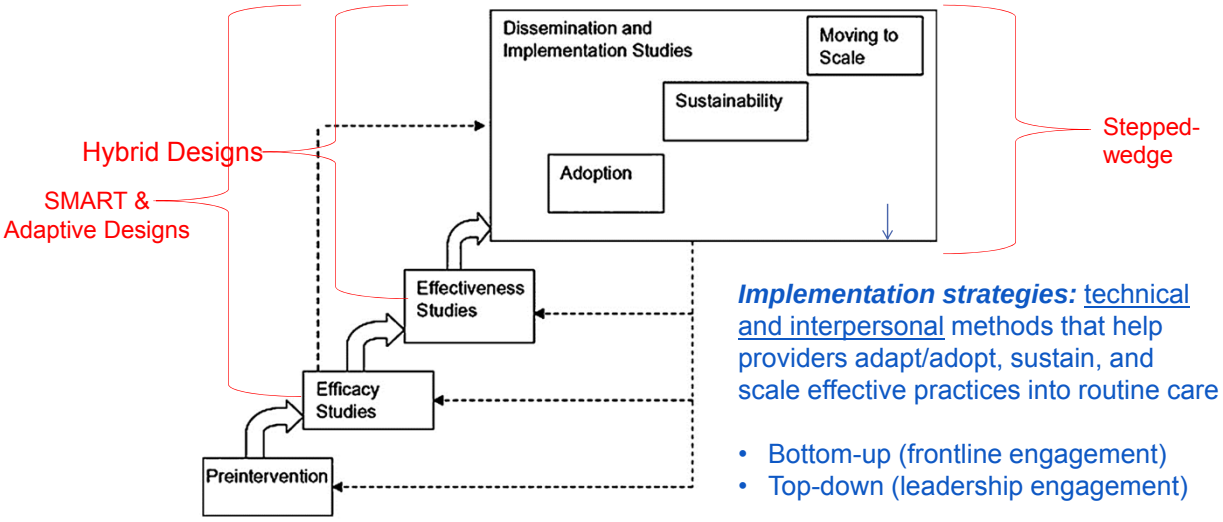
- ***Implementation research*** is the scientific study of the use of strategies to promote the uptake of evidence-based health interventions in clinical and community settings in order to improve patient/population outcomes.

From: NIH PAR 16-238: Dissemination and Implementation Research in Health (R01)

Designs for Implementation & Dissemination Intervention Research

- Randomized controlled trial (RCT)
- Pragmatic clinical trials (PCT)
- Interrupted time series (ITS)
- Dynamic wait list design (DWLD)
- Regression point displacement design (RPDD)
- Stepped-wedge designs
- Hybrid Effectiveness/Implementation Designs
- Sequential Multiple Assignment Randomized Trial/adaptive designs

Study Designs for Implementation Strategies



Implementation strategies: technical and interpersonal methods that help providers adapt/adopt, sustain, and scale effective practices into routine care

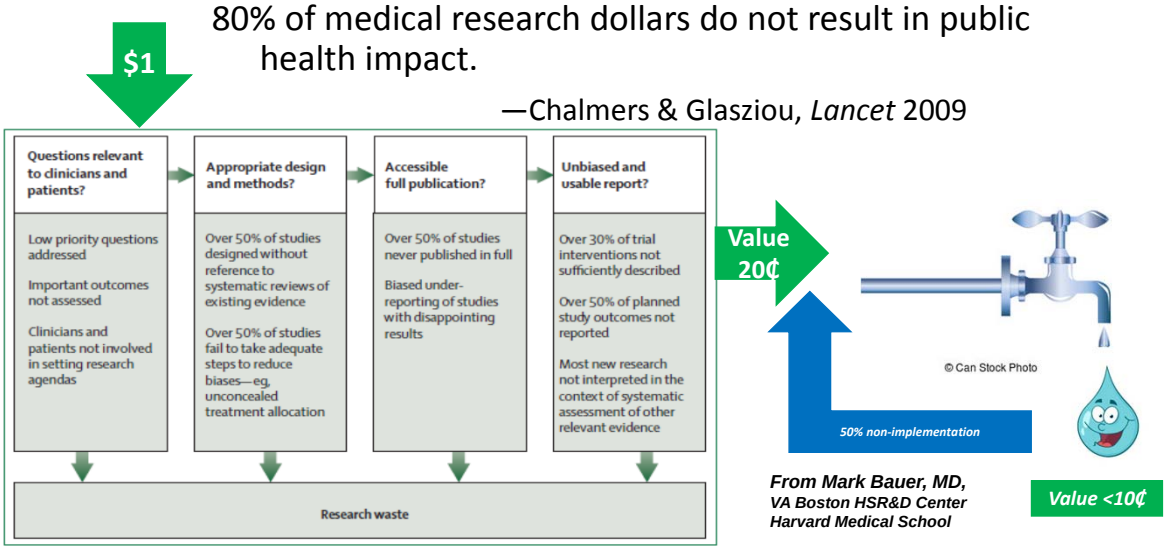
- Bottom-up (frontline engagement)
- Top-down (leadership engagement)

Why Research on Implementation Strategies?

Effective Practices are Not Routinely Implemented for Small Populations

80% of medical research dollars do not result in public health impact.

—Chalmers & Glasziou, *Lancet* 2009



From Mark Bauer, MD,
VA Boston HSR&D Center
Harvard Medical School

Implementation Science Addresses the Research-to-Practice Gap

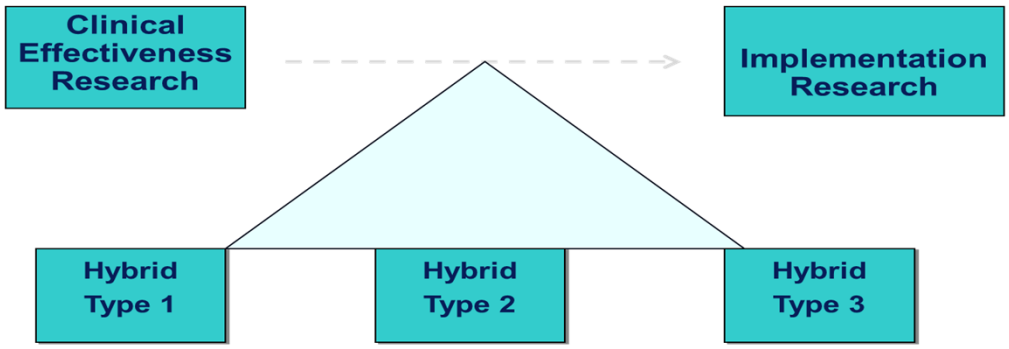
Challenge	Implementation Strategies to Consider	Design Barrier
Interventions not designed for small populations	Tools to adapt to local settings/populations	Sufficient numbers of sites
Interventions rolled out with limited planning	Provider training, facilitation, community engagement	Policy imperative, urgency to "do something"
Intervention reach hard to sustain	Policy incentives, organizational change	Data access/reliability

Hybrid Effectiveness/ Implementation Designs

- Compare implementation strategies
- Address limits of step-wise research (speed research → practice)
- Promote external validity
- Blend effectiveness, implementation stages

Curran, et. al. Effectiveness-Implementation Hybrid Designs. Med Care 2012

Types of Hybrid Designs



Hybrid Type 1: test clinical intervention, observe/gather information on implementation

Hybrid Type 2: test clinical intervention, test implementation strategy

Hybrid Type 3: test implementation strategy, observe/gather information on clinical intervention outcomes

Hybrid Effectiveness/Implementation Designs

	Type I	Type II	Type III
Design Characteristic	Test clinical intervention	Test clinical & implementation strategies	Test implementation strategy
Question	Is treatment effective versus usual care (UC)?	Is treatment delivered through tailored provider coaching effective vs UC?	Does provider coaching vs. training alone improve treatment uptake?
Unit of analyses	Patient	Providers/clinics	Providers/clinics
Primary outcomes	Health outcomes	Process measures	Provider Uptake, Sustainability
Key Advantage	"Cleanest" in determining intervention effectiveness	Ideal when there is time-sensitive need to roll out intervention	All participants get intervention, focus on what will it take to sustain

Hybrid Type I Example:

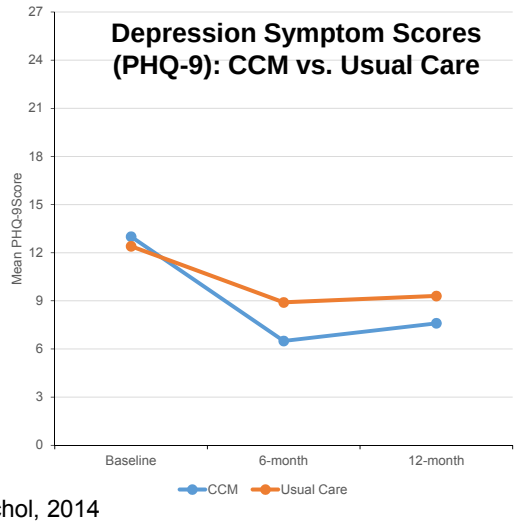
National Implementation of Collaborative Care Model (CCM) for Aetna Enrollees with Mood Disorders from Small Group Practices

Collaborative Care Model

- Care Management
- Self-management support

Usual Care

- Wellness mailings



Kilbourne AM et al, BMC Psychol, 2014

Hybrid Type II Example:

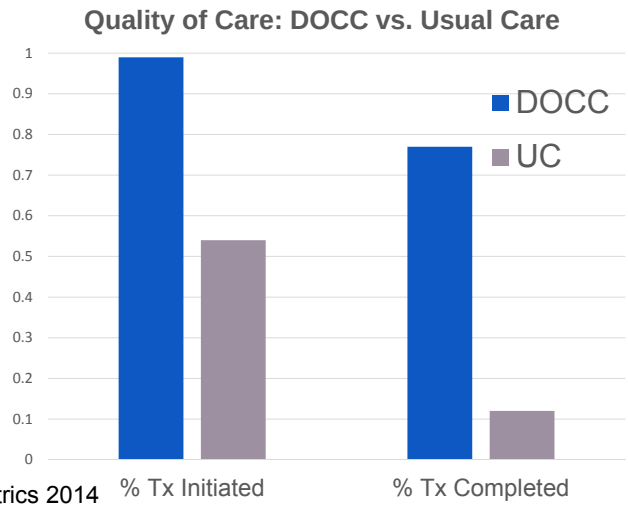
Implementing Doctor-Office Collaborative Care to Improve Pediatric Behavioral Health Outcomes

DOCC

- Care Management
- Family Support
- **Provider Training & Consultation**

Usual Care

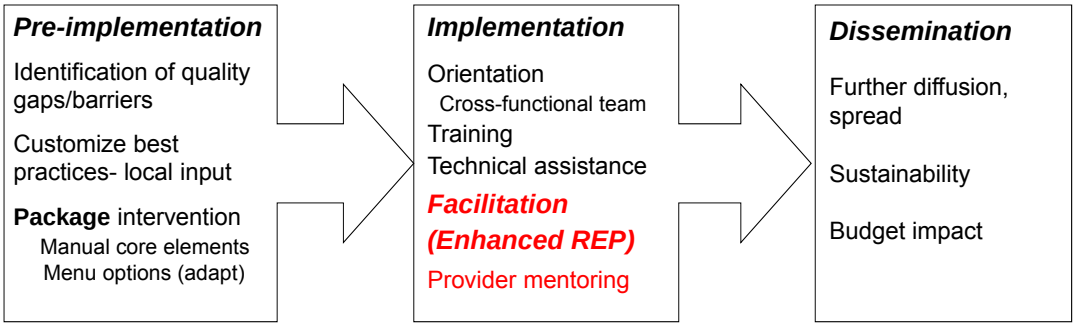
- Standard primary care



Kolko DJ et al, Pediatrics 2014

Hybrid Type III Examples:

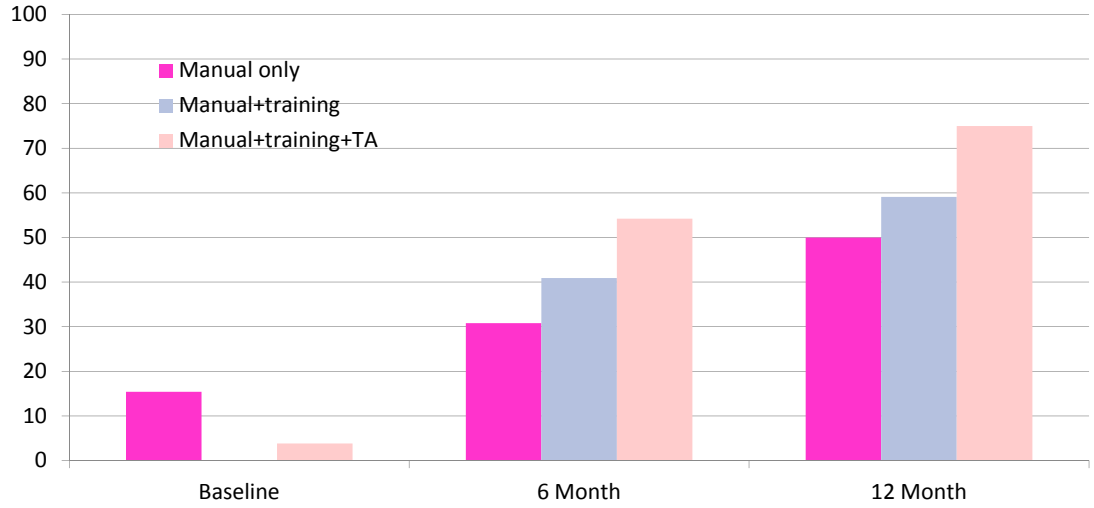
Enhanced Replicating Effective Programs (REP) Implementation Strategy



- *REP was developed by the Centers for Disease Control to rapidly translate prevention programs to community-based settings (Social Learning Theory, Rogers' Diffusion model) (Kegeles 2000; Kilbourne 2007)*
- *Enhanced REP added Facilitation (regular coaching by implementation expert) to support providers in implementation self-efficacy through identifying/mitigating barriers to adoption, building coalitions at sites, and enhancing communication with leaders (Kilbourne et al Implementation Science 2014)*

Hybrid Type III Example #1:

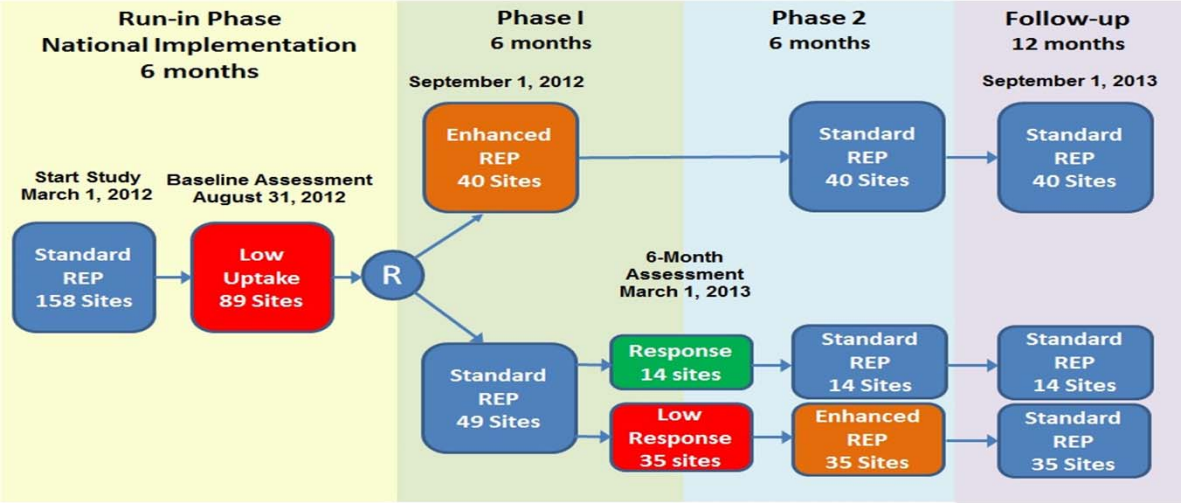
Implementation Strategies and Uptake of HIV Prevention Interventions in AIDS Service Organizations



Kelly J, et al. AJPH 2000

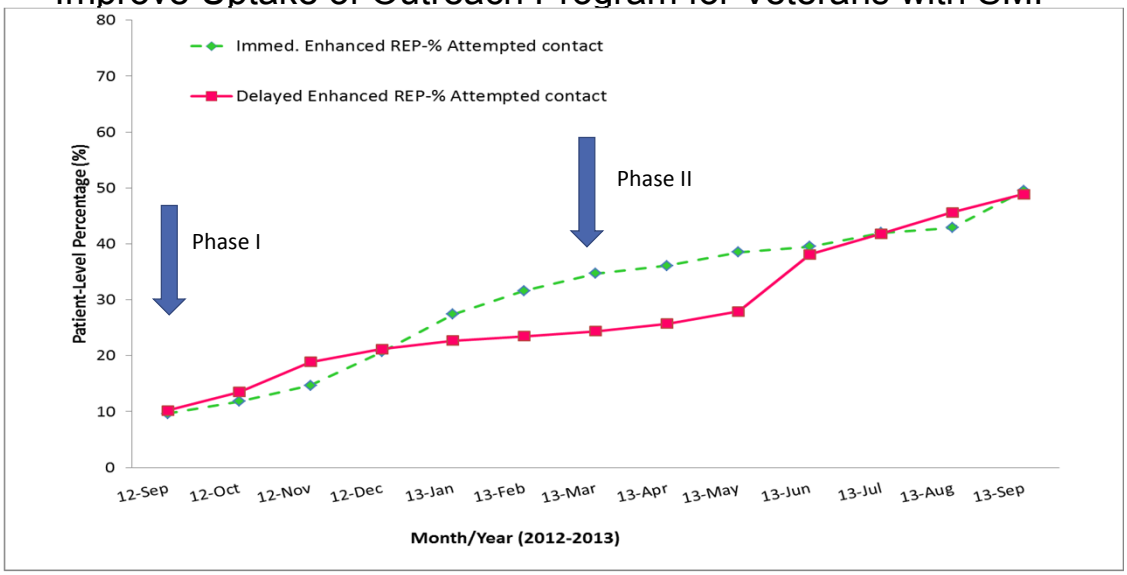
Hybrid Type III Example #2:

Immediate vs. Delayed Enhanced REP Implementation Strategy to Improve Uptake of Outreach Program for Veterans with SMI



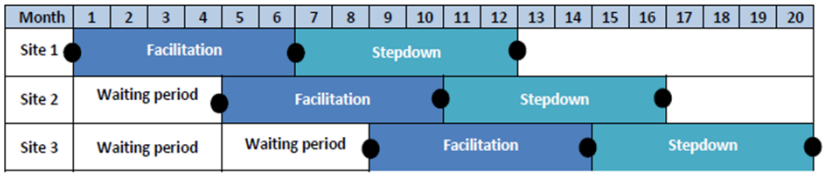
Hybrid Type III Example #2:

Immediate vs. Delayed Enhanced REP Implementation Strategy to Improve Uptake of Outreach Program for Veterans with SMI



Stepped-Wedge Designs Overview

- All participants receive uniform intervention
- Start-time is randomized
- Ideal when resources are too limited to intervene at same time



Stepped-Wedge Design Advantages

Budgetary:

- Resources too limited to intervene at the same time at all participants/sites

Policy:

- Policy imperative to have all participants receive intervention

Pragmatic:

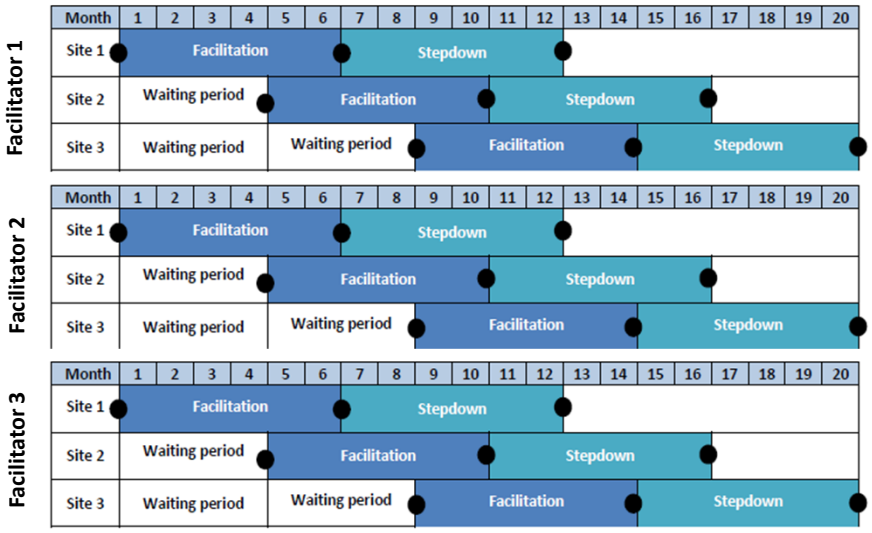
- Advantageous for recruiting & retention to have all participants receive intervention

Ethical:

- Intervention clearly causes more good than harm for participants, rather than equipoise

Stepped-Wedge Design Example:

Provider Facilitation -Collaborative Care in Mental Health Clinics



Sequential Multiple Assignment Trials (SMART)

Towards Precision Implementation

- Multi-stage trials; same subjects throughout
- Each stage corresponds to a critical decision point
- Pre-specified measure of responsiveness
- Treatment options at randomization restricted depending on history of responsiveness
- Subjects randomized to set of treatment options

The goal of a SMART is to inform development of adaptive intervention strategies

When to Use SMART Designs for Implementation

Often insufficient evidence/theory to decide:

- Which implementation strategy(ies) should I **start with**?
- What should I do for sites that are ***non-responsive*** to first-line implementation strategy?
- What should I do for sites that are ***responsive*** to first-line implementation?

SMART designs

can help to answer these questions.

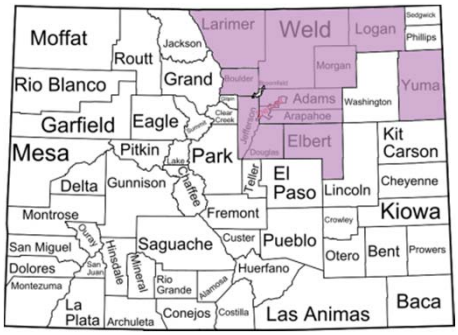
Adaptive Implementation Interventions: Example: Adaptive Implementation of Effective Programs Trial (ADEPT) Study

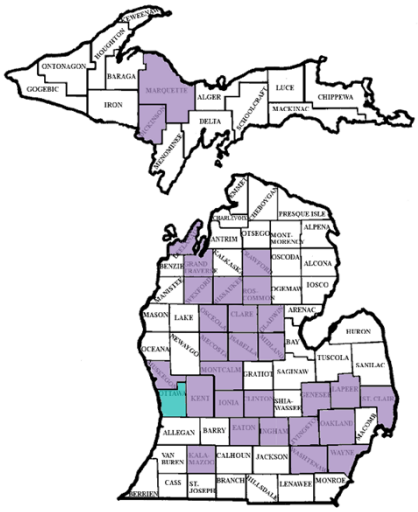
The question:

What is the best way to implement a collaborative care model (Life Goals) in community-based practices to improve patient mental health outcomes?

Kilbourne AM et al. (2014). Implementation Science, 9(1), 132; R01 MH 099898

ADEPT Setting:
Small Practices in Michigan & Colorado

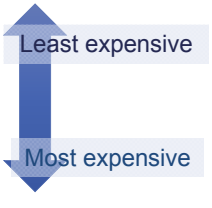




Example: Adaptive Implementation of Effective Programs Trial (ADEPT)

Implementation strategy options:

- Replicating Effective Programs (REP)
- External Facilitation (EF)
- External + Internal Facilitation (EF/IF)



Adaptive Implementation Interventions: Rationale for ADEPT

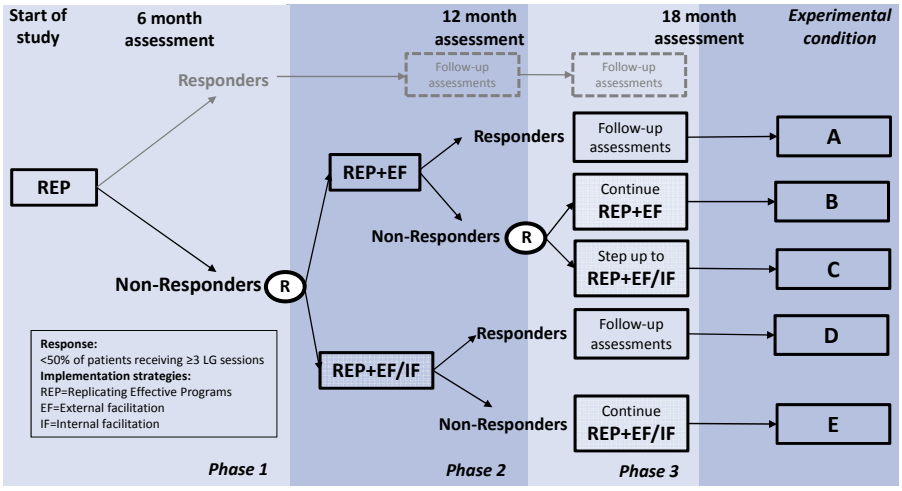
Prior evidence says:

- **REP** will work for **some** sites, but likely not **most**
  - But we don't really know which…
- Most sites will need **more support than REP**

But we don't know:

- What do we do when **REP doesn't work**?
  - Step up directly to **EF/IF** or to **EF**? (Aim 1)
  - What if we step up to **EF** but sites still don't respond? (Aim 2)

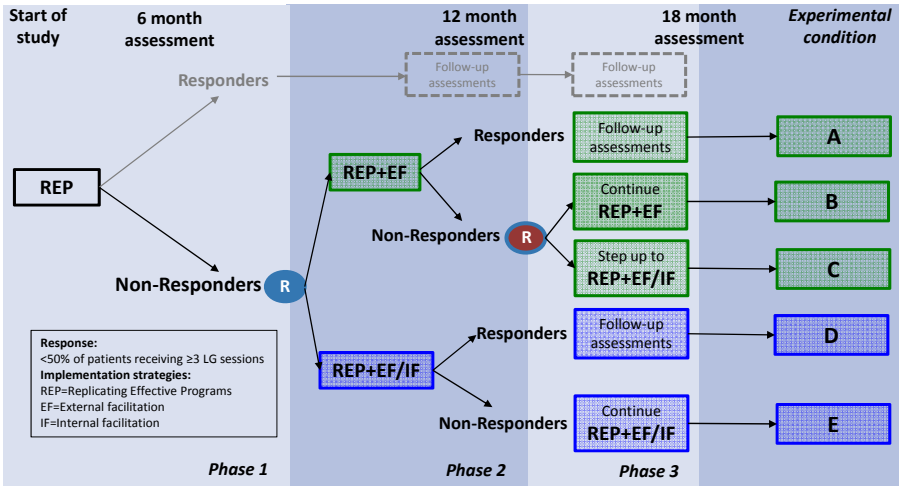
ADEPT Study Design



Kilbourne AM et al. (2014). Implementation Science, 9(1), 132.

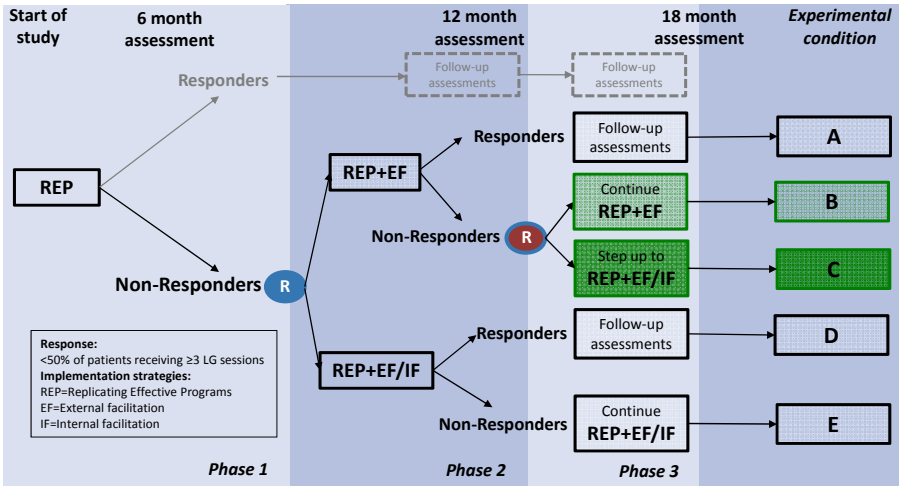
ADEPT Study Design: Aim 1

Is EF+IF better than EF alone for non-responding sites?



ADEPT Study Design: Aim 2

Is continuing EF+IF or EF alone better for non-responding sites?



Future Directions

- Enhancing reach: community organizations, schools, etc.
- Implementation strategies: everyone gets something
- Randomization: stakeholder timelines
- Data capture strategies

THANK YOU!

Contributors:
Shawna Smith, PhD, UM Dept. of Psychiatry, Institute for Social Research
Daniel Almirall, PhD, UM Institute for Social Research
Mark Bauer, MD, VA Boston and Harvard Medical School

Funding: NIMH R01 MH099898, R01 MH114203, VA HSRD 11-232

Disclosure: The views expressed are those of the authors and do not necessarily represent the views of the Veterans Administration

Contact: amykilbo@umich.edu

Designs for Dissemination and Implementation Research for Small Populations
Amy M. Kilbourne

References

Hybrid Implementation Designs:

- Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care. 2012;50(3):217-26. doi: 10.1097/MLR.0b013e3182408812.
- Hybrid Type I example: Kilbourne AM, Nord KM, Kyle J, Van Poppelen C, Goodrich DE, Kim HM, Eisenberg D, Un H, Bauer MS. Randomized controlled trial of a health plan-level mood disorders psychosocial intervention for solo or small practices. BMC Psychol. 2014;2(1):48. doi: 10.1186/s40359-014-0048-x.
- Hybrid Type II example: Kolko DJ, Campo J, Kilbourne AM, Hart J, Sakolsky D, Wisniewski S. Collaborative care outcomes for pediatric behavioral health problems: a cluster randomized trial. Pediatrics. 2014;133(4):e981-92. doi: 10.1542/peds.2013-2516. PMID: 24664093
- Hybrid Type III example: Kilbourne AM, Almirall D, Goodrich DE, Lai Z, Abraham KM, Nord KM, Bowersox NW. Enhancing outreach for persons with serious mental illness: 12-month results from a cluster randomized trial of an adaptive implementation strategy. Implement Sci. 2014;9:163. doi: 10.1186/s13012-014-0163-3.

Stepped-wedge designs- Implementation studies:

- Grayling MJ, Wason JM, Mander AP. Group sequential designs for stepped-wedge cluster randomised trials. Clin Trials 2017;14(5):507-517. doi: 10.1177/1740774517716937. PMID: 28653550
- Grayling MJ, Wason JM, Mander AP. Stepped wedge cluster randomized controlled trial designs: a review of reporting quality and design features. Trials. 2017;18(1):33. doi: 10.1186/s13063-017-1783-0.
- Bauer MS, Miller C, Kim B, Lew R, Weaver K, Coldwell C, Henderson K, Holmes S, Seibert MN, Stolzmann K, Elwy AR, Kirchner J. Partnering with health system operations leadership to develop a controlled implementation trial. Implement Sci. 2016 Feb;11:22. doi: 10.1186/s13012-016-0385-7.

SMART Implementation Designs:

- Kilbourne AM, Almirall D, Eisenberg D, Waxmonsky J, Goodrich DE, Fortney JC, Kirchner JE, Solberg LI, Main D, Bauer MS, Kyle J, Murphy SA, Nord KM, Thomas MR. Protocol: Adaptive Implementation of Effective Programs Trial (ADEPT): cluster randomized SMART trial comparing a standard versus enhanced implementation strategy to improve outcomes of a mood disorders program. Implement Sci. 2014;9:132. doi: 10.1186/s13012-014-0132-x.

- NeCamp T, Kilbourne A, Almirall D. Comparing cluster-level dynamic treatment regimens using sequential, multiple assignment, randomized trials: Regression estimation and sample size considerations. Stat Methods Med Res. 2017;26(4):1572-1589. doi: 10.1177/0962280217708654. Epub 2017 Jun 19.
- Penn State Methodology Center SMART design examples: https://methodology.psu.edu/ra/adap-inter/projects

Other references on implementation trial designs:

- Brown CH, Curran G, Palinkas LA, Aarons GA, Wells KB, Jones L, Collins LM, Duan N, Mittman BS, Wallace A, Tabak RG, Ducharme L, Chambers DA, Neta G, Wiley T, Landsverk J, Cheung K, Cruden G. An Overview of Research and Evaluation Designs for Dissemination and Implementation. Annu Rev Public Health. 2017;38:1-22. doi: 10.1146/annurev-publhealth-031816-044215.
- Wyman PA, Henry D, Knoblauch S, Brown CH. Designs for Testing Group-Based Interventions with Limited Numbers of Social Units: The Dynamic Wait-Listed and Regression Point Displacement Designs. Prev Sci. 2015;16(7):956-66. doi: 10.1007/s11121-014-0535-6. PMID: 25481512 Free PMC Article

UW Medicine
BIOMEDICAL INFORMATICS AND MEDICAL EDUCATION

ADDRESSING THE CHALLENGES OF RESEARCH WITH SMALL POPULATIONS

Improving Health Research for Small Populations Workshop
Washington, DC, Jan. 18-19, 2018

Diane M. Korngiebel, University of Washington, Seattle

UW Medicine
BIOMEDICAL INFORMATICS AND MEDICAL EDUCATION

Focus for presentation (Srinivasan, et al., 2015)

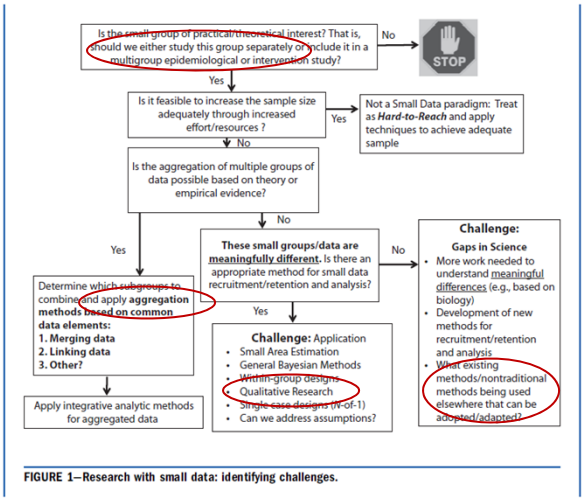


FIGURE 1—Research with small data: identifying challenges.

UW Medicine
BIOMEDICAL INFORMATICS AND MEDICAL EDUCATION

Talk agenda

- Bioethics and small populations research
- Small populations and data aggregation
  - The "data cycle"
  - Data challenges when population numbers are small
  - Qualitative methods and data aggregation
  - Summary of case study findings
- Role of co-production and some approaches

UW Medicine
BIOMEDICAL INFORMATICS AND MEDICAL EDUCATION

Should we study this group separately?

Visiting three "pillars" of bioethics in a small populations context

Beneficence and non-maleficence

- Does the population benefit from separate study?
  - Relevant data
  - Tailored interventions
  - Resources to address local needs
- Could there be harm from separate study?
  - Inadequate numbers for meaningful results
  - Potential use of data to stigmatize group
- Could there be harm if not studied separately?
  - Invisibility to research agendas, resource allocators
  - Inappropriate interventions with low uptake
  - Perpetuation of disparity

Respect for autonomy

The idea of autonomy and respect should be expanded beyond the traditional application to individuals.

- Does the population have an ethno-cultural community identity?
- Other community identity (i.e., beyond hard to reach)?
- Health disparities research must:
  - Respect individual autonomy
  - And community autonomy and identity

UW Medicine
BIOMEDICAL INFORMATICS AND MEDICAL EDUCATION

Justice and equity

Opportunities to address injustice and inequity should guide health disparities research.

- Has the small population experienced disadvantage as a population or group?
  - For example, consider the social determinants of health (SDOH)
- Could health disparities research inform resource distribution to address SDOH?
  - Not necessarily more healthcare (Woolf, et al., 2007)

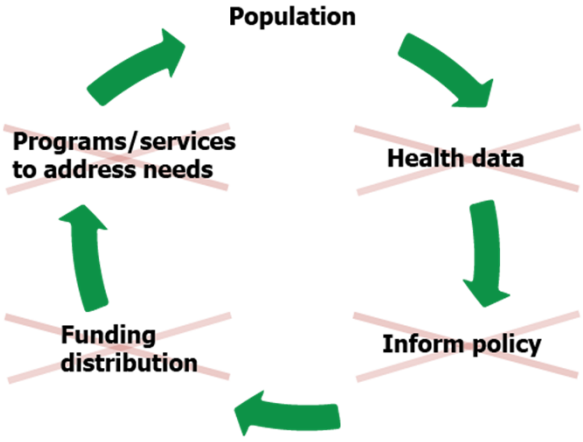
UW Medicine
BIOMEDICAL INFORMATICS AND MEDICAL EDUCATION

Small populations & data aggregation

Justice and equity and the role of qualitative methods

The (lack of) data cycle (Tauali et al., 2014)

- Lack of data → inequitable distribution of resources → increased health disparities



Goal **BREAK THE CYCLE**

Data collection challenges (Korngiebel, et al., 2015)

Current methods do not support the collection of accurate data for small populations.

- Low survey response rates (state, regional, national)
- Ethnic and racial misclassification
  - Including *ad hoc* assigning of "category" by outsiders

Challenge **REVISE CURRENT METHODS**

Data aggregation challenges

UW Medicine BIOMEDICAL INFORMATICS AND MEDICAL EDUCATION

Current methods do not support relevant aggregation of data for small populations.

- Groups of unequal size are collapsed
  - E.g., Asians (96%)/Pacific Islanders (4%)
  - Issues of smaller group subsumed by larger
- Or imposed categories neglect context
  - E.g., Native Hawaiians and other Pacific Islanders
  - Some SDOH may be shared but some may not

Challenge → **DISPARITIES ARE MASKED**

What is valid data?

UW Medicine BIOMEDICAL INFORMATICS AND MEDICAL EDUCATION

Some data are considered more "valid" than others.

- Defining valid data
  - Privileging types of data → gatekeeping
  - Aggregation → gatekeeping
- Ways of knowing (See for example, Walker and Bigelow, 2011)
- Role of qualitative data
  - Context
  - Perspective

Challenge → **EXPAND IDEA OF "DATA"**

"THE TAKE HOME POINT IS THAT DATA SHOULD BE ETHICAL—AND DO NO HARM. SMALL, LARGE, WHATEVER FORM IT TAKES, IT SHOULD NEVER INFLICT HARM ON A PEOPLE.

THAT ETHICAL STANCE SOMETIMES REQUIRES US TO WORK WITH DATA IN WAYS WE MIGHT NOT HAVE LEARNED IN BIOSTATS COURSES THAT VALUED THE NORMAL DISTRIBUTION.

WE SOMETIMES HAVE TO DIG DEEPER, AND ALWAYS WITH HUMILITY, RESPECT, AND KINDNESS."

~DR. MAILE TAUALII

UW Medicine
BIOMEDICAL INFORMATICS AND MEDICAL EDUCATION

Case study: aggregation insights

What tribal partners recommended

Mixing it up: a case study (Van Dyke, et al. 2016)

UW Medicine BIOMEDICAL INFORMATICS AND MEDICAL EDUCATION

Study timeline

- 2009 conference with Indigenous and Tribal health leaders to identify the issue
- 2011 bioethics administrative supplement to U54
- 2012-2013: Data collection and analysis
- 2014-2016: Tribal review and publications

Mixing it up: a case study

UW Medicine BIOMEDICAL INFORMATICS AND MEDICAL EDUCATION

Communities share their criteria for improved data aggregation.

- Five tribes
  - Varying sizes
  - Engagement approach: Tribal Participatory Research (Fisher and Ball, 2003)
- What characteristics should be considered when data are aggregated?

Goal → **MORE RELEVANT DATA**

The qualitative approach

UW Medicine BIOMEDICAL INFORMATICS AND MEDICAL EDUCATION

Qualitative methodologies → direct engagement.

- Data collection
  - Key informant interviews and focus groups
- Analysis
  - Single coding with study team review
  - Consensus resolution
  - Member checking

Goal → **MORE RELEVANT DATA**

What we learned

UW Medicine BIOMEDICAL INFORMATICS AND MEDICAL EDUCATION

Many factors might inform data aggregation.

- Tribal partners identified significant variables
  - Geographic proximity
  - Community type (urban/rural; coastal/inland)
  - Culture
  - Presence/absence of contaminated environment
  - Type/severity of health concerns
  - Access to health care
  - Generational cohort

Result → **ADDED RELEVANCE**

Geographic proximity was important…

…but was not the whole story.

- Community type (urban/rural; coastal/inland)
- Presence/absence of contaminated environment

Goal → **ADDED RELEVANCE**

Health-related

Communities can already identify priority health concerns.

- Types of health concerns
- Severity of health concerns
- Access to health care

Goal → **ADDED RELEVANCE**

UW Medicine
BIOMEDICAL INFORMATICS AND MEDICAL EDUCATION

How do we leverage the community wisdom of small populations?

By focusing on co-production and co-creation in our approaches, frameworks, and methodologies.

UW Medicine
BIOMEDICAL INFORMATICS AND MEDICAL EDUCATION

The future is co-production (Turakhia and Combs, 2017)

Collaborative co-creation is the future of health research and health care interventions and delivery— ***and may have particular relevance for small populations.***

- Generating value together
  - The data aggregation method above is an example of co-production
- Users and communities co-shape and co-make interventions/products/services
- Such approaches prioritize and invest in collaborations with those most affected by data, research, interventions.

Goal **INVESTMENT → OUTCOMES**

Co-production (table adapted from Israilov and Cho, 2017)

UW Medicine BIOMEDICAL INFORMATICS AND MEDICAL EDUCATION

Challenges	Benefits
Addressing data "hierarchy"	Qualitative context improves local relevance
Engagement takes time	Stakeholder investment in activity/intervention/policy
Recognizing diverse expertise	Stakeholders learn from each other; no "one" expert
Achieving consensus	Development of transparent and inclusive process

FOCUS → **PARTNERSHIPS ADD VALUE**

Approaches from the social sciences…

UW Medicine BIOMEDICAL INFORMATICS AND MEDICAL EDUCATION

An example: Community-based participatory research (CBPR) (Israel, et al., 1998)

- A values-based approach
- All partners contribute expertise to defining the issue and determining the action to take
- Communities are constantly consulted
  - Before: what is the priority?
  - During: Set-up, data, collection, analysis
  - After: Review and dissemination of results; what next?

Goal → **IMPROVE HEALTH**

…and industry…

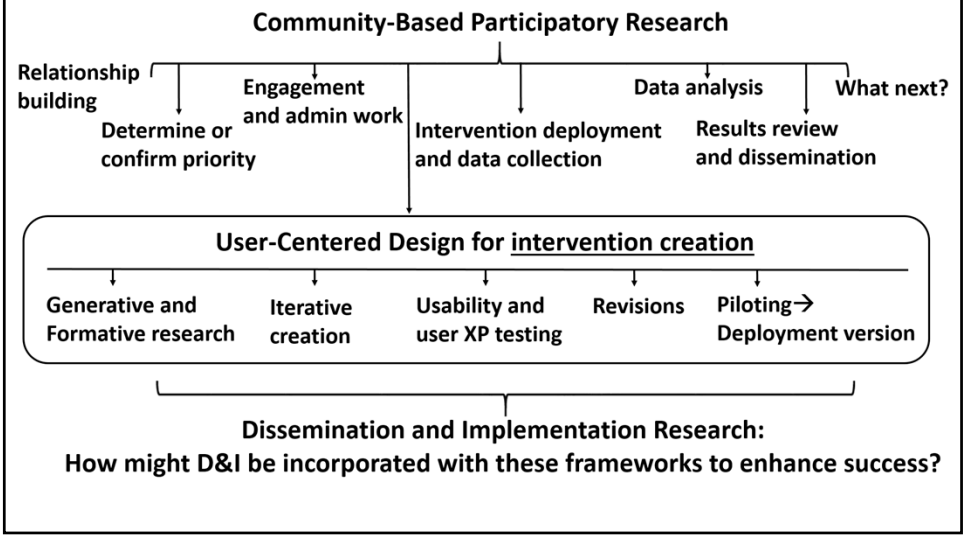
Some industry approaches may be particularly helpful in partnerships with small populations.

- Example: User-Centered Design (contextual design; user-centered system design; user experience) (Nielsen Norman Group)
  - Researches the "lived" context of an intervention
  - Focus on end users & key stakeholders working together to create and refine
  - Use of diverse data collection techniques (IDEO)

→ improved resonance of collection methods

Goal → **IMPROVE HEALTH**

…interwoven.



References

UW Medicine
BIOMEDICAL INFORMATICS AND MEDICAL EDUCATION

Fisher PA, Ball TJ. Tribal participatory research: mechanisms of a collaborative model. American journal of community psychology. 2003;32(3-4):207-216.

IDEO Methods. http://www.designkit.org/methods.

Israel BA, Schulz AJ, Parker EA, Becker AB. Review of community-based research: assessing partnership approaches to improve public health. Annual review of public health. 1998;19:173-202.

Israilov S, Cho HJ. How Co-Creation Helped Address Hierarchy, Overwhelmed Patients, and Conflicts of Interest in Health Care Quality and Safety. AMA journal of ethics. 2017;19(11):1139-1145.

Korngiebel DM, Taualii M, Forquera R, Harris R, Buchwald D. Addressing the Challenges of Research With Small Populations. American journal of public health. 2015;105(9):1744-1747.

Nielsen Norman Group. User experience. https://www.nngroup.com/.

Srinivasan S, Moser RP, Willis G, et al. Small is essential: importance of subpopulation research in cancer control. American journal of public health. 2015;105 Suppl 3:S371-373.

Taualii M, Quenga J, Samoa R. Understanding diversity among indigenous people: Conducting research with Native Hawaiians and Pacific Islanders. In: Solomon T, ed. *Conducting health research in Native American communities*. Washington, DC: American Public Health Press; 2014.

Turakhia P, Combs B. Using Principles of Co-Production to Improve Patient Care and Enhance Value. AMA journal of ethics. 2017;19(11):1125-1131.

Van Dyke ER, Blacksher E, Echo-Hawk AL, Bassett D, Harris RM, Buchwald DS. Health Disparities Research Among Small Tribal Populations: Describing Appropriate Criteria for Aggregating Tribal Health Data. American journal of epidemiology. 2016;184(1):1-6.

Walker RD, Bigelow DA. A constructive Indian country response to the evidence-based program mandate. Journal of psychoactive drugs. 2011;43(4):276-281.

Woolf SH, Johnson RE, Phillips RL, Jr., Philipsen M. Giving everyone the health of the educated: an examination of whether social change would save more lives than medical advances. American journal of public health. 2007;97(4):679-683.

SPECIAL THANKS TO

COMMUNITY COLLEAGUES

FOR THEIR INVALUABLE CONTRIBUTIONS

Questions? dianemk@uw.edu

G

TAB G

SESSION 5: Recruitment, Retention, and Collection of Data with a Focus on Small or Hard to Reach Populations

Presentations in this tab:

Issues and Challenges Associated with Recruitment and Retention for Health Research
Vetta Sanders-Thompson, Washington University in St. Louis

Improving Health Research in Rural Areas
F. Douglas Scutchfield, University of Kentucky

Using Technology for Recruitment, Retention. Data Collection, and Intervention Delivery
Kathi Mooney, University of Utah

Recruitment, Retention, and Collection of Data with a Focus on Small or Hard to Reach Populations
Tracy Onega, Dartmouth University Geisel School of Medicine

ISSUES AND CHALLENGES ASSOCIATED WITH RECRUITMENT AND RETENTION FOR HEALTH RESEARCH

Vetta L. Sanders Thompson,
Brown School
January 19, 2018

Department of Surgery
Division of Public Health Sciences

SITEMAN CANCER CENTER

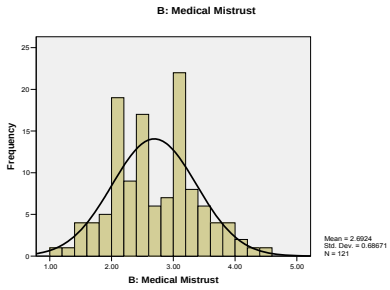
Objectives

- Outline the issues that affect recruitment and retention of participants to research.
- Identify solutions to these issues
- Provide examples of solutions

SITEMAN CANCER CENTER

Challenges to Retention and Recruitment

- Attitudes towards:
  - Research
  - Researchers
  - Institutions
  - Universities



- Research knowledge and literacy
- Outreach strategies and engagement

SITEMAN CANCER CENTER

Basic Requirements for Success

- Flexibility to adjust study recruitment to account for differences in location, behavior, media, technology use, etc.
  - Basic knowledge of a client socio-cultural position
- Communities and participants have to be met with openness and acceptance.
  - Right staff, materials and approach
- Long lasting partnerships are helpful.
  - The challenge for researchers is to work in such a way that trust is developed and maintained.

SITEMAN CANCER CENTER

Basic Requirements for Success

- Make the ask
- Know your audience
- Go where they are
- Build the relationship

SITEMAN CANCER CENTER

Make The Ask

- Staffs need to come from a variety of backgrounds;
- Staff needs to have people skills.
- Maintained continuing contact with study participants;
  - Birthdays, holidays, Facebook and twitter sites that were fun.
- Case management model – a source of referral and support.

SITEMAN CANCER CENTER

Make the Ask

- Health provider connections and support
- Endorsements can be important – political, health, business, church, community

- Right media for the market – racial & ethnic minority media; right frequency –Are you talking to me?
- Right incentives, including non-cash incentives

SITEMAN CANCER CENTER

Know Your Audience

- Education, research literacy
- Media preferences
- Region of the country
- Group diversity
  - Immigration status
  - Level of acculturation
  - Identity
- Gender, age and generation
  - Technology use
  - Media

SITEMAN CANCER CENTER

Learning About Your Audience: The NCS Experience

1. PARTICIPATION WITH AND WITHIN THE COMMUNITY
   - Carving out a respected place in the community for the study through awareness raising efforts.

2. MIRROR THE COMMUNITY
   - The makeup of the NCS staff mirrors the demographics and norms of the community;
   - Community norms and demographics should be considered for mass communication and outreach, screening and recruitment, and when maintaining continuous contact with study participants.

3. TRAINING
   - Consistent relationships with study participants and staff who were responsive and informed often assisted in participants' ability to see as trustworthy.

SITEMAN CANCER CENTER

Go Where They Are: CECCR

- We go where people are everyday;
- Identify in ZIP codes with racial/ethnic, low-income residents.
  - –health centers,
  - –laundry mats,
  - –beauty salons & barber shops

- Neighborhood Voice Mobile Unit (NVMU)
  - –The NVMU is a shuttle-type vehicle customized for research
  - –Allows engagement in the communities.

SITEMAN CANCER CENTER

Go Where They Are:

Survey of Multicultural Factors Affecting African American Colorectal Cancer Screening

- Effective call lists
  - –Targeted list sample, created using random digit dial (RDD) generated lists
  - –Matched to a market research data
  - –Developed to assure that proportional to the geographical distribution of the African Americans
  - –Used a separate RDD list in calling to reduce biases produced by a listed sample.

SITEMAN CANCER CENTER

Go Where They Are: Feasibility Study

- Internet use is widespread and could be a channel to reach and disseminate health information to AA men;
- There are disparities in internet use and limited literature exists on how to best address this divide.
- Our data suggest that disseminating information online is not a very effective way to reach older African American men, with limited education.
- We do not recommend using websites among this population,
- Email was effective in getting participants to the website, even though they expressed a preference for phone messages.

SITEMAN CANCER CENTER

Build the Partnership

Implemented a community based participatory research training program for community members.

Promoted the role of underserved populations in research by enhancing the capacity for community based participatory research.

Bridge Washington University in St. Louis (WUSTL) researchers and community based organizations and community health workers serving the St. Louis Greater Metropolitan area to address health disparities.

SITEMAN CANCER CENTER

Community Research Fellows Training Program (CRFT)

- Train community members to become good consumers of research.
- Understand how to use research as a tool in improving health outcomes in communities.
- Increase community members understanding of how to work with academic researchers.

SITEMAN CANCER CENTER

CRFT

17 Multidisciplinary faculty have trained four cohorts (122 community members) through a semester long public health research training course.

Created a pool of trained community members who collaborate with academic researchers and other health practitioners on community research advisory boards, councils and institutional review boards.

Patient Research Advisory Board (PRAB)

SITEMAN CANCER CENTER

Goal and Objectives

The St. Louis Patient Research Advisory Board (PRAB)

1. Serve in an advisory role to academic researchers on issues of community engagement, building trust, and ethical considerations of research and study design.

2. Provide a forum that allows for mutually beneficial communication between community stakeholders and academic researchers on meaningful, relevant clinical concerns

3. PRAB informs, guides and reviews grant proposals

4. PRAB will fosters academic community linkages and disseminates information about clinical research findings pertinent to the community.


UK
UNIVERSITY OF
KENTUCKY
Improving Health Research in Rural Areas: The Case of Kentucky
F. Douglas Scutchfield, MD
Bosomworth Professor Emeritus
Colleges of Public Health and Medicine
Scutch@uky.edu
see blue.
in everything we do.
An Equal Opportunity University


UK
UNIVERSITY OF
KENTUCKY
The Case of Kentucky
4 M people, 120 Counties; Largest Jefferson 762,000
Smallest Robertson 2155; Median Marion 19,820
Kenton
Boone
Campbell
Gallatin
Bracken
Carroll
Pendleton
Grant
Mason
Trimble
Owen
Robertson
Lewis
Greenup
Henry
Harrison
Oldham
Nicholas
Fleming
Carter
Boyd
Scott
Jefferson
Shelby
Franklin
Bourbon
Rowan
Bath
Elliott
Woodford
Montgomery
Lawrence
Meade
Bullitt
Spencer
Anderson
Fayette
Clark
Menifee
Morgan
Jessamine
Hancock
Nelson
Mercer
Powell
Johnson
Martin
Henderson
Breckinridge
Hardin
Washington
Madison
Wolfe
Union
Daviess
Estill
Magoffin
Boyle
Garrard
Lee
Floyd
Webster
McLean
Ohio
Grayson
Larue
Marion
Lincoln
Breathitt
Pike
Crittenden
Taylor
Rockcastle
Jackson
Owsley
Knott
Livingston
Hopkins
Hart
Casey
Perry
Muhlenberg
Butler
Edmonson
Green
Caldwell
Clay
Ballard
Adair
Laurel
Leslie
Letcher
McCracken
Lyon
Warren
Metcalfe
Russell
Pulaski
Carlisle
Marshall
Christian
Logan
Barren
Knox
Harlan
Trigg
Todd
Cumberland
Wayne
Graves
Simpson
Allen
Monroe
Clinton
Whitley
Bell
Hickman
Calloway
McCreary
Fulton
see blue.
in everything we do.
An Equal Opportunity University

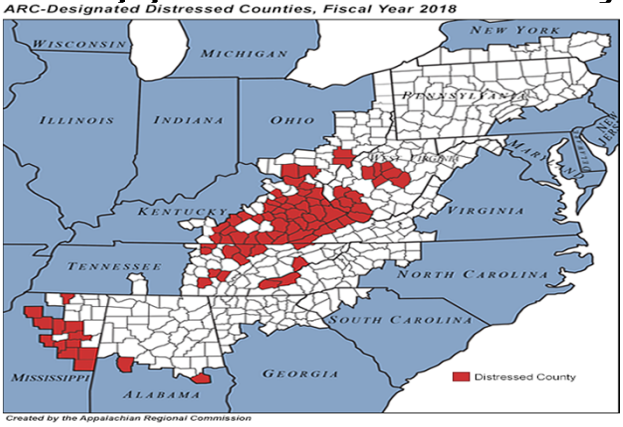
UNIVERSITY OF KENTUCKY

Kentucky Area Development Districts
KY BRFS Data

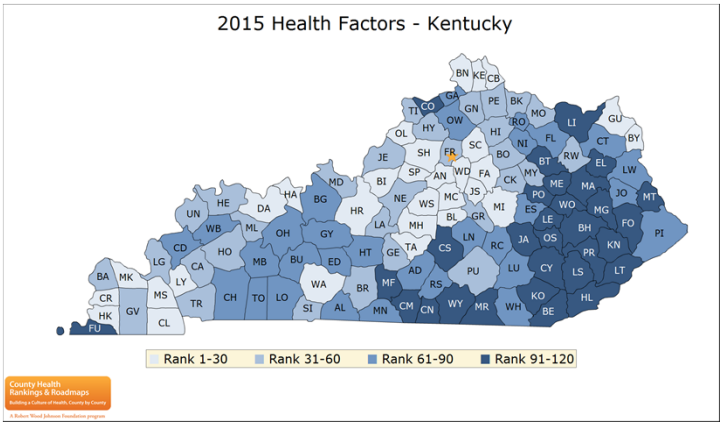
Kentucky Area Development District (ADD) Profiles Kentucky Behavioral Risk Factor Survey (KyBRFS)

UNIVERSITY OF KENTUCKY

Appalachia Kentucky



County Health Rankings



Community engaged research

Community Engaged Research

- is a collaborative process between the researcher and community partner designed to benefit the community and advance knowledge
- it identifies the assets of stakeholders and incorporates them in the design and conduct of the research process; asset mapping
- includes community based participatory research and participatory action research.

Community Asset Mapping

- asset mapping is a tool that relies on a core belief of asse based community development; assets suited to advancin those communities
- assets include physical and economic assets, stories, local residents, local associations, local institutions.

Community Engaged Research

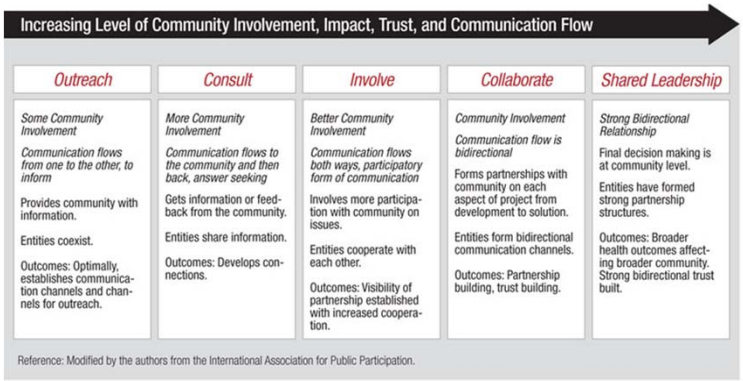


Figure 1.1. Community Engagement Continuum

Location of Existing UK Outreach and Community Activities

UK Research Outreach: Patient and Population Illustrations

- UK Center of Excellence in Rural Health (Hazard)
- CTSA/CCTS Research Liaison (ATRN-ETSU, Marshall, WVU, OSU, Ohio University, UC) (Hazard and Morehead-Research Study Coordinator)
- UK/Kentucky Regional Medical School Program (Morehead, Bowling Green, Covington, Hazard)
- Markey Cancer Center Outreach
- UK Cooperative Extension (Project HEEL)
- Area Health Education Centers (Morehead, Hazard, Covington, Mt Vernon)
- Research Networks (KAN, KPHREN, Dentistry, Rehabilitation)

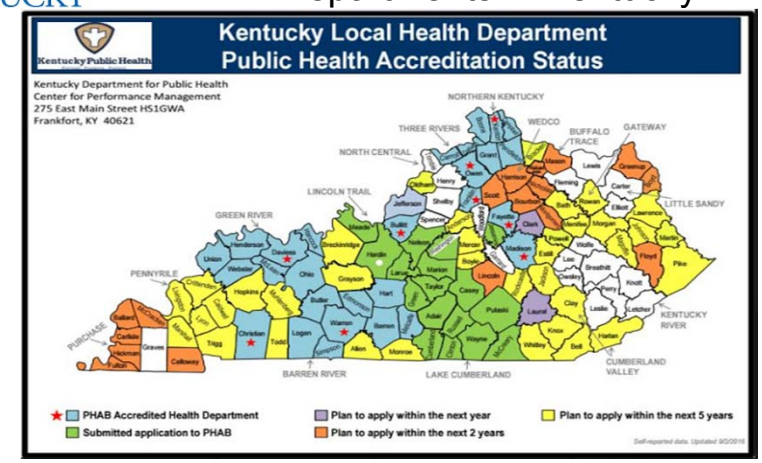
Illustrations of opportunities and Mature Coalitions

PHAB Accreditation Standards

Standard 1.1 Participate in or Lead a Collaborative Process Resulting in a Comprehensive Community Health Assessment

Standard 5.2 Conduct a Comprehensive Planning Process Resulting in a Tribal/State/Community Health Improvement Plan

PHAB Accreditation of Local Health Departments in Kentucky



Non-profit Hospitals and CHNA

- **Recent changes in legislation (ACA) now require that non-profit hospitals explicitly and publicly demonstrate community benefit by conducting a community health needs assessment (CHNA) and adopting an implementation strategy to meet the identified community health needs.**
- **ACA added new requirements that 501(c)(3) hospitals must conduct a CHNA at least once every three years in order to assess community need and annually file information regarding progress toward addressing identified needs.**
- **This can involve partnerships with other clinical, public health, and population health focused organizations**

MODELS OF COLLABORATION INVOLVING HOSPITALS, PUBLIC HEALTH DEPARTMENTS, AND OTHERS
Improving Community Health through Successful Partnerships

- To identify, compare, and contrast exceptional models of collaboration involving community hospitals, public health departments, and other stakeholders who share commitment to improving community health and determine the key lessons learned from their experience.
- Identify models of collaboration in improving community health that are operational and considered to be highly successful;
- Produce insights that will assist policy makers and leaders of public and private organizations in building strong, successful partnerships designed to improve community health.
- http://www.uky.edu/publichealth/studyOverview.php

Illustrative Coalition Effort: Markey Cancer Center and Cancer Coalitions

Kentucky Regional Comprehensive Cancer Control

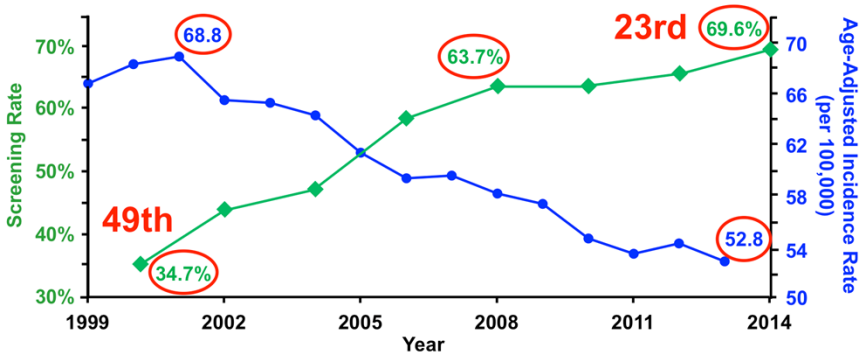
- The Kentucky Cancer *Consortium* (KCC) focuses on multi-regional and state-level efforts in cancer control. KCC is funded through CDC.
- The Kentucky Cancer Program (KCP) is a state-funded, university-affiliated, and community-based regional cancer control program, focused at the regional and local level.
- KCP operates through a network of 13 regional offices staffed by cancer control specialists who lead cancer prevention and control initiatives for all of Kentucky's 120 counties.
- KCP works closely with 15 District Cancer Councils across the state to analyze local cancer data, identify and prioritize the community's cancer needs, and develop interventions/solutions.
- KCP is jointly administered by the University of Kentucky Lucille Parker Markey Cancer Center and the University of Louisville James Graham Brown Cancer Center.

Kentucky Cancer Consortium

- The mission of the Kentucky Cancer Consortium shall be to achieve reductions in the incidence, morbidity and mortality of cancer in Kentucky through a comprehensive, integrated and coordinated approach to cancer control. This approach covers the cancer continuum from prevention, early detection, treatment and care.
- The Consortium is Kentucky's state comprehensive cancer control coalition - a statewide partnership of 70+ diverse organizations united to reduce the burden of cancer in Kentucky.
- The Consortium provides a common forum for like-minded organizations to take collective action. Through group consensus at Consortium meetings, statewide cancer control events, and evaluation, the Consortium determines common priorities, prevents overlap, maximizes resources, and evaluates impact.

Markey Cancer Center Clinical Liaison

Measures of Success: Colorectal Cancer Screening and Incidence



Keys to Success

- Rural focused dissemination and implementation science
  - Extensive formative research, training, resources, funding, technical assistance
- Sustainability
- Coordination of activities and players / silos of funding and initiatives within CRC
  - Benefit of KCC to *bring together and lead* all of these partners
- Innovative use of "other settings", community-based networks and staff, and health communication *in combination* with personal-touch
- Capitalize on community-clinical linkages
- Patient-centered communication / patient navigation

That's all folks!

Questions?

Scutch@uky.edu

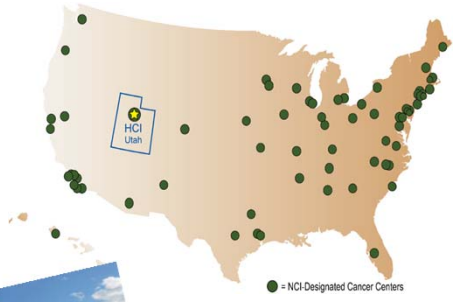
Huntsman Cancer Institute

Using Technology for Recruitment, Retention, Data Collection and Intervention Delivery

Kathi Mooney, PhD, RN
Huntsman Cancer Institute
University of Utah
Salt Lake City, Utah

HUNTSMAN CANCER INSTITUTE UNIVERSITY OF UTAH

HEALTH UNIVERSITY OF UTAH COLLEGE of NURSING



- Utah population- 3 million people
- Utah encompasses nearly 85,000 mi^2
- 96% of Utah is rural (<100 persons/mi^2)
- 70% of Utah is frontier (<7 persons/mi^2)
- Utah is home to 7 Native American tribes/nations

The Huntsman Cancer Institute Catchment Area

- Includes 5 Intermountain West states covering 17% of the US continental landmass
- 30% of patients being treated at the Huntsman Cancer Institute live in rural/frontier communities
- Sparse population densities:
  - Utah = 35.5 people/mi^2
  - Nevada = 26.3 people/mi^2
  - Idaho = 20.0 people/mi^2
  - Montana = 7.1 people/mi^2
  - Wyoming = 6.0 people/mi^2

Recruitment

- Connecting to the target population- trust
- Marketing the opportunity
- Engaging the target population
- Social media
  - Methods of recruitment
  - Examples: Army of Women Susan Love Foundation; Apple/Stanford Heart Study
- Patient-facing portal of the electronic health record
- Video/Video sharing
  - Example: ORIEN Total Cancer Care Cohort
- Combine person-based and technology-based methods

- Social media use issues
  - Network and venue bias
  - Snowball sampling bias
  - Accuracy of reported data
  - Abuse of incentives

Retention

- Automated reminders; encouragement from influentials
- Updates, boosters, newsletters
- Just enough- not too much
- Use of technology to track accrual and retention
  - Research management systems

Technology delivery modes:

- Mobile phone text
- Automated telephone message- smart or not
- Email
- Patient-facing portals of the electronic health record
- Social media
- Telecommunication

Data Collection

- Electronic capture of patient-reported data-
  - Multiple platforms- phone, internet, app, research management systems
  - Ecological Momentary Assessment (EMA)
  - Computer Adaptive Testing (CAT)
  - Electronically Activated Recorder (EAR)
- Automated monitoring- wearable, home, community sensor data
- Telecommunication

- Advantage to collect many data points very quickly

Intervention Delivery

- Multiple platforms
- Treatment fidelity
- Easily adapted
- Scalability

- Use of adaptive designs to test a variety of interventions
- Can combine data collection with intervention delivery

An example of technology-assisted retention, data collection, and intervention delivery

Symptom Care at Home (SCH)- a remote symptom monitoring and automated self-management coaching platform with alerts to clinicians for poorly controlled cancer symptoms

NCI funding: RO1CA120558, R01 CA89474, PO1CA138317
Publications: Mooney et al. Cancer Med 2017; Mar. 6(3):537-546; Mooney et al. Support Care Cancer 2014; 2(9):2343-2350.

Extending Care beyond the Cancer Center Walls

Symptom Care at Home

Telephone based- automated voice response system (IVR)- soon to include web and app platforms

1. **Daily automated monitoring of common symptoms (presence, severity (1-10), drill-down for rapid triage) of patient and caregiver**
2. **Automated algorithm-based patient or caregiver coaching based on reported symptoms and intensity. Short-term and long-term behavioral change coaching**
3. **Automated alerting of clinicians for poorly controlled symptoms- symptom graphs for patterns and guideline-based decision support system for intensifying symptom management**

Significant Benefit for Patients

Calls 5 min. avg. length
90% daily call adherence

For Chemotherapy (n=358)

- **Significantly less symptom severity** than usual care; p <.001
  - **67% less severe symptom days** than UC (8-10 severity, (0-10 scale); p<.001
  - **40% less moderate symptom days** than UC (4-7 severity); p<.001
  - **60% more mild days** than UC (1-3 severity); p=.006
  - **25% more asymptomatic days** than UC;(0- not present) p=.006
- **Benefit extended across geography and race**

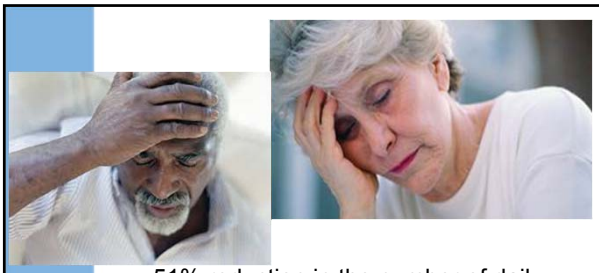
Calls 11 min. avg. length
73% daily call adherence

For Hospice/End of Life (n=298)

- **Significantly less symptom severity for patients** as reported by the family caregiver than usual hospice care; p=.03
- **Rapid onset** of patient benefit compared to usual hospice care; p<.02

Large Mental Health Benefit for Men
Potential value of technology over face to face

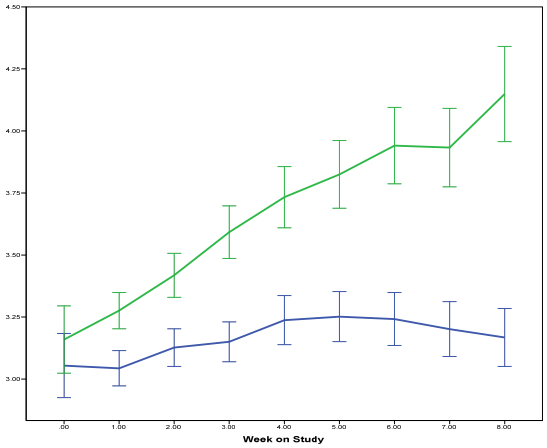
•Men gain a significant mental health advantage from automated monitoring and support for emotional concerns during treatment (SF36 mental health subscale)
• Gender x benefit interaction favoring men (p=.016)
• SCH men gained 5.2 scale points per month (p=.003), 21 scale points overall (4 months)
• 21 scale points overall (0-100)= 11.7 gain in normed T-score where 3.0 is the minimally important difference (MID)



Family Caregiver vitality maintained during caregiving

Lower fatigue, better sleep, and less activity disruption (p<.001)

- 51% reduction in the number of daily moderate-to-severe symptoms for family caregivers (p<.001)
- In SCH (but not UC), caregiver symptom reduction mediated a reduction in patient symptoms, p=.027
- Supporting caregiver's health translates to improved patient symptom outcomes; both are benefited
- 6 months after death of spouse, SCH spouses showed better bereavement outcomes than UC spouses (p=.01)



People will engage and benefit from technology

Hospice Family Caregiver post-intervention interviews:

- I did my calls at the end of the day and it was a release of sorts for me…the time I spent alone at night to reflect on mom's day and how she did.
- Good outlet/input for me-pointing out I wasn't alone and she was not really unusual.
- It gave me a sense of confidence that what I was seeing and feeling was 'normal'.
- It helped calm me when I was having a bad day.

- Being able to anonymously tell someone what is going on made it easier to be helped.
- It felt like someone else was listening to what I had to say. Another person on the team.
- It made me realize I was forgetting who he had been. I was just seeing him as a sick person- that was so helpful so I could change.
- It got me through the hardest time in my life.

Technology can assist in improving health research in small, hidden, and hard to reach populations

- Technology has been used successfully in each and across research phases
- Use technology that is familiar to the target population
- Health technology is a growth industry, we need equivalent advances in health research use
- Engage participants/communities in how to improve the technology
- If it didn't work, don't assume it was the technology- technology is the vehicle not the content or intervention
- There is a need for further research examining best practices in technology use for recruitment, retention, data collection and intervention delivery

Using Technology for Recruitment, Retention, Data Collection and Intervention Delivery
Kathi Mooney

References

1. Alkhaldi G, Hamilton FL, Lau, R, Webster R, Michie S, Murray E (2016) The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 18(1).
2. Cantrell MA, Lupinacci P (2007) Methodological Issues in Online Data Collection. J Adv Nurs 60(5)544-549.
3. Gibbons CJ (2016) Turning the page on pen-and paper questionnaires: combining ecological momentary assessment and computer adaptive testing to transform psychological assessment in the 21st century. Front Psychol 7:1933
4. Gupta A, Calfas KJ, Marshall SJ, Robinson TN, Rock CL, Huang JS, et al (2015) Clinical trial management of participant recruitment, enrollment, engagement, and retention in the SMART study using a Marketing and information Technology (MARKIT) model Contemp Clin Trials. 43:185-195.
5. Hall EW, Sanchez TH, Stein AD, Stephenson R, Ziotorzynska M, Sineath RC, Sullivan PS (2017) Use of videos improves informed consent comprehension in web-based surveys among internet-using men who have sex with men: a randomized controlled trial. J Med Internet Res. 19(3)
6. Kaye J, Whitley EA, Lund D, Morrison M, Teare H, Melham K (2015) Dynamic consent: a patient interface for twenty-first century research networks. Eur J Hum Gen 23:141-146.
7. Kaye J, Curren L, Anderson N, Edwards K, Fullerton K, Kanellopoulou N, et al (2013) From patients to partners: participant-centric initiatives in biomedical research. Nat Rev Genet 13(5): 371-376.
8. Lane TS, Armin J, Gordon JS (2015) Online recruitment methods for web-based and mobile health studies: a review of the literature. J Med Internet Res. 17(7).
9. Marcano Belisario JS, Jamsek J, Huckvale K, O'Donoghue J, Morrison CP, Car J (2015) Comparison of self-administered survey questionnaire responses collected using mobile apps versus other methods. Cochrane Database of Systematic reviews issue 7. Art. No.:MR000042.
10. Mehl MR (2017) The electronically activated recorder (EAR): a method for the naturalistic observation of daily social behavior Curr Direc Psychol Sc. 26(2):184-190.
11. Rupert DJ, Poehlman JA, Hayes JJ, Ray SE, Moultrie RR (2017) Virtual versus in-person focus groups: comparison of costs, recruitment, and participant logistics. J Med Internet Res 19(3).
12. Sonne SC, Andrews JO, Gentilin SM, Oppenheimer S, Obeid J, Brady K, et al (2013) Development and pilot testing of a video-assisted informed consent process. Contemp Clin Trials 36(1):25-31.

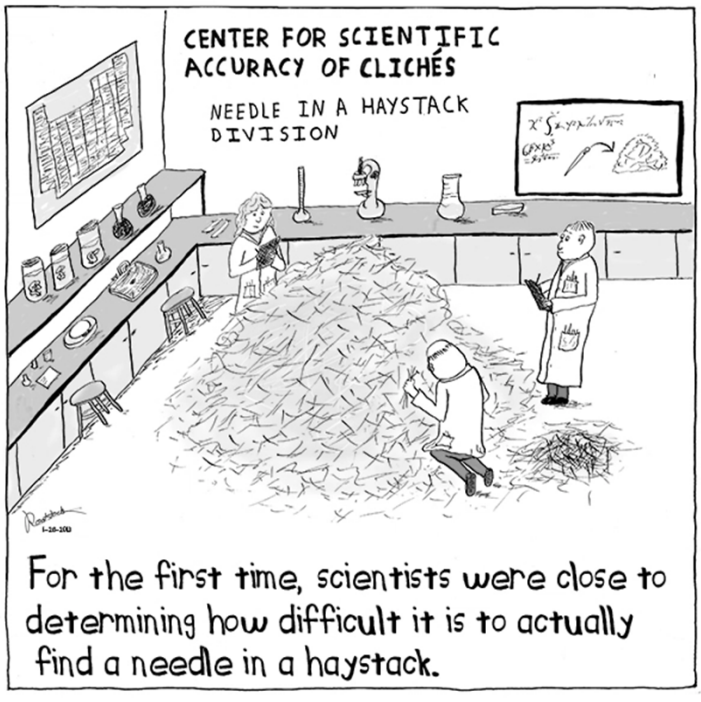
13. Voruganti T, Grunfeld E, Makuwaza T, Bender J (2017) Web-based tools for text-based patient-provider communication in chronic conditions: scoping review. J Med Internet Res. 19(10).
14. Whitaker C, Stevelink S, Fear N (2017) The use of Facebook in recruiting participants for health research purposes: a systematic review. J Med Internet Res. 19(8).
15. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K et al (2016) Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 51(5):833-842.

Dartmouth
GEISEL SCHOOL OF MEDICINE

Dartmouth-Hitchcock
NORRIS COTTON CANCER CENTER

Tracy Onega, PhD, MA, MS
Geisel School of Medicine at Dartmouth & the Norris Cotton Cancer Center

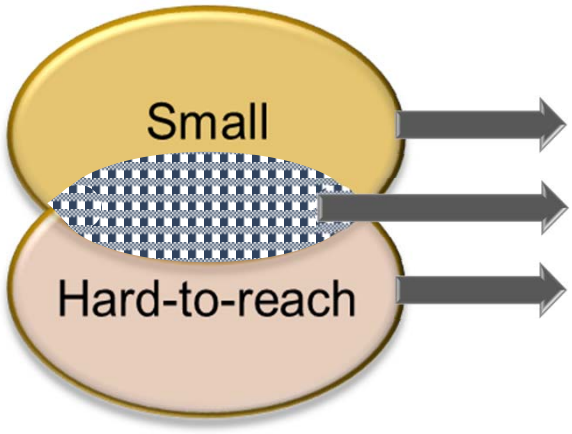
Recruitment, Retention and Collection of Data with a Focus on Small or Hard to Reach Populations



Dartmouth GEISEL SCHOOL OF MEDICINE

Dartmouth-Hitchcock NORRIS COTTON CANCER CENTER

Framing the Issues



Three distinct populations?

With distinct issues and approaches?

Dartmouth GEISEL SCHOOL OF MEDICINE

Dartmouth-Hitchcock NORRIS COTTON CANCER CENTER

Improving Research in Small/HTR Populations ……..a tale of two tasks…….

Identify commonalities to move forward with joint approaches

Identify important distinctions that need to be approached in unique ways

Dartmouth GEISEL SCHOOL OF MEDICINE

Dartmouth-Hitchcock NORRIS COTTON CANCER CENTER

Commonalities in Reaching Populations

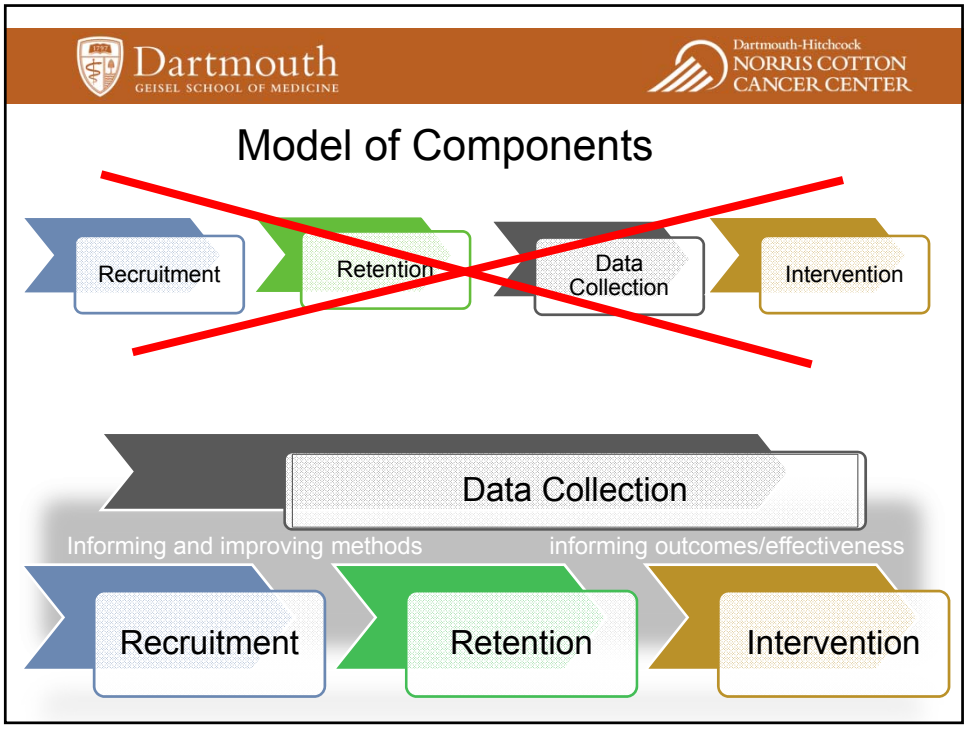
- Mixture (albeit varying) of 'boots on the ground' with remote reach
- Increasingly relying on technology
- Always predicated on knowledge of population
- Data collection / measurement objective(s)
- Must work across phases: recruitment, retention, etc.

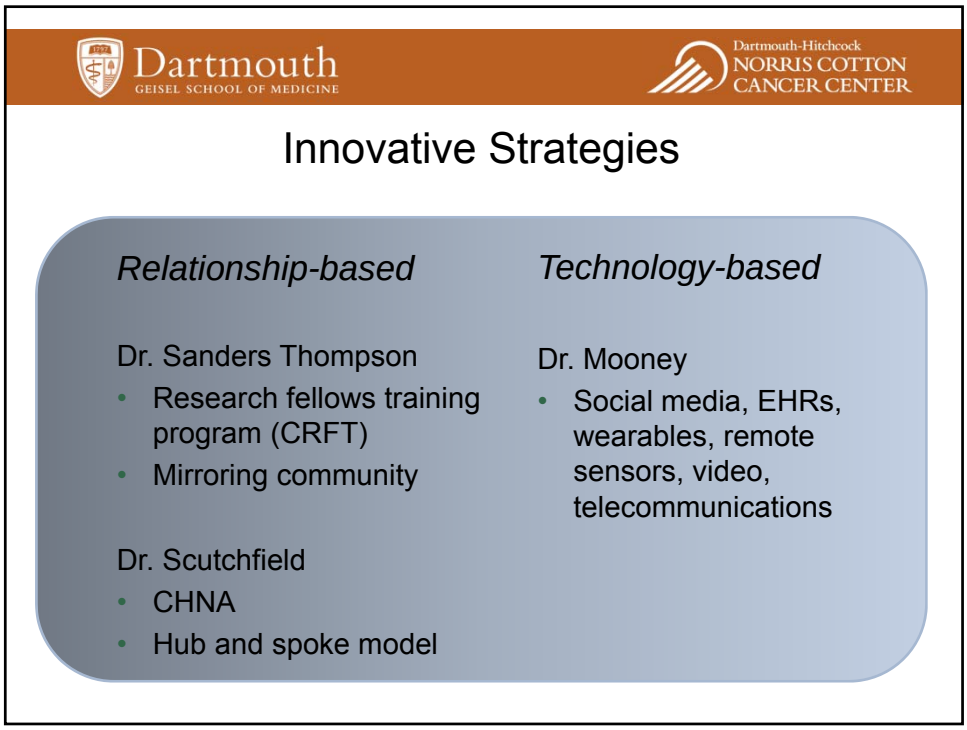
Dartmouth GEISEL SCHOOL OF MEDICINE

Dartmouth-Hitchcock NORRIS COTTON CANCER CENTER

Distinctions in Reaching Populations

- Settings – urban, rural, specific venues, distributed
- Sampling frame – individuals, providers, communities
- Sampling strategy – snowball, RDS, IFWS, etc.
- Technology v. human components variably effective
- Barriers vary: linguistic, cultural, technological, geographic, etc.
- Heterogeneous criteria for "small and/or hard-to-reach"


Dartmouth
GEISEL SCHOOL OF MEDICINE
Dartmouth-Hitchcock
NORRIS COTTON
CANCER CENTER
Model of Components
Recruitment
Retention
Data Collection
Intervention
Data Collection
Informing and improving methods
informing outcomes/effectiveness
Recruitment
Retention
Intervention


Dartmouth
GEISEL SCHOOL OF MEDICINE
Dartmouth-Hitchcock
NORRIS COTTON
CANCER CENTER
Innovative Strategies
Relationship-based
Dr. Sanders Thompson
• Research fellows training program (CRFT)
• Mirroring community
Dr. Scutchfield
• CHNA
• Hub and spoke model
Technology-based
Dr. Mooney
• Social media, EHRs, wearables, remote sensors, video, telecommunications

Dartmouth GEISEL SCHOOL OF MEDICINE

Dartmouth-Hitchcock NORRIS COTTON CANCER CENTER

Posing Key Questions

For Dr. Mooney

- You have a nice example of combining person-based and technology-based methods. Are there populations and/or settings when person precedes technology or vice-versa for best effectiveness?
- You incorporate data collection across phases. This seems crucial, but what unique challenges does this pose?

Dartmouth GEISEL SCHOOL OF MEDICINE

Dartmouth-Hitchcock NORRIS COTTON CANCER CENTER

Posing Key Questions

For Dr. Scutchfield

You gave excellent examples of community partnerships. These efforts seem broadly targeted, which can maximize 'reach'. Can you comment on whether 'casting the wide net' misses some populations of interest, and how you would know.

The 'hub and spoke' model, such as with the Markey Cancer Center and Cancer Coalitions seems to work well. What are its best applications and limitations in terms of reaching small/HTR populations?

Dartmouth GEISEL SCHOOL OF MEDICINE

Dartmouth-Hitchcock NORRIS COTTON CANCER CENTER

Posing Key Questions

For Dr. Sanders Thompson

You gave a wonderful example of matching the right media for your 'market' and knowing your audience

- Is this all done a priori, iteratively, …….
- What are the implications for cost/feasibility and 'getting it right' as well as potentially alienation populations/individuals if you don't tailor correctly and how do you balance that?

For populations for which you can't "go where they are" and/or mirror the audience – what then?

- Virtual v actual "going where they are"
- Can something similar be adapted to online communities – technology/social media-savvy embedded individuals?

Dartmouth GEISEL SCHOOL OF MEDICINE

Dartmouth-Hitchcock NORRIS COTTON CANCER CENTER

Posing Key Questions

For All Speakers

- What are we doing about populations we can't reach? Do we know who and/or where they are?
- Is there a comprehensive compilation of small and hard-to-reach populations, such that we can track/address:
  - Which have been reached and how?
  - Which haven't?
  - For which do we have evidence – or even information – on how to recruit, retain, and intervene?

Dartmouth GEISEL SCHOOL OF MEDICINE

Dartmouth-Hitchcock NORRIS COTTON CANCER CENTER

Where Do We Go From Here?

- What technologies actually work, and according to what factors (age, race, geography, etc.)
- Tall order to determine effective strategies specific to populations, data needs, AND by phase.
  What intentionality should we as researchers bring to this? (need-based prioritization, low hanging fruit, piecemeal, etc.)

Dartmouth GEISEL SCHOOL OF MEDICINE

Dartmouth-Hitchcock NORRIS COTTON CANCER CENTER

Additional Questions

- What existing data & resources can we leverage:
  - Web content mining
  - Existing geospatial or governmental resources
  - Online communities
- What data can/should we generate to inform best strategies?
- How can we best leverage/maximize what we learn?

H

TAB H

SESSION 6: Analysis Techniques for Small Population Research

Presentations in this tab:

Design and Analysis Considerations in Research with Small Samples
Rick H. Hoyle, Duke University

Bayesian Methods for Small Population Analysis
Thomas A. Louis, Johns Hopkins Bloomberg School of Public Health

Estimating the Size of Hidden Populations
Katherine R. McLaughlin, Oregon State University

Design and Analysis Considerations in Research with Small Samples

Rick H. Hoyle
Department of Psychology & Neuroscience
Duke University

NAS Committee on National Statistics
Improving Health Research for Small Populations: A Workshop
January 18-19, 2018

Topics

- When are analyses informative?
- What do we mean by "small"?
- Finite population correction
- Research strategies that address some concerns
- Multivariate models

[2]

Analyses are informative when they . . .

- address the question that motivated the research . . .
  - or, address a narrower or more preliminary question for which an analytic strategy can be justified given *N*
- use data that satisfy the assumptions of the analytic strategy
- are sufficiently powered to detect meaningful effects . . .
  - or, reveal descriptively promising patterns in the data so that a new, more focused and informative study can be run
- produce results likely to generalize to the target population

[3]

What do we mean by "small"?

- Sample size is small when . . .
  - estimates and tests would be unduly influenced by a small number of cases;
  - it falls at or below the minimum required for valid estimates of parameters and/or standard errors;
  - estimation results in nonconvergence or problematic parameter estimates;
  - statistical tests are insufficiently powered to detect meaningful effects.

[4]

Small *and* constrained

- small sample solutions are for circumstances when samples are small and constrained
  - population of cases is small
  - reaching cases requires substantial resources
- proposed solutions are not for circumstances in which sample size is not constrained; in such cases, increasing sample size is the preferred solution
  - the compromises required when using small data are not justified when sample size is small but could be increased with reasonable time and effort

[5]

When constrained by population size

- sampling fraction, $f = \frac{n}{N}$, where n = sample size and N = population size
- $f = 1$ = census
- as f approaches 1, standard error is adjusted downward to reflect reduction in sampling error due to large proportion of population in sample
- as f approaches 0, tests mirror those for samples assumed to be infinitely large
- when $f > .05$, power of statistical test can be improved through use of the finite population correction factor

$$FPC = \sqrt{\frac{N-n}{N-1}}$$

[6]

Finite population correction factor (Cochran, 1977)

$$FPC = \sqrt{\frac{N-n}{N-1}}$$

- applied to standard error for tests of parameter estimates
- example, $\sigma_M = 10$, $N = 200$, n varies

n	f	FPC	σ_M
175	.875	.354	3.54
150	.750	.501	5.01
125	.625	.614	6.14
100	.500	.709	7.09
75	.325	.793	7.93
50	.250	.868	8.68
25	.125	.938	9.38
10	.050	.977	9.77

[7]

Finite population correction factor

- can be used for study planning when working with finite population
- determine required sample size, n_r, if assuming infinite population sampled with replacement
- derive sample size adjusted for planned use of FPC, n_a

$$n_a = \frac{n_r}{1 + \frac{(n_r - 1)}{N}}$$

N	n_r	n_a	n_a/n_r
200	150	86	.57
200	125	77	.62
200	100	69	.69
200	75	55	.73
200	50	41	.82

[8]

Finite population correction factor

- assumes random sampling without replacement
- accounts for reduction in sampling error as *f* increases toward 1.0
- allows inference about state of population at that point in time; not prediction of state of other populations or state of current population at a later point in time

[9]

Research strategies for addressing some concerns

$$T = \frac{parameter\ estimate}{standard\ error}$$

- options
  - increase parameter estimate
  - decrease standard error

[10]

Increasing parameter estimate

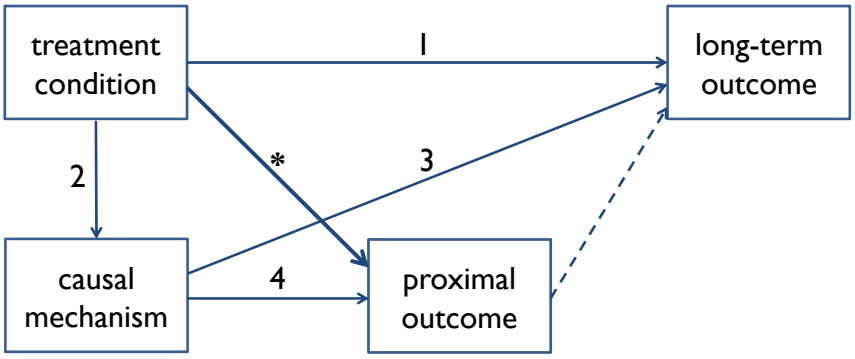
$$T = \frac{parameter\ estimate}{standard\ error}$$

- strengthen treatment condition
  - increase "dosage" of treatment
  - choose inactive control condition
- focus treatment directly on causal mechanism
- choose reliable but sensitive outcome measure
  - minimize attenuation due to unreliability
  - maximize odds of detecting difference or change by using outcome that is responsive to change in conditions

[11]

Increasing parameter estimate

$$T = \frac{parameter\ estimate}{standard\ error}$$



[12]

Decreasing standard error

$$T = \frac{parameter\ estimate}{standard\ error}$$

- examples of standard error

$$\sigma_M = \sqrt{\frac{s^2}{n}}$$

$$\sigma_{M_1 - M_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- options when sample size is constrained

[13]

Leave no data unanalyzed

$$T = \frac{parameter\ estimate}{standard\ error}$$

$$\sigma_M = \sqrt{\frac{s^2}{n}}$$

- ensure that the full sample is the analysis sample
  - minimize attrition in prospective studies
  - use modern methods for managing missing data
    - multiple imputation
    - model-based methods
      - e.g., FIML in SEM
      - incorporate missing data mechanism in model
      - inclusion of auxiliary variables

[14]

Account for unexplained variance in outcome

$$Y = \beta_0 + \beta_1 x_1 + e_F$$

$$F = \frac{(e_R - e_F)/(df_R - df_F)}{e_F/df_F}$$

- reduce e_F by including covariates associated with e_F (i.e., variance in *Y* not accounted for by predictors of interest)

$$Y = \beta_0 + \beta_1 x_1 + \underbrace{\beta_2 x_{c1} \ldots \beta_i x_{cj}}_{\text{covariates}} + e_F$$

[15]

Multivariate models

- multilevel modeling
- growth modeling
- structural equation modeling
- person-level dynamic modeling

[16]

Multivariate models

- generally considered large sample methods
- yet, increasing evidence of use with small samples
- reviews of behavioral science applications
  - *N* of higher-level groups < 30 for 21% of MLM studies
  - *N* < 100 for 33% of growth models
  - *N* < 100 for 40% and < 200 for 60% of EFA studies
  - *N* < 100 for 18% of SEM studies
- suggests research questions that . . .
  - require data that are clustered;
  - concern unobserved influences;
  - focus on patterns of change over repeated assessments

[17]

Multilevel modeling

- making the following assumptions
  - continuous measures
  - ICC ≈ .20
  - 4-8 predictors
  - no missing data
  - 2 or fewer cluster-level random effects
- < 40 clusters is considered small
- < 20 clusters should not be analyzed using standard methods
- clusters should have at least 5 observations

[18]

Multilevel modeling

- solutions (McNeish, 2017)
  - restricted maximum likelihood (REML)
  - REML with Kenward-Roger correction
  - wild cluster bootstrap

[19]

Growth modeling

- making the following assumptions
  - continuous measures
  - 4-8 observations per person
  - random intercepts and slopes
  - linear growth
  - < 5 time-varying covariates
- *N* < 100 is considered small
- *N* < 50 should not be analyze using standard methods

[20]

Growth modeling

- solutions
  - depends on analytic framework
  - typically SEM, discussed next

[21]

Structural equation modeling

- making the following assumptions
  - continuous measures
  - near-normal multivariate distribution
- and considering the following model characteristics
  - magnitude of loadings on latent variables
  - number of latent variables
  - number of indicators per latent variable
- *N* < 200 is considered is small for moderate loadings (.5-.7) and moderately complex models (3-4 indicators of 3-4 latent variables)
- *N* < 100 should not be analyzed using standard methods

[22]

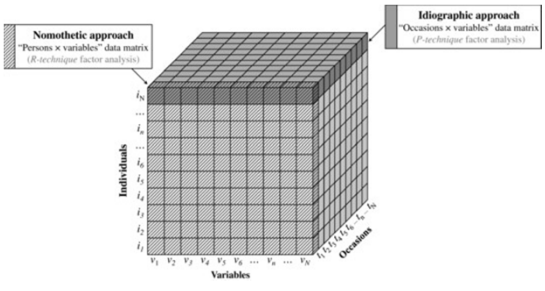
Structural equation modeling

- solutions
  - do not interpret raw "χ^2" value; use Bartlett, Yuan, or Swain correction, which include *N*; Yuan-correction performs well at *N*s of 25 and 50
  - use adjunct fit indices that perform well with small *N*s—comparative fit index
  - limit model complexity

[23]

Person-level dynamic modeling

- p-technique factor analysis—modeling of intraindividual variability across intensive, repeated observations



[24]

Person-level dynamic modeling

- traditional p-technique factor analysis
  - "sample" is a set of observations of one person (p) on a set of variables (e.g., measure of affect administered daily for three months yielding ≈ 90 observations)
  - factor analysis of latent structure for the person
  - multiple people can be "chained"
  - limitation: does not account for effects of time

[25]

Person-level dynamic modeling

- dynamic p-technique factor analysis (Nelson et al., 2011)
  - explicitly incorporates time to allow for modeling of intraindividual change over time
  - uses lagged covariance matrices, permitting modeling of
    - within-lag covariances between variables
    - autoregressive covariances (stability)
    - cross-lagged covariances (prospective relations between variables)
  - person-level data can be chained or analyses done using multigroup SEM

[26]

Person-level dynamic modeling

"DPT can allow for complex models that match the complexity of research hypotheses. Simply stated, DPT allows researchers to conduct sophisticated analyses, despite small numbers of participants. . . . Repeated measurement of a small number of individuals over time is often more feasible than studying large numbers of participants."

Nelson, Aylward, & Rausch (2011)

[27]

Summary

- the outcome of statistical analysis/modeling should be informative; informative results are challenging to produce for small sample data
- what qualifies as a small sample varies as a function of a number of features of a study
- when *N* is small and constrained, the goal is to maximize the yield of the study through careful consideration of design, measurement, and analysis options
- health research often concerns patterns, processes, or structures that require the use of multivariate methods; such methods can sometimes produce informative results when *N* is small

[28]

CNSTAT Workshop (January 18-19, 2018):
Improving Health Research for Small Populations

Bayesian Methods for Small Population Analysis

Thomas A. Louis, PhD
Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health
www.biostat.jhsph.edu/~tlouis/
tlouis@jhu.edu

Pre-Summary

- Seldom can inferences from small populations stand on their own, because estimates are unstable (have low precision)
  - Also, large-sample assumptions/conveniences may not apply
- Modeling or other stabilization/(information enhancement) is necessary, and there are a wide variety of strategies, including:
  - Aggregation
  - Regression both within and across populations
  - Hierarchical (Bayesian/EB) modeling to 'borrow information' within and between data sources
  - Trimming survey weights
- Stabilization/enrichment targets include,
  - Estimated regression slopes and residual variances
  - A control group, using historical data
  - Clinical trial subgroup estimates (Henderson et al., 2016)
  - Transporting, e.g., adults $\longrightarrow$ children
  - Small Area (Domain) estimates (SAEs)
  - Estimated SMRs and the challenges of low information
  - Survey weights (Gelman, 2007)
- The Bayesian formalism is effective in meeting these goals

Preview

The Bayesian formalism
- Modern Contraceptive Rates in Uganda
- Inferences on rates of bone loss
- Stabilizing variance estimates

Combine, don't pool
- Historical controls in carcinogenicity testing

Making use of Big Data
- Embed a high-resolution study in a larger, lower-resolution one

Design-based inference loosens its grip on the survey world
- Combine survey estimates: SIPP aided by the ACS
- Small Area Income and Poverty Estimates (SAIPE)
- Alternative language determinations as required by Section 203 of the voting rights act

Health Provider Profiling
- Shrinkage/stabilization can be controversial
- The challenges of low information

Closing

Trading off Variance and Bias (for the linear model)

- K units (individuals, clusters, institutions, studies, regions, domains, ...)
- Each with an underlying feature of interest (θ_k):
  - Poverty Rate, RR/SMR, TxEff, Residual Variance, ...
- A direct (unbiased) estimate of it (Y_k), with estimated variance ($\hat{\sigma}_k^2$)
- Unit-specific attributes $\mathbf{X}_k$ (tax data, age, exposure) produce,

$$\begin{aligned} \text{regression prediction} &= \hat{\boldsymbol{\beta}} X_k \quad (\text{e.g., } \hat{\beta}_0 + \hat{\beta}_1 X_k) \\ \text{residual} &= Y_k - \hat{\boldsymbol{\beta}} X_k \end{aligned}$$

- Inviting three choices for estimating the θ_k:

  Direct: Use the Y_k (unbiased, but possibly unstable)
  Regression: Use the regression (stable, but possibly biased)
  Middle ground: A weighted average of Regression and Direct

$$\begin{aligned} \hat{\theta}_k &= \text{regression prediction} + (1 - \hat{B}_k) \times \text{residual} \\ &= \hat{\boldsymbol{\beta}} \mathbf{x}_k + (1 - \hat{B}_k) \cdot (Y_k - \hat{\boldsymbol{\beta}} \mathbf{x}_k) \\ \hat{B}_k &= \hat{\sigma}_k^2 / (\hat{\sigma}_k^2 + \hat{\tau}^2) \\ \hat{\tau}^2 &= \text{residual/unexplained variance, model lack of fit} \end{aligned}$$

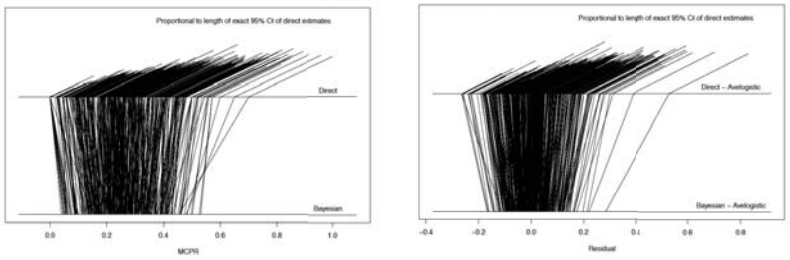
- For general models use the Bayesian formalism
  (Carlin and Louis, 2009; Gelman et al., 2013; Kadane, 2015)

Small Area Estimates

Modern Contraceptive Prevalence Rate (MCPR) in Uganda

- Performance Monitoring and Accountability 2020 (PMA2020) survey data
- Woman-specific information:≈ 13,100 inputs
  (109 areas) ×(4 rounds) ×(≈ 30 women per round)
- Logistic regression with covariates and an area-specific random effect

Direct → Bayes (Direct - Regression) → (Bayes - Regression)



Age-specific rate of bone loss Hui and Berger (1983)

- Woman/age–specific, locally linear slope estimates (+ means loss)
- Short follow-up, so slope and residual variance estimates are imprecise
- Use empirical Bayes to calm the variation

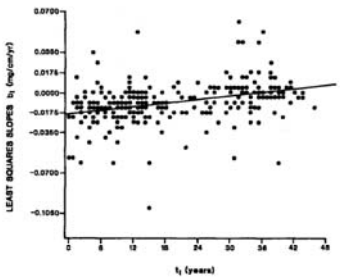


Figure 2. Individual least squares estimates of rate of bone loss b_i vs. t_i, where the t_i are suitably chosen points in the follow-up intervals.

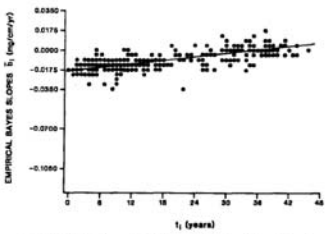


Figure 4. Individual empirical Bayes estimates of rate of bone loss $\tilde{b}_i$ versus t_i.

Stabilizing Variance Estimates

(Less controversial than stabilizing attributes of primary interest)

- The woman/age-specific, estimated residual variance is $\hat{\sigma}_k^2$
- With degrees of freedom, $d_k = \#\{\text{measurements}\} - 2$
- The σ_k^2 come from a (Gamma) prior with,
  - Estimated mean $\hat{m}$
  - Estimated effective sample size $\hat{M}$
- The empirical Bayes estimates are,

$$\begin{aligned} \tilde{\sigma}_k^2 &= \hat{m} + (1 - B_k)(\hat{\sigma}_k^2 - \hat{m}) \\ B_k &= \hat{M}/(\hat{M} + d_k) \\ \tilde{d}_k &\approx B_k d_+ + (1 - B_k) d_k \end{aligned}$$

- The distribution of $\tilde{\sigma}_k^2$ isn't chi-square, and a fully Bayesian analysis produces the appropriate (hybrid) distribution

Historical Controls (combine, don't 'pool')

	C	E	Total
Tumor	0	3	3
No Tumor	50	47	97
	50	50	100

- Fisher's exact one-sided P = 0.121
- But, pathologists get excited:
  - The 3 tumors are 'Biologically Significant'
- Statisticians protest:
  - But, they aren't 'Statistically Significant'

We need to stop using these terms!

Include Historical Data

- There may be historical information for the same species/strain, same Lab, recent time period with 0 tumors in 450 control rodents
- Pooling gives,

Pooled Analysis			
	C	E	Total
Tumor	**0**	3	3
No Tumor	500	47	547
	500	50	550

- Fisher's exact one-sided P $\doteq$ .0075
- Convergence between biological and statistical significance
- The Bayesian formalism should be be used to bring in history, in general, giving it only partial credit

Bringing in history

Identify 'relevant' experiments, and use the Bayesian formalism

- Control rates come from a Beta distribution with

$$\begin{aligned} \text{mean} &= \mu \\ \text{Variance} &= \frac{\mu(1-\mu)}{M+1} \end{aligned}$$

- Use all the data to produce $\hat{\mu}$ and $\widehat{M}$
- Augment concurrent control group by pseudo-data with mean $\hat{\mu}$ and sample size $\widehat{M}$ (adaptive down-weighting of history)
- Female, Fisher F344 Male Rats, 70 historical experiments (Tarone, 1982)

Tumor	N	$\widehat{M}$	$\hat{\mu}$	$\frac{\widehat{M}}{N}$
Lung	1805	513	.022	28.4%
Stromal Polyp	1725	16	.147	0.9%

See Ibrahim et al. (2014) for a clinical trials example

Big Data and Data Synthesis

Chatterjee et al. (2016)

- Have a fine-grained study, with internally valid estimates
- And have stable, reduced dimension, external information
  - e.g, a joint distribution of a subset of the within-study variables
- Constrain the within-study estimates to be compatible with the externally determined (marginal) distributions in the spirit of,
  - Stabilizing estimates in a contingency table by 'benchmarking' to marginal distributions estimated from other data
  - Using external prevalence data so that a case-control study can estimate relative risk (RR) or a risk difference
- The key issue is whether stochastic features of the external data are sufficiently similar to those for the internal data so that in the end MSE is reduced
- Resonates with external validity, representativity of a sample, transporting within-sample estimates to a reference population, ...

  See, Keiding and Louis (2016); Keiding and Louis (2018)
  Pearl and Bareinboim (2014); National Academies (2017)

Combining Surveys

With other data, see Lohr and Raghunathan (2017)

Combining Estimates from Related Surveys via Bivariate Models

(Application: using ACS estimates to improve estimates from smaller U.S. surveys)

William R. Bell and Carolina Franco, U.S. Census Bureau

2016 Ross-Royall Symposium

February 26, 2016

Application I: 2010 Disability Rates for U.S. States: SIPP borrowing from ACS

y_{1i} = SIPP disability estimate, y_{2i} = ACS disability estimate

Smoothing of SIPP direct sampling variance estimates is applied.

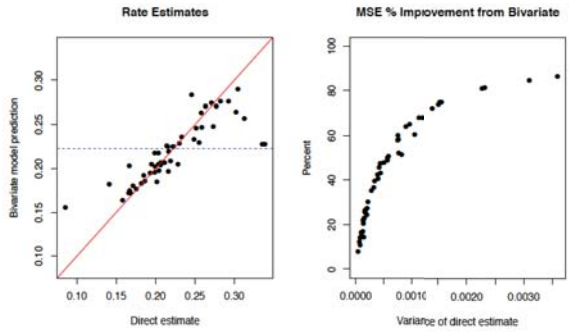
$\hat{\rho}$= **.82**

- Univariate shrinkage yields an MSE decrease of $2\% - 67\%$ from direct, with a median of 19%
- The MSE decrease from bivariate vs. univariate model is $6\% - 59\%$ with a median of 29%
- The MSE decrease from bivariate vs. direct is **$8 - 86\%$, with a median decrease of 43%**

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Disability Rates for U.S. States, 2014

Bivariate model for SIPP and ACS estimates



JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

SAIPE and Section203

(Bayesian) hierarchical modeling is essential

SAIPE: Small Area income and Poverty Estimates (Bell et al., 2016)

- Allocate $12+ billion a year
- 'Direct' Data are from the ACS and other surveys
- Xs are tax rates, etc.

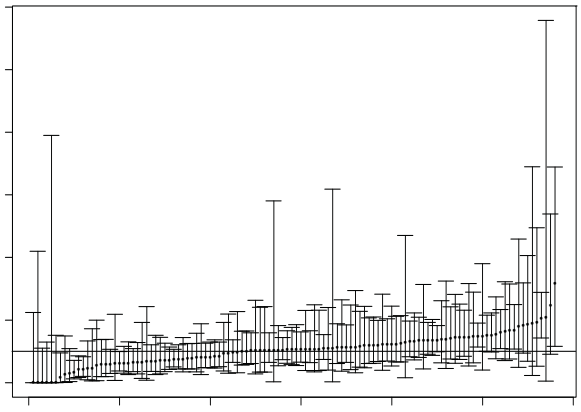
Section 203 of the Voting Rights Act (Slud and Ashmead, 2017)

- In order to make the determinations, it is necessary to estimate the total population of voting age persons who are citizens, of citizens who have limited English proficiency, and of citizens with limited English proficiency who are illiterate in approximately 8000 jurisdictions, 570 American Indian and Alaska Native Areas (AIA/ANAs), and 12 Alaska Native Regional Corporations (ANRCs), separately for 68 Language Minority Groups
- Total, potential, estimation domains $\approx 560{,}000 = 70 \times 8000$
- Allowed to use only the census and the ACS

USRDS, SMRs: MLEs and exact CIs

(1, 41, 81, . . . ordered MLEs)

- SMR = Standardized Mortality Ratio = observed/expected deaths

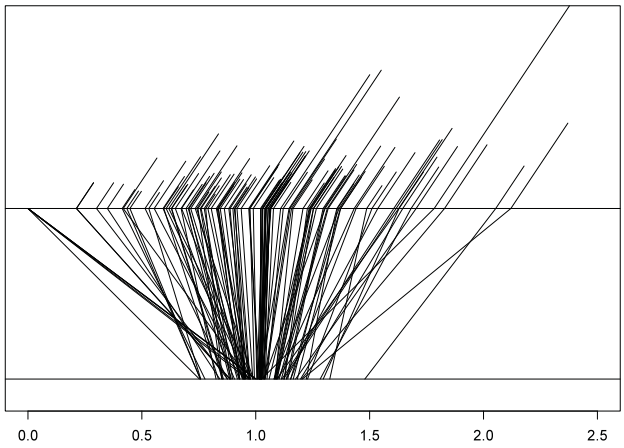


- Sampling variability has a wide range over units

Bayesian analysis, ρ = SMR (Lin et al., 2009)

$\hat{\rho}^{mle}$, $\hat{\rho}^{pm}$, SE($\hat{\rho}^{mle}$) using USRDS dialysis data

middle = MLE :: whisker = SE :: bottom = Posterior Mean



Shrinkage can be controversial (Normand et al., 2016)

- Direct estimates with greatest uncertainty are shrunken closest to the regression surface, potentially conferring undue benefits or punishments
- Especially troublesome when the model is mis-specified (always true!) and sample size is informative so that the degree of shrinkage is 'connected at the hip' to the underlying truth
- Standard model fitting gives more weight to the stable units, consequently the units that 'care about' the regression model have less influence on it
- Recent approaches increase the weights for the relatively unstable units, paying some variance, but improving estimation performance for mis-specified models (Chen et al., 2015; Jiang et al., 2011)

Closing

- Statistics has always been about combining information; think $\bar{X}$
- Careful development and assessment is necessary, and the Bayesian formalism is an effective aid to navigation and inferential framework
- Advances in data science (annotation, harmonization, storage and retrieval), computing (hardware & software), and statistical methods; make evermore relevant,

> *All of statistics involves combining evidence over basic units to make inferences for a population. The current challenge involves broadening the scope of inputs and inferences in a scientifically valid and credible manner. Development and application of these meta-modeling strategies will challenge and inform in the next and subsequent decades.* *(Louis, 1989)*

- However,

  Space-age procedures will not rescue stone-age data

#thank you

Bibliography

Bell, W., Basel, W., and Maples, J. *Analysis of Poverty Data by Small Area Estimation*, chapter An Overview of the U. S. Census Bureau's Small Area Income and Poverty Estimates Program. John Wiley & Sons (2016).

Carlin, B. P. and Louis, T. A. *Bayesian Methods for Data Analysis, 3rd edition*. Boca Raton, FL: Chapman and Hall/CRC Press, 3$^{\text{nd}}$ edition (2009).

Chatterjee, N., Chen, Y. H., Maas, P., and Carroll, R. J. "Constrained Maximum Likelihood Estimation for Model Calibration Using Summary-level information from External Big Data Sources (with discussion)." *J Am Stat Assoc*, 111(513):107–131 (2016).

Chen, S., Jiang, J., and Nguyen, T. "Observed Best prediction for small area counts." *Journal of Survey Statistics and Methodology*, 3:136–161 (2015).

Gelman, A. "Struggles with survey weighting and regression modeling." *Statistical Science*, 22(2):153–164 (2007).

Gelman, A., Carlin, J., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. *Bayesian Data Analysis, 3rd edition*. Boca Raton, FL: Chapman and Hall/CRC Press (2013).

Henderson, N., Louis, T., Wang, C., and Varadhan, R. "Bayesian Analysis of Heterogeneous Treatment Effects for Patient-Centered Outcomes Research." *Health Services and Outcomes Research Methodology*, 16:213–233 (2016).

Hui, S. L. and Berger, J. O. "Empirical Bayes estimation of rates in longitudinal studies." *Journal of the American Statistical Association*, 78:753–759 (1983).

Ibrahim, J., Chen, M., Lakshminarayanan, M., Liu, G., and Heyse, J. "Bayesian probability of success for clinical trials using historical data." *Statistics in Medicine*, 34:249–264 (2014).

Jiang, J., Nguyen, T., and Rao, J. S. "Best Predictive Small Area Estimation." *Journal of the American Statistical Association*, 106:732–745 (2011).

Kadane, J. "Bayesian Methods for Prevention Research." *Prevention Science*, 16:1017–1025 (2015).

Keiding, N. and Louis, T. "Web-based Enrollment and other types of Self-selection in Surveys and Studies: Consequences for Generalizability." *Annual Review of Statistics and Its Application*, 5:to appear (2018).

Keiding, N. and Louis, T. A. "Perils and potentials of self-selected entry to epidemiological studies and surveys (with discussion and response)." *J Roy Statist Soc, Ser A*, 179:319–376 (2016).

Lin, R., Louis, T., Paddock, S., and Ridgeway, G. "Ranking of USRDS, provider-specific SMRs from 1998-2001." *Health Services and Outcomes Reseearch Methodology*, 9:22–38 (2009).

Lohr, S. and Raghunathan, T. "Combining Survey Data with Other Data Sources." *Statistical Science*, 32:293–312 (2017).

Louis, T. "Meta-modeling." In *Challenges for the 90's, ASA Sesquicentennial visioning* (1989).

National Academies. *Federal Statistics, Multiple Data Sources and Privacy Protection: Next Steps*. National Academies Press (2017).

Normand, S.-L., Ash, A. S., Fienberg, S. E., Stukel, T., Utts, J., and Louis, T. A. "League Tables for Hospital Comparisons." *Annual Review of Statistics and Its Application*, 3:21–50 (2016).

Pearl, J. and Bareinboim, E. "External Validity: From do-calculus to Transportability across Populations." *Statistical Science*, 29:579–595 (2014).

Slud, E. and Ashmead, R. "VRA Section 203 Determinatinos: Statistical Methodology Summary." Technical report, U. S. Census Bureau (2017).

Tarone, R. "The use of historical control information in testing for a trend in proportions." *Biometrics*, 38:215–220 (1982).

Bayesian Methods for Small Population Analysis
Thomas A. Louis

References

Bell, W., Basel, W., and Maples, J. Analysis of Poverty Data by Small Area Estimation, chapter An Overview of the U. S. Census Bureau's Small Area Income and Poverty Estimates Program. John Wiley & Sons (2016).

Carlin, B. P. and Louis, T. A. Bayesian Methods for Data Analysis, 3rd edition. Boca Raton, FL: Chapman and Hall/CRC Press, 3nd edition (2009).

Chatterjee, N., Chen, Y. H., Maas, P., and Carroll, R. J. "Constrained Maximum Likelihood Estimation for Model Calibration Using Summary-level information from External Big Data Sources (with discussion)." J Am Stat Assoc, 111(513):107–131 (2016).

Chen, S., Jiang, J., and Nguyen, T. "Observed Best prediction for small area counts." Journal of Survey Statistics and Methodology, 3:136–161 (2015).

Gelman, A. "Struggles with survey weighting and regression modeling." Statistical Science, 22(2):153–164 (2007).

Gelman, A., Carlin, J., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. Bayesian Data Analysis, 3rd edition. Boca Raton, FL: Chapman and Hall/CRC Press (2013).

Henderson, N., Louis, T., Wang, C., and Varadhan, R. "Bayesian Analysis of Heterogeneous Treatment Effects for Patient-Centered Outcomes Research." Health Services and Outcomes Research Methodology, 16:213–233 (2016).

Hui, S. L. and Berger, J. O. "Empirical Bayes estimation of rates in longitudinal studies." Journal of the American Statistical Association, 78:753–759 (1983).

Ibrahim, J., Chen, M., Lakshminarayanan, M., Liu, G., and Heyse, J. "Bayesian probability of success for clinical trials using historical data." Statistics in Medicine, 34:249–264 (2014).

Jiang, J., Nguyen, T., and Rao, J. S. "Best Predictive Small Area Estimation." Journal of the American Statistical Association, 106:732–745 (2011).

Kadane, J. "Bayesian Methods for Prevention Research." Prevention Science, 16:1017–1025 (2015).

Keiding, N. and Louis, T. "Web-based Enrollment and other types of Self-selection in Surveys and Studies: Consequences for Generalizability." Annual Review of Statistics and Its Application, 5: to appear (2018).

Keiding, N. and Louis, T. A. "Perils and potentials of self-selected entry to epidemiological studies and surveys (with discussion and response)." J Roy Statist Soc, Ser A, 179:319–376 (2016).

Lin, R., Louis, T., Paddock, S., and Ridgeway, G. "Ranking of USRDS, provider-specific SMRs from 1998-2001." Health Services and Outcomes Reseearch Methodology, 9:22–38 (2009).

Lohr, S. and Raghunathan, T. "Combining Survey Data with Other Data Sources." Statistical Science, 32:293–312 (2017).

Louis, T. "Meta-modeling." In Challenges for the 90's, ASA Sesquicentennial visioning (1989). National Academies. Federal Statistics, Multiple Data Sources and Privacy Protection: Next Steps. National Academies Press (2017).

Normand, S.-L., Ash, A. S., Fienberg, S. E., Stukel, T., Utts, J., and Louis, T. A. “League Tables for Hospital Comparisons.” Annual Review of Statistics and Its Application, 3:21–50 (2016).

Pearl, J. and Bareinboim, E. “External Validity: From do-calculus to Transportability across Populations.” Statistical Science, 29:579–595 (2014).

Slud, E. and Ashmead, R. “VRA Section 203 Determinatinos: Statistical Methodology Summary.” Technical report, U. S. Census Bureau (2017).

Tarone, R. “The use of historical control information in testing for a trend in proportions.” Biometrics, 38:215–220 (1982).

Estimating The Size Of Hidden Populations

CNSTAT Workshop on Improving Health for Small Populations, January 18-19, 2018

Katherine R. McLaughlin
Assistant Professor, Department of Statistics

mclaugka@oregonstate.edu

http://stat.oregonstate.edu/~mclaugka/

Outline

- What is a hidden population?
- Challenges of estimating the size of hidden populations
- Population size estimation methods
  - Capture-recapture methods
  - Object and service multipliers
  - Network scale-up methods
  - Successive Sampling-Population Size Estimation (SS-PSE) for respondent-driven sampling

1

Hidden Populations

- Hidden populations may also be called:
  - Hard-to-reach
  - Hard-to-sample
- Members of hidden populations may engage in behaviors that are sometimes illegal or stigmatized and thus may tend to avoid disclosure of their membership and be unwilling to participate in surveys.
- Examples:
  - Key populations at high risk for HIV, such as female sex workers (FSW), people who inject drugs (PWID), men who have sex with men (MSM)

2

Why Populations Size Estimation?

- To assess the existence or magnitude of an issue relating to the population
- To assess how resources should be allocated for better program planning and management
- To aid other estimation methods for these populations, which may require knowledge of *N*
- If repeated over time, to assess population dynamics

3

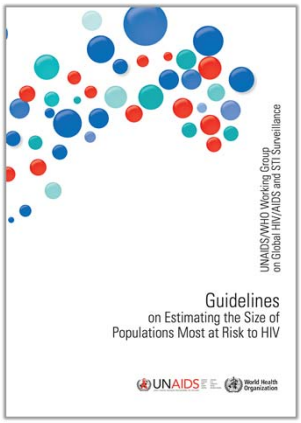
Challenges of Population Size Estimation

- A sampling frame may not exist
- Members of hidden populations may not identify themselves and members of the general population may not know whether or not their friends are members of the hidden population
- Those who participate in a sample may be different than those that do not
  - May be more likely to observe people who are more visible/ highly connected
  - Non-participants may be more isolated or even completely separate from those who do participate
- Populations are dynamic – both in time/space, and membership
  - Timing matters for methods that rely on two samples

4

Current Methods for PSE

- No "gold standard" currently exists and many methods have been used, each with different strengths and weaknesses
- The particular approach chosen will depend on the population of interest and resources available



UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance (2010). Guidelines on Estimating the Size of Populations Most At Risk to HIV.

5

How to Reach a Hidden Population

If we have no sampling frame, what do we do?

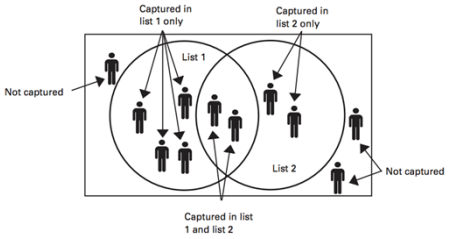
- Rely on a general population survey.
  - Inefficient if we only are interested in the hidden population.
  - Individuals may be unlikely to disclose status as a member of hidden population, and people may not know their friends' statuses.
- Venue-based / time-location sampling: identify locations where members of a hidden population are likely to congregate. Sample locations instead of people.
  - May be difficult to identify a list of venues.
  - May miss individuals who do attend venues.
- Respondent-driven sampling: identify a few "seed" members of the hidden population, use restricted peer-recruitment to grow sample chains.
  - May be biased by initial choice of seeds, volunteerism, dependence between individuals

6

Capture-Recapture

- Procedure
  1. Map all the sites where the population can be found
  2. Go to the sites and "tag" all the members of the population at the site
  3. Return to the sites at a later date and retag all members of the population
  4. Record size of each sample and overlap

Figure 3.2. Illustrating the capture-recapture method



Guidelines on Estimating the Size of Populations Most At Risk to HIV, p.17.

$$\widehat{N} = \frac{n_1 n_2}{m}$$

- Simplistic and requires many assumptions, e.g.
  - Every member has an equal chance of being sampled
  - Matching is reliable
  - The two samples are independent

7

Multiplier Methods

- Service Multiplier Method (SMM) or Unique Object Multiplier Method (UOM)
- Relies on two sources of data
  1. A count of population members who received some service (e.g. attended a clinic or program) or object (unique, memorable)
  2. A representative survey of the population, such as RDS. In the survey, ask each individual if they received the service or object

$$\hat{N} = \frac{\#\ of\ people\ receiving\ the\ service\ or\ object}{\%\ of\ sample\ who\ reported\ receiving\ the\ service\ or\ object}$$

8

Multiplier Methods

- Challenges
  - Requires that the two data sources be independent
  - Obtaining a random sample of a population lacking a sampling frame
  - Timing between the service/object distribution and the sample
  - Everyone receiving the service/object must be a member of the hidden population

9

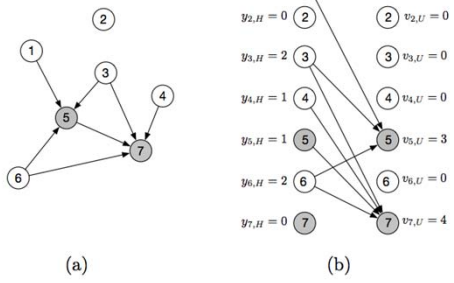
Network Scale-Up Method (NSUM)

- Procedure:
  1. In a general population survey, ask how many individuals each person knows and how many of those are in the hidden population.
  2. Proportion of respondents' contacts who are members of the hidden population is assumed to be equal to the population proportion. Multiply this by the known general population size.
- Requires the assumption that people in the general population are aware of whether or not their network members are members of the hidden population.
- Assumes network connections are formed at random.

10

Network Scape-Up Method (NSUM)

- Generalized scale-up estimator
  - Relies on both a general population survey and a hidden population survey (RDS)
  - Total out-reports equals total in-reports
  - Still assumes that hidden population members have aggregate awareness about visibility



Feehan, D. M. and Salganik, M. J. (2016). Generalizing the Network Scale-Up Method: A New Estimator for the Size of Hidden Populations.

11

Successive Sampling-Population Size Estimation (SS-PSE)

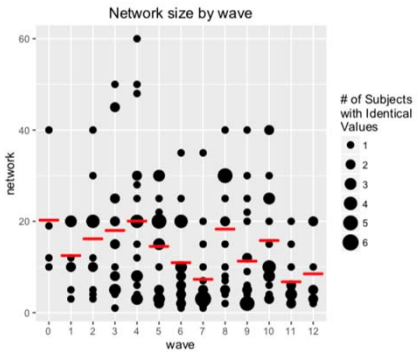
- Use with data collected via one RDS study
  - Cost-effective
  - Can be done retroactively
- Bayesian framework where prior information about population size can be incorporated
  - Good way to combine data from different sources
  - Statistical model for uncertainty in estimates

Handcock, M. S., Gile, K. J., and Mar, C. M. (2014, 2015)

12

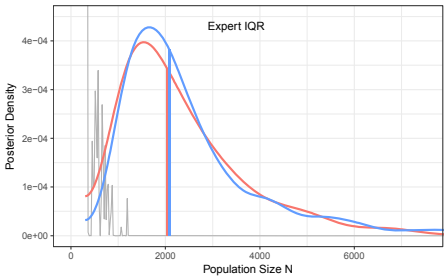
Successive Sampling-Population Size Estimation (SS-PSE)

- Conceptual overview
  - People that are more *visible* (tend to have larger network size) are more likely to be sampled in RDS, and be sampled earlier
  - Consider network size by wave
    - If the frequency of larger network sizes decreases over RDS waves, the population is likely being depleted
      - Population size likely larger
    - If the frequency of larger network sizes does not decrease, there are still many people who have not yet been sampled
      - Population size likely smaller



13

Successive Sampling-Population Size Estimation (SS-PSE)



Example posterior distribution for population size

- Challenges:
  - Network sizes in RDS may not contain a lot of information about population size
  - Relies on quality of RDS data
  - Possibility of inconsistent expert prior beliefs

14

Future Directions for PSE Work

- Further sensitivity analysis, validation, and diagnostics for existing methods
- Further work on uncertainty estimates for existing methods
- Methods that incorporate multiple estimates
- Opportunities to develop new methods that incorporate
  - New technology
  - Social media data

15

References

- Fearon, E., Chabata, S. T., Thompson, J. A., Cowan, F. M., and Hargreaves, J. R. (2017). Sample Size Calculations for Population Size Estimation Studies Using Multiplier Methods With Respondent-Driven Sampling Surveys. *JMIR Public Health and Surveillance*, 3(3):e59.
- Feehan, D. M. and Salganik, M. J. (2016). Generalizing the Network Scale-Up Method: A New Estimator for the Size of Hidden Populations. *Sociological Methodology*, 46:153–186.
- Handcock, M. S., Gile, K. J., and Mar, C. M. (2014). Estimating Hidden Population Size Using Respondent-Driven Sampling Data. *Electronic Journal of Statistics*, 8(1):1491–1521.
- Handcock, M. S., Gile, K. J., and Mar, C. M. (2015). Estimating the Size of Populations at High Risk for HIV Using Respondent-Driven Sampling Data. *Biometrics*, 71(1):258–266.
- Johnston, L. G. (2013). Introduction to Respondent-Driven Sampling. World Health Organization, Geneva, Switzerland. Available at http://applications.emro.who.int/dsaf/ EMRPUB_2013_EN_1539.pdf.
- Johnston, L. G., McLaughlin, K. R., El Rhilani, H., Latifi, A., Toufik, A., Bennani, A., Alami, K., Elomari, B., and Handcock, M. S. (2015). Estimating the Size of Hidden Populations Using Respondent-driven Sampling Data: Case Examples from Morocco. *Epidemiology*, 26:846–852.
- Johnston, L. G., Prybylski, D., Raymond, H. F., Mirzazadeh, A., and McFarland, W. (2013). Incorporating the service multiplier method in respondent-driven sampling surveys to estimate the size of hidden and hard-to-reach populations: case studies from around the world. *Sexually Transmitted Diseases*, 40(4):304–310.
- Maltiel, R. A., Raftery, A. E., McCormick, T. H., and Baraff, A. J. (2015). Estimating Population Size Using the Network Scale Up Methods. *The Annals of Applied Statistics*, 9(3):1247–1277.
- UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance (2010). Guidelines on Estimating the Size of Populations Most At Risk to HIV. World Health Organization, Geneva, Switzerland. Available at http://apps.who.int/iris/bitstream/10665/ 44347/1/9789241599580_eng.pdf.

16

Estimating the Size of Hidden Populations

Katherine R. McLaughlin

References

Fearon, E., Chabata, S. T., Thompson, J. A., Cowan, F. M., and Hargreaves, J. R. (2017). Sample Size Calculations for Population Size Estimation Studies Using Multiplier Methods With Respondent-Driven Sampling Surveys. *JMIR Public Health and Surveillance*, 3(3):e59.

Feehan, D. M. and Salganik, M. J. (2016). Generalizing the Network Scale-Up Method: A New Estimator for the Size of Hidden Populations. *Sociological Methodology*, 46:153–186.

Handcock, M. S., Gile, K. J., and Mar, C. M. (2014). Estimating Hidden Population Size Using Respondent-Driven Sampling Data. *Electronic Journal of Statistics*, 8(1):1491–1521.

Handcock, M. S., Gile, K. J., and Mar, C. M. (2015). Estimating the Size of Populations at High Risk for HIV Using Respondent-Driven Sampling Data. *Biometrics*, 71(1):258–266.

Johnston, L. G. (2013). *Introduction to Respondent-Driven Sampling*. World Health Organization, Geneva, Switzerland. Available at `http://applications.emro.who.int/dsaf/EMRPUB_2013_EN_1539.pdf`.

Johnston, L. G., McLaughlin, K. R., El Rhilani, H., Latifi, A., Toufik, A., Bennani, A., Alami, K., Elomari, B., and Handcock, M. S. (2015). Estimating the Size of Hidden Populations Using Respondent-driven Sampling Data: Case Examples from Morocco. *Epidemiology*, 26:846–852.

Johnston, L. G., Prybylski, D., Raymond, H. F., Mirzazadeh, A., and McFarland, W. (2013). Incorporating the service multiplier method in respondent-driven sampling surveys to estimate the size of hidden and hard-to-reach populations: case studies from around the world. *Sexually Transmitted Diseases*, 40(4):304–310.

Maltiel, R. A., Raftery, A. E., McCormick, T. H., and Baraff, A. J. (2015). Estimating Population Size Using the Network Scale Up Methods. *The Annals of Applied Statistics*, 9(3):1247–1277.

UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance (2010). *Guidelines on Estimating the Size of Populations Most At Risk to HIV*. World Health Organization, Geneva, Switzerland. Available at `http://apps.who.int/iris/bitstream/10665/44347/1/9789241599580_eng.pdf`.