National Institutes of Health Pathways to Prevention Workshop: Methods for Evaluating Natural Experiments in Obesity

Karen M. Emmons, Ph.D.[1], Chyke A. Doubeni, M.D., M.P.H.[2], Maria E. Fernandez, Ph.D.[3], Diana L. Miglioretti, Ph.D.[4], Jonathan M. Samet, M.D., M.S.[5]

[1]Harvard T.H. Chan School of Public Health, Boston, MA; [2]Department of Family Medicine and Community Health, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; [3]Center for Health Promotion and Prevention Research, Department of Health Promotion and Behavioral Sciences, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX; [4]Division of Biostatistics, Department of Public Health Sciences, School of Medicine, University of California, Davis, Davis, CA; [5]Office of the Dean, Colorado School of Public Health, Aurora, CO.

On December 5–6, 2017, the National Institutes of Health (NIH) convened the Pathways to Prevention Workshop: Methods for Evaluating Natural Experiments in Obesity to identify the status of methods used to assess natural experiments to reduce obesity, areas in which these methods could be improved, and research needs for advancing the field. This paper considers the findings from a systematic evidence review related to methods used for evaluating natural experiments in obesity, presentations by experts and stakeholders, and public comment.

Research gaps are identified and recommendations provided related to four key issues. Recommendations related to *population-based data sources and data integration* include maximizing use and sharing of existing surveillance and research databases, and ensuring there is significant effort to integrate and link existing databases. In the area of *measurement,* recommendations include use of standardized and validated measures of obesity-related outcomes and exposures, systematic measurement of co-benefits and unintended consequences, and expanded use of validated technologies for measurement. *Study design recommendations* include improving guidance, documentation, and communication about methods used, increased use of designs that minimize bias in natural experiments, and more careful selection of control groups. *Cross-cutting recommendations* target activities that the NIH and other funders might undertake to improve the rigor of natural experiments in obesity, including training and collaboration related to modeling and causal inference, promoting the importance of community engagement in conduct of natural experiments, ensuring maintenance of relevant surveillance systems, and supporting extended follow-up assessments for exemplar natural experiments.

To combat the significant public health threat posed by obesity, researchers should continue to take advantage of natural experiments. The recommendations in this report aim to strengthen evidence from such studies.

Introduction

The rapid increase in overweight and obesity poses a significant and worsening public health threat. About 70% of U.S. adults are overweight or obese; racial/ethnic minorities and those from low socioeconomic backgrounds are disproportionately affected. The complexity of obesity prevention and control requires addressing its multilevel and interrelated drivers. Evidence on effectiveness is needed to guide interventions to reduce overweight and obesity.

Structured, public health-oriented frameworks have considered all influences on energy balance, including individual factors, behavioral settings, community- and systems-level sectors of influence, and social norms and values. A 2012 Institute of Medicine report[1] focused on accelerating progress in obesity control recommended a systems approach to target critical environments for change. Policy actions or programs designed to change "real-world" environments to influence obesity-related outcomes can be evaluated as "natural experiments." As defined by the Medical Research Council, natural experiments examine natural or unplanned variation in exposures using design and analytical features that can support causal inference.[2] Natural experiments include laws or policies (e.g., taxes on sugar-sweetened beverages), programs implemented at a community level (e.g., park-based physical activity programs), and environmental change (e.g., construction of separated bike lanes).

On December 5–6, 2017, the National Institutes of Health (NIH) convened the Pathways to Prevention Workshop: Methods for Evaluating Natural Experiments in Obesity to identify the status of methods used to assess natural experiments to reduce obesity, areas in which these methods could be improved, and research needs for advancing the field. The workshop was co-sponsored by the NIH Office of Disease Prevention, the National Cancer Institute, the National Institute of Diabetes and Digestive and Kidney Diseases, and the National Heart, Lung, and Blood Institute. A multidisciplinary working group developed the agenda, and the Johns Hopkins Evidence-based Practice Center prepared a systematic evidence review (EPC Report)[3] that

addressed methods used for evaluating natural experiments in obesity. During the workshop, there were presentations by experts and stakeholders; panel members asked questions of the experts; and onsite and online participants commented during open discussions. After considering the EPC Report findings and the workshop proceedings, an independent panel prepared a draft report for public comment and then revised it based on the comments received.

In formulating this report, the panel specifically acknowledged the temporally and spatially dynamic nature of the obesity epidemic, which varies in its severity across populations and contexts. For example, the epidemic of childhood obesity varies widely across groups defined by race, ethnicity, and income, and its rise over time antedates the sedentary behavior associated with video game use and screen time. This heterogeneity is an overarching challenge in translating research findings into action. The goal of research on obesity is to produce actionable evidence leading to interventions that will slow and ultimately end an epidemic that is forecasted to continue deep into the 21st century. To generate the requisite evidence, an integrated, networked, and contextually responsive research enterprise is needed. Creating and sustaining that broad enterprise should be a goal for the NIH, other funders, and stakeholders.

This panel report summarizes the workshop, identifies research gaps, and provides recommendations for further enhancing the methodological rigor of natural experiments. The recommendations, summarized in Tables 1–4, are organized around four issues: population-based data sources and data integration; obesity-related measures; study design issues; and cross-cutting issues that transcend the key questions of the EPC review and are essential for achieving the goal of generating useful evidence for decision-making in an ongoing fashion.

Population-Based Data Sources and Data Integration

Table 1 summarizes recommendations to improve data sources and data integration. The EPC Report found a lack of shared comprehensive surveillance and research databases. Most currently available databases contain information on schools and communities. Few databases had a formal or detailed data dictionary that would permit efficient shared use. Presenters noted that many existing data systems are inadequate to evaluate the complex socioecological and biological factors driving the obesity epidemic at the level of granularity needed. Primary and secondary data sources differed in their instruments and approaches used for data collection and in their follow-up time frames. Data linkage and integration of multiple data sources could provide measurements across multiple levels. However, such integrated data systems are unavailable for widespread use.

It was also noted that data development efforts are disparate, siloed, and often duplicative because of limited cross-sector collaboration. Workshop presenters and participants conveyed a need for collaborative platforms and partnerships for more efficient evaluations of natural experiments with both retrospective and prospective data. They also emphasized the need to improve communication between researchers and policymakers, to enhance engagement of policymakers, and to prioritize availability of timely data to drive decisions.

Recommendations to Increase Integration of Population-Based Data Sources

To accelerate progress and catalyze innovation, funding agencies should make investments in a data ecosystem that lessens the burden of regulatory, privacy, and proprietary concerns that restrict access to and sharing of individual-level data and/or geospatial data at smaller levels of aggregation than zip codes. Building on existing infrastructure (e.g., National Collaborative on Childhood Obesity Research) would help support integration of data systems on objective obesity-related measures across the lifespan. The panel also recommends

enhancing linkages of electronic health record data with community-level data by leveraging advances in interoperability and health information exchanges.

The panel recommends fostering ongoing public-private partnerships to harness and coalesce the combined power of population-level data in both the private (e.g., Google) and the public (e.g., crime and school data) sectors with the goal of establishing new and robust surveillance platforms that will enable dynamic and longer-term monitoring of natural experiments. The panel urges the recognition of relevant data sources outside the health sector, such as those related to agriculture, transportation, and city planning, and the importance of multisector collaboration. The integration of high-dimensional data from mobile devices equipped with positioning systems offers promise for capturing dynamic data, which may improve the efficiency and reduce the cost of data collection.

Measurement of Obesity-Related Outcomes

Table 2 summarizes recommendations to improve the measurement of obesity-related outcomes. Studies typically measure body composition, dietary intake, or physical activity. Gold standard measures of these outcomes (e.g., underwater weighing, doubly labeled water) are rarely used in natural experiments, given high costs and limited feasibility in community settings. Studies reviewed in the EPC Report typically measured body composition using body mass index z-score or percentile in children, adjusting for child's age and sex, and body mass index for adults. Children's height and weight were typically measured by trained staff; studies of adults largely used self-report. Other measures of body composition—such as waist-to-hip ratio, skinfold thickness—were not used.

The EPC Report found that dietary intake was typically measured as change in total daily caloric intake or intake of fruits and vegetables, sugar-sweetened beverages, fast food, or fiber. Studies used 24-hour recall, food frequency questionnaires, short dietary questionnaires,

record logs, or food diaries. Questionnaires varied widely across studies, limiting comparability and complicating pooling of data.

Presenters highlighted the need to validate questionnaires in diverse populations and to develop web-based registries of standardized, valid, and reliable obesity-related measures to encourage use of common measures. The National Collaborative on Childhood Obesity Research Measures Registry (http://www.nccor.org, accessed March 5, 2018) lists assessment instruments for individual diet, food environment, individual physical activity, and physical activity environment in children. There are no similar resources for measures designed for use with adults.

A workshop presenter discussed the need for more accurate measures of food intake, such as measuring "plate waste." Promising new technologies currently exist and others are being developed to improve the efficiency and accuracy of collecting dietary measures, such as assessment of dietary intake from digital images of meals, digital images accompanied by a voice recording describing the food, and smartphone diary applications.

Studies in the EPC Report generally assessed single aspects of physical activity, such as counts of steps, rather than multidimensional aspects of total physical activity. Assessment strategies used included questionnaires, electronic monitoring, a record or log, and direct observation. Electronic monitoring and mobile devices (e.g., video cameras and tracking devices) are increasingly used to improve efficiency and accuracy of data collection.

The EPC Report found that some studies evaluated co-outcomes, including food purchasing behavior, commuting behavior, food environment, and physical activity environment.

Recommendations for Improving Measurement of Obesity-Related Outcomes (Table 2)

The panel recommends that existing registries of relevant measures for children should be expanded to document the validity and reliability of the instruments, to offer guidance on the best measures for specific types of studies, and to include measures of exposures commonly collected in obesity-related studies. This resource should include a detailed data dictionary for both outcomes and exposures. A similar resource is needed for measures of obesity-related variables in adults.

There are promising new technology-based measures, although they need careful validation and comparison to existing established measures in order to understand the additional information or efficiencies gained. Assessment of economic, safety, environmental, health, and social measures is needed to measure both co-benefits and unintended consequences of interventions.

Study Design Issues

Table 3 summarizes recommendations to improve design of natural experiments, including strategies to reduce bias. The EPC Report used the Effective Public Health Practice Project Risk of Bias Assessment Tool to classify study design. Of the 294 studies reviewed, about one-half were natural experimental studies. The most common designs used included cross-sectional comparisons of exposed and unexposed groups, difference-in-differences approaches, and pre/post designs. Other non-experimental designs, such as instrumental variable approaches, regression discontinuity approaches, and interrupted time series designs, were used infrequently. Most studies involved individual-level analyses, and most used methods such as multilevel modeling or robust standard errors to account for the hierarchical structure of the data. However, some studies did not account for clustering in the analytic approach, which

may result in an inflated type 1 error rate.[4,5] Also, when there is only one site per study arm, it is not possible to separate variability due to treatment from variability due to site.[6]

A key design issue relates to strategies to reduce bias. Observational studies are the pillar for assessing the effects of "real-world" interventions, but face multiple threats to validity, including from confounding, measurement error, and selection bias. Efforts to minimize confounding are critical in the design and analysis of natural experiments. To infer that confounding affects the results of a particular study, there should be sufficient understanding of the suite of relevant potential confounding factors and indications that the confounding factors are associated with the intervention under investigation. Threats to validity in observational studies can also arise from non-comparability of the treatment and comparison groups on one or more covariates.

In estimating intervention effects, the counterfactual model of causation is useful. This approach compares the risk of the outcome under the condition of exposure to the intervention with that risk on the assumption of no exposure—the counterfactual—with all other factors contributing to risk being comparable. Such comparability is sought through study design and by data analysis (e.g., adjustment and use of propensity scores). Failure to achieve comparability of groups receiving an intervention with comparison groups is a well-known source of bias in observational studies.

Workshop presenters emphasized ways to improve comparability between treatment and comparison conditions. Broad pragmatic principles proposed include choice of the nonequivalent comparison population, use of a pretest measure of outcome, and use of a "rich" set of covariates to control for unmeasured sources of bias. Presenters provided emerging evidence that combining the right design and analytical methods can produce results comparable to randomized experiments. Workshop presentations also highlighted the advantages of some approaches for establishing causal inference and for reducing potential

bias. These quasi-experimental approaches can help estimate the average treatment effect in studies in which randomization is undesirable or not feasible. One recommended method is the regression-discontinuity approach, a quasi-experimental design that assigns a threshold below or above which an intervention is assigned and examines causal effects of interventions by observing either side of the threshold. Speakers provided evidence from social and behavioral studies that have directly compared randomized controlled trial and regression discontinuity estimates at cutoff, holding constant treatment, measurement specifics, and the estimand (that which is to be estimated in a statistical analysis), demonstrating comparability in estimates and outcomes.

Comparative interrupted time series designs evaluate program impacts by observing whether deviations from baseline trends in the treatment group are greater than in the comparison group. Pre-intervention data at multiple time points are critical for describing the nature of the selection difference between the treatment and comparison groups, both in general and as it changes over time.

Other presenters described novel strategies for estimating effects of natural experiments, including use of synthetic control methods (using a weighted combination of multiple sample populations as the comparison group) and instrumental variables (selection of a third variable that induces changes in the explanatory variable but has no independent effect on the dependent variable to overcome the problem of correlation among variables in a multilevel model and allow the researcher to uncover the causal effect). Near/far matching combines standard matching with instrumental variables analysis by simultaneously matching individuals who are similar on observable characteristics (e.g., age, race, sex) and maximally different on the instrumental variable. This approach can help correct for both measured and unmeasured confounders that lead to differences in outcomes between exposed and unexposed participants. Bounding approaches facilitate estimating effects of obesity on any measurable outcome and

associated statistics using observational data non-parametrically. Presenters also recommended choosing analytic strategies depending on the type of data available, such as near/far matching for individual-level data when the control group is imperfect; synthetic control analysis for population-level data that has an imperfect control group; and distributional decomposition approaches for individual or population-level data in order to estimate subgroup disparities.

The EPC Report principally focused on risk of bias from confounding, as assessed by the Effective Public Health Practice Project Risk of Bias Assessment Tool. This assessment revealed a global rating of "weak" for nearly 80% of the natural experiments; however, this classification reflects in part the properties of the tool, which rates only randomized controlled trials and controlled clinical trials as "strong" in design. Evaluation of the handling of confounding was based on which variables the investigators considered as confounders, without distinguishing actual from potential confounders or considering the underlying causal processes. Withdrawals and losses to follow-up were a further consideration leading to a "weak" classification for the majority of studies. Non-blinding was also a limitation of observational studies, although not achievable for participants in natural experiments. Tools are clearly needed to estimate bias in natural experiments.

Table 3 provides specific recommendations regarding study design and methodology. As a foundation, the panel recommends use of explicit conceptual frameworks depicting assumed causal relationships to guide decisions about design and analysis.

Cross-Cutting Issues That Could Strengthen Natural Experiments (Table 4)

Methodologic and Analytic Strategies to Address Complexity. The use of natural experiments for obesity prevention is challenging due to the complex nature of energy balance and the many factors affecting it, leading to analytic challenges in reflecting this complexity.

Complex system models focus on uncovering pathways and mechanisms that underlie successful interventions and address multiple interacting intervention elements. These models can help identify the data that are needed to understand complex systems and facilitate extrapolation from one context to diverse settings and long time horizons. Simulations also can help anticipate how environmental changes will affect behavior and support policymakers as they consider different policy options. A critical factor in ensuring that modeling approaches fulfill their potential will be the development of mechanisms for dialogue between modelers and the obesity research community, which will help identify key uncertainties in models that are critical to resolve if models are to be internally and externally valid.

In addition to estimates of the effectiveness of interventions, decision-makers need estimates of population impact, cost, cost-effectiveness, and impact on equity. A well-tested but largely underutilized strategy for evaluating outcomes of natural experiments is cost-effectiveness analysis. There are some examples, including the Childhood Obesity Intervention Cost-Effectiveness Study model presented at the workshop—a microsimulation modeling effort to project the population impact of interventions and their cost-effectiveness over 10 years in the United States. Such modeling is a useful strategy for projecting future population impact, cost, cost-effectiveness, and equity.

Capacity for Improving Methods Used in Natural Experiments. Over the past decade, the importance of quasi-experimental designs, including natural experiments, has been increasingly recognized for enhancing the impact of research in the real world. As a result, NIH-sponsored training programs have provided opportunities to develop skills in a wide range of design approaches. Efforts to develop skills in analysis of large and multilevel data from multiple sources have been limited. More training efforts in complex system models and causal inference methods are needed.

Public-private partnerships could also expand data sources and the capacity for evaluation of natural experiments. Privacy, regulatory, and proprietary concerns hinder information and data sharing as well as access to novel and existing data systems. For example, many natural experiments use commercial data systems that must be purchased separately for each study, leading to duplicative efforts in data cleaning and limiting sharing of data enhancements.

Community-Level Issues That May Affect the Validity of Natural Experiments. Community engagement is important for both the planning, conduct, and interpretation of findings from natural experiments. Community members likely have a deep understanding of key contextual factors that will be particularly important for natural experiments at the population or community level. Community engagement can also help ensure adequate representation of subpopulations, such as racial/ethnic minority groups and rural populations. Inclusion of these groups has been limited in evaluating natural experiments, and more diverse representation is critical. Cultural and social connections often transcend geographic boundaries, and understanding the ways that communities define themselves is important context for population-level experiments. Another key issue for improving the rigor and generalizability of natural experiments relates to the definitions of population-level units, such as neighborhoods.

The Critical Role of Surveillance Systems and Long-Term Follow-Up. Population-level surveillance systems—such as the U.S. Census, the National Health and Nutrition Examination Survey (NHANES), and the Behavioral Risk Factor Surveillance System (BRFSS)—represent cost-efficient resources for evaluation of natural experiments in obesity prevention. It is critical that these resources continue to be publicly available and supported in ways that maintain their rigor and maximize scientific utility. The panel also notes the importance of the electronic health record as a key resource. Considerable effort is needed to ensure curation of record data for use in research studies, but the approaches are often not standardized across individual health

systems to allow broader use and sharing and ensure the comparability of measures across health systems.

Research studies, including natural experiments that assemble cohorts and population-based samples, often have a truncated period of follow-up, largely due to the vagaries of funding. However, long-term follow-up is needed to determine if intervention effects on obesity are sustained. In many cases, it would be cost-effective to continue surveillance and follow-up in these valuable cohorts, which would allow the assessment of long-term impact and cost-effectiveness. It is important to develop criteria for such continued support, to ensure that resources are put into the best opportunities for added value. The question as to which entities would fund such ongoing follow-up needs consideration.

Conclusions

The panel's recommendations (Tables 1–4) are directed at various partners of the research enterprise: funders, researchers, public health and clinical practitioners, decision-makers, and communities. There are many ways to implement the recommendations, including convening workshops, conducting evidence reviews, providing funding for targeted research and demonstration projects, and developing and supporting data repositories. The panel did not link each recommendation to an implementation path, but does offer general suggestions:

- **Data catalogs and repositories.** Recommendations relate to identification of key data resources and their use for research and evaluation. The panel urges funders and stakeholders to support the development and maintenance of the requisite data platforms.
- **Data access and sharing.** The panel identified issues around data access that need attention from funders and federal and private agencies—such as obtaining data with

sufficient geographic granularity for meaningful linkages and assuring that key data systems are maintained and even augmented. Funders also might act to enhance access to commercial data sources and facilitate engagement to address longstanding privacy, regulatory, and proprietary concerns that hamper appropriate data sharing.

- **Workshops.** The panel's recommendations call for standardization and harmonization related to both outcomes and exposure variables. Such topics often are addressed effectively through workshops with comprehensive reports and ongoing working groups that obtain broad input. Workshop formats also could be used to advance methods and to compare results using alternative design and analytical approaches by drawing on diverse disciplines.

- **Consortia.** For many questions, the most informative evidence will come from research that reaches across diverse populations and geographies. This is often best accomplished by establishing consortia to assemble cohorts with broadly representative populations to provide generalizable evidence and serve as a resource or platform for individual researchers.

- **Targeted funding opportunities.** These can encourage research on specific topics or to develop methodology. Funding might be specifically targeted at some of the recommendations around study methodology and novel data collection approaches, an area that is generally poorly supported through the usual funding processes.

- **Evidence reviews and synthesis reports.** Periodic authoritative reviews (e.g., Surgeon General's Reports) have proved invaluable, both for taking stock of the state of evidence on a variety of topics and for advancing research and cross-sector dialogue.

The obesity epidemic is likely to worsen for many decades. To combat this significant public health threat, researchers should continue to take advantage of natural experiments. The recommendations in this report aim to strengthen evidence from such studies.

References

1.   Committee on Accelerating Progress in Obesity Prevention; Food and Nutrition Board; Institute of Medicine; Glickman D, Parker L, Sim LJ, et al, eds. *Accelerating Progress in Obesity Prevention: Solving the Weight of the Nation*. Washington, DC: The National Academies Press; 2012. Accessed at https://www.ncbi.nlm.nih.gov/books/NBK201141/ on 5 March 2018.

2.   Medical Research Council. *Using Natural Experiments to Evaluate Population Health Interventions: Guidance for Producers and Users of Evidence*; 2009. Accessed at http://www.mrc.ac.uk/documents/pdf/natural-experiments-guidance on 5 March 2018.

3.   Bennet W, Cheskin LJ, Wilson RF, et al. *Methods for Evaluating Natural Experiments in Obesity: A Systematic Evidence Review. Comparative Effectiveness Review No. 204.* (Prepared by the Johns Hopkins University Evidence-based Practice Center under Contract No. 290-2012-00007-I). AHRQ Publication No. 18-EHC006-EF. Rockville, MD: Agency for Healthcare Research and Quality; 2017. Accessed at https://effectivehealthcare.ahrq.gov/topics/obesity-research-methods/systematic-review on 5 March 2018.

4.   Turner EL, Li F, Gallis JA, Prague M, Murray DM. Review of recent methodological developments in group-randomized trials: part 1-design. American Journal of Public Health. 2017;107(6):907–915. doi:10.2105/AJPH.2017.303706.

5.   Turner EL, Prague M, Gallis JA, Li F, Murray DM. Review of recent methodological developments in group-randomized trials: part 2-analysis. American Journal of Public Health. 2017;107(7):1078–1086. doi:10.2105/AJPH.2017.303707.

6.   Varnell SP, Murray DM, Baker WL. An evaluation of analysis options for the one-group-per-condition design. Can any of the alternatives overcome the problems inherent in this design? Evaluation Review. 2001;25(4):440–453. doi:10.1177/0193841X0102500402.

Table 1. Recommendations for Improving Data Systems

1. **Maximize use and sharing of existing surveillance and research databases.**

   A. Address long-standing regulatory, privacy, and proprietary hurdles to the sharing of existing surveillance and research-generated databases.

   B. Support web-based platforms that provide a registry of population-level databases, with supporting documentation such as data dictionaries, to facilitate data sharing.

   C. Identify and address gaps in existing data systems.

2. **Enhance the value of existing databases through data integration and linkages.**

   A. Develop and use integrated data sources to enhance use of innovative designs (e.g., integrate health care or device data with community-level health policy data).

   B. Create novel surveillance infrastructures to enable long-term monitoring of obesity-related outcomes and exposures, including detailed spatiotemporal data, before and after natural experiments. This is best enabled through strong public-private or cross-sector partnerships.

Table 2. Recommendations for Improving Measurement of Obesity-Related Outcomes

1. **Promote use of common standardized, valid, and reliable measures of obesity-related outcomes and exposures.**

   A. Convene stakeholder group to standardize terminology and definitions for obesity-related outcomes and exposures.

   B. Convene stakeholder group to identify or develop a core set of measures for obesity-related outcomes and exposures, validated for diverse populations.

   C. Expand or develop web-based registries of obesity-related measures for children and adults.

2. **Encourage measurement of co-benefits and unintended consequences.**

   A. Identify and measure co-benefits and unintended consequences of natural experiments that are important to community members and other stakeholders in relevant sectors to inform decision-making and policymaking.

3. **Further develop and integrate use of new technologies for measuring obesity-related outcomes and exposures.**

   A. Support efforts to improve the efficiency, accuracy, and analysis of detailed dietary and physical activity data collected via new technologies.

   B. Promote collection of data using established measures, along with measures from new technologies, until new measures are validated in diverse populations.

Table 3. Recommendations for Improving Methods for Design and Analysis

1. **Improve guidance, documentation, and communication about methods.**

   A. Develop reporting guidelines for critical study elements including selection or construction of the comparison population and the assessment and control of confounding.

   B. Develop and use standard nomenclature related to natural experiment designs and analytic methods.

   C. Develop reporting standards with guidance to fully describe the approaches taken to reduce the potential for bias in protocols and publications. This should go beyond a listing of potential confounders and address those specific to the topic area and populations under investigation.

   D. Describe analytical strategies for addressing bias in sufficient detail to allow for replication, and explicitly state the statistical models employed.

2. **Increase awareness and use of designs to minimize bias.**

   A. Advance awareness and use of study designs that reduce bias and support causal inference.

   B. Develop a tool specifically designed to assess risk of bias in natural experiments.

   C. Convene a workshop to evaluate the suitability of emerging analytic approaches for addressing potential bias from non-comparability of treatment and comparison groups in natural experiments.

   D. Researchers should consider combining multiple methods for controlling for potential

sources of bias, including (1) the approach to selection of the comparison population or of multiple comparison populations, (2) the use of a pretest measure of outcome in intervention and comparison populations, and (3) the use of a robust set of covariates.

E. Researchers should acknowledge and, where possible, adjust for the hierarchical nature of data in most natural experiments. For example, researchers should identify additional intervention and comparison sites (or only additional comparison sites if additional intervention sites are not available). If only two sites are available, researchers should refrain from making any causal inference, because it is impossible to separate variability due to the intervention from variability due to the sites.

3. **Use informed approaches for the selection of comparison groups.**

A. Use community-based methods to guide design decisions such as selection of appropriate comparison groups or communities.

B. Consider and further evaluate emerging approaches for control group selection, generation, and analysis. These may include bounding, instrumental variable approaches including the near/far approach, and use of synthetic comparison groups.

Table 4. Recommendations for Cross-Cutting Issues

1. Develop NIH-sponsored training programs in modeling studies and causal inference relevant to obesity control and prevention, and support collaborations that build this capacity within the nation's prevention research community across disciplines as well as across multiple sectors.

2. Increase awareness and capacity for incorporating community engagement and context into design and implementation of natural experiments.

3. Undertake efforts to ensure that crucial surveillance systems are maintained.

4. Develop criteria and a funding mechanism for extending the follow-up period of exemplar natural experiments, and identify organizations or agencies who are well-paced to lead efforts to achieve the recommendations outlined in this report.

National Institutes of Health (NIH) Pathways to Prevention Workshop:
Methods for Evaluating Natural Experiments in Obesity

Panel Roster

Workshop & Panel Chair: Karen M. Emmons, Ph.D.
Professor of Social and Behavioral Science
Harvard T.H. Chan School of Public Health

Chyke A. Doubeni, M.D., M.P.H., FRCS
Chair and Presidential Professorship of Family
Medicine and Community Health
Department of Family Medicine and
Community Health
Perelman School of Medicine
University of Pennsylvania

Maria Fernandez, Ph.D.
Director
Center for Health Promotion and Prevention
Research
Professor
Department of Health Promotion and
Behavioral Sciences
School of Public Health
University of Texas Health Science Center at
Houston

Diana L. Miglioretti, Ph.D.
Dean's Professor in Biostatistics
Division of Biostatistics
Department of Public Health Sciences
School of Medicine
University of California, Davis

Jonathan M. Samet, M.D.
Dean
Professor
Department of Epidemiology
Office of the Dean
Colorado School of Public Health