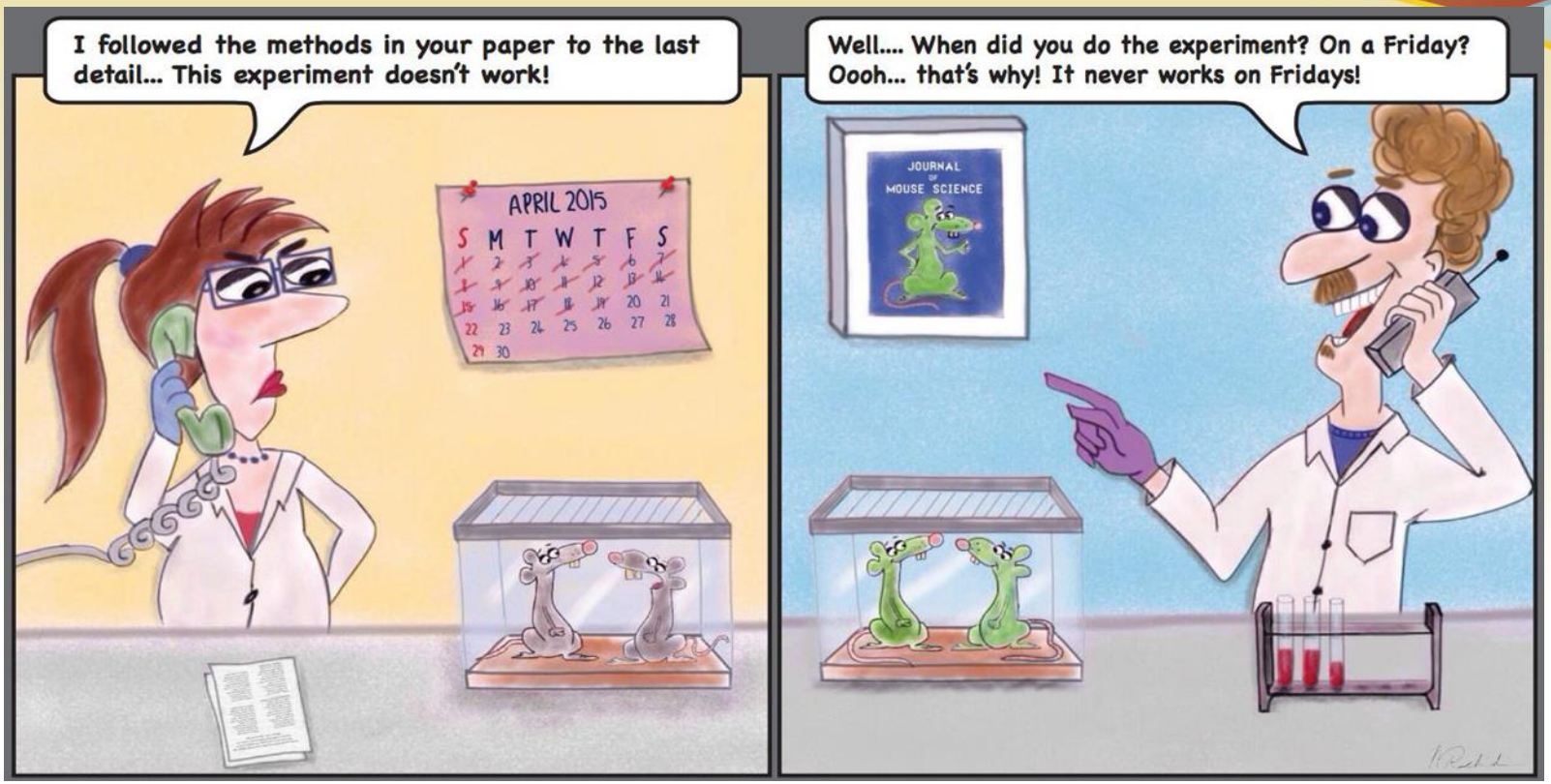


Borrowed from Columbia

Rigor, Transparency and Reproducibility

Carrie Dykes, PhD

Research Engagement Specialist

Why is NIH Making More Work for Me?

- **NIH Mission**
  - To seek fundamental knowledge about the nature and behavior of living systems and the application of that knowledge to enhance health, lengthen life and reduce illness and disability.
  - Key to this is scientific rigor: and one of NIH's goals is to 'exemplify and promote the highest level of scientific integrity, public accountability and social responsibility in the conduct of science'.

Key items

- **Rigor**
  - scientific premise
- **Reproducibility**
  - quality system in your lab
- **Transparency**
- **Robust and unbiased results**
  - sex factored into design
- **Authentication of Key Biological and/or Chemical Resources**
  - New attachment

Borrowed from San Diego

Rigorous Experimental Design

- **Scientific rigor is the strict application of the scientific method to ensure robust and unbiased experimental design, methodology, analysis, interpretation and reporting of results.**
- **NIH expects applicants to describe how they will achieve robust and unbiased results when describing the experimental design and proposed methods. Features of experimental design may include:**
  - Use of standards
  - Sample size estimation
  - Randomization
  - Blinding
  - Appropriate replicates
  - Controlling for inter-operator variability
  - Statistical methods planned
  - Inclusion and exclusion criteria
  - Subject retention and attrition
  - How missing data will be handled
  - And others, as appropriate to the science

HOW SCIENTISTS FOOL THEMSELVES – AND HOW THEY CAN STOP

Humans are remarkably good at self-deception. But growing concern about reproducibility is driving many investigators to seek ways to fight their own worst instincts.

COGNITIVE FALLACIES IN RESEARCH

HYPOTHESIS MYOPIA

Collecting evidence to support a hypothesis, not looking for evidence against it, and ignoring other explanations.

TEXAS SHARPSHOOTER

Seizing on random patterns in the data and mistaking them for interesting findings.

ASYMMETRIC ATTENTION

Rigorously checking unexpected results, but giving expected ones a free pass.

JUST-SO STORYTELLING

Finding stories after the fact to rationalize whatever the results turn out to be.

Adapted from Nature (go.nature.com/nqyohl)

Borrowed from NYU

HOW SCIENTISTS FOOL THEMSELVES – AND HOW THEY CAN STOP

Humans are remarkably good at self-deception. But growing concern about reproducibility is driving many investigators to seek ways to fight their own worst instincts.

DEBIASING TECHNIQUES

DEVIL'S ADVOCACY

Explicitly consider alternative hypotheses – then test them out head-to-head.

PRE-COMMITMENT

Publicly declare a data collection and analysis plan before starting the study.

TEAM OF RIVALS

Invite your academic adversaries to collaborate with you on a study.

BLIND DATA ANALYSIS

Analyze data that look real but are not exactly what you collected – and then lift the blind.

Adapted from Nature (go.nature.com/nqyohl)

Borrowed from NYU

Transparency

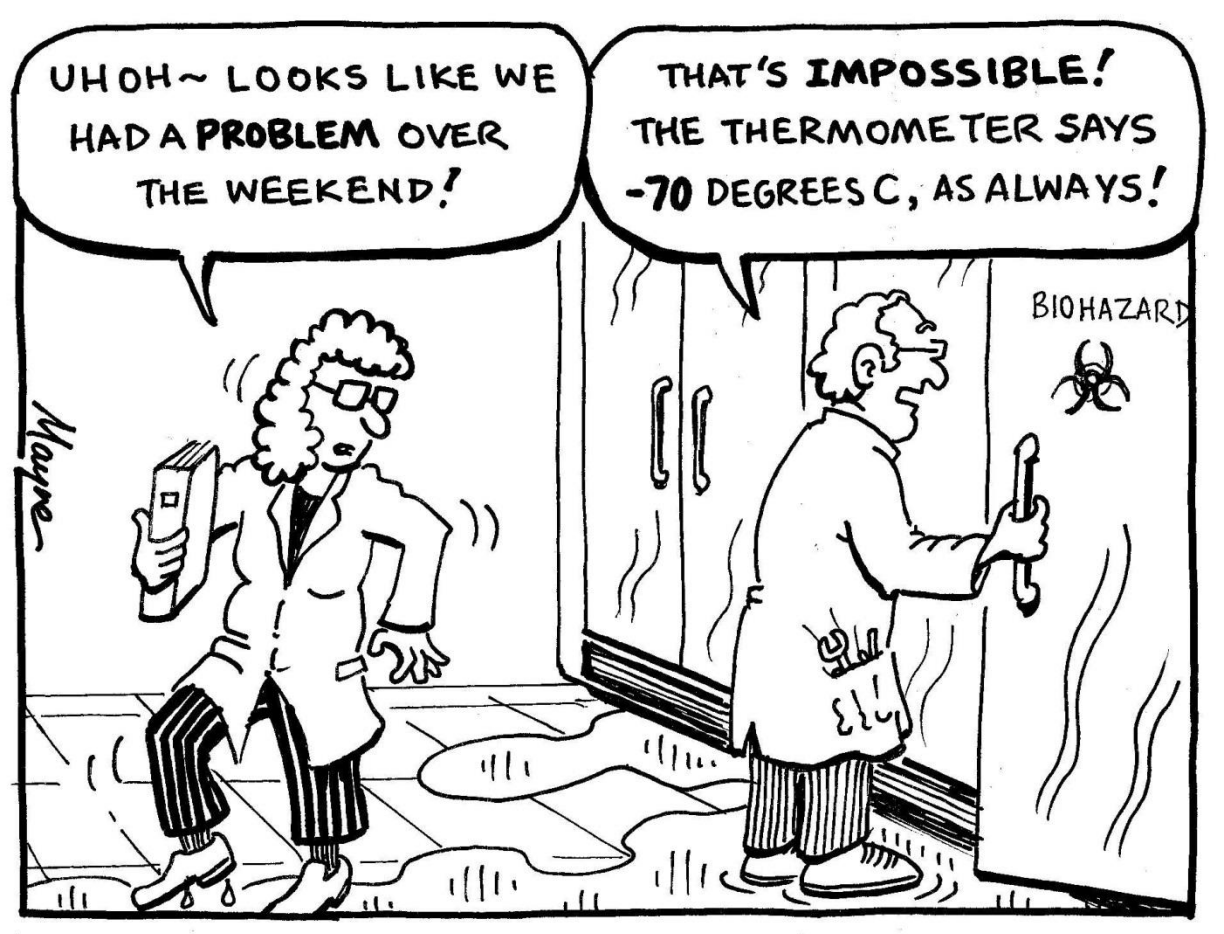
- **Full Transparency in reporting experimental details so that others can reproduce the results**
- **Sharing of raw data sets and programs**
- **Data Archiving**
  - process of moving data that is no longer actively used to a separate storage device for long-term retention. Archive data consists of older data that is still important to the organization and may be needed for future reference, as well as data that must be retained for regulatory compliance.
- **Device Archiving**
  - Same idea as data

Reproducibility-Quality Lab System

- **Equipment**
  - Has equipment been maintained properly and calibrated?
- **Management**
  - Who reports to who and who helps who with problems?
- **SOPs**
  - Are protocols documented are SOPs followed? Are deviations documented and corrected? Are the materials properly made, labeled, stored, and not expired.
- **Training**
  - Are people properly trained and is training documented?
- **Documentation**
  - How are lab notebooks kept and results documented and reviewed? Are lab note books signed off on?

Your Grant

- **Significance**
- **Approach**
- **Authentication of Key Resources Plan**



Significance section

- **Explicitly state the *scientific premise* for the proposed project.**
  - The general strengths and weaknesses of the prior research cited by the applicant, which form the basis for the proposed research
  - Not your hypothesis. This separates the premise from the hypothesis your grant is trying to address; the premise leads to the hypothesis.
- **We recommend starting the Significance section of your grant application with a paragraph or subsection entitled "Scientific Premise".**
  - Consider a separate section in the Significance entitled "Strengths and Weakness of Supporting Data" or, alternatively, a 1-3 sentence appraisal of the data at the end of each section where it is presented.

Strengths and Weaknesses of Supporting Data: Studies of inter-individual differences in leukocyte telomere length (LTL) have focused largely on middle age and elderly persons. These studies have established that adult LTL is influenced by heredity (17-22), by paternal age at conception (PAC) (1, 3-5, 23), and by environmental exposures (24-28) which augment oxidative stress. They have also provided compelling evidence that shortened LTL is related to cardiovascular disease (CVD), principally atherosclerosis (29-36), and reduced longevity (37-40). Yet empirical observations (41-46) and simulations (47) suggest that LTL at birth is a major determinant of LTL throughout the human lifespan, such that individuals endowed with short (or long) LTL at birth are likely to have short (or long) LTL later in life. Therefore, we posit that determinants of LTL at birth impact the evolution of health and disease throughout the life course. By identifying these determinants, we will provide a foundation for linking experience from conception to birth with health and longevity in later life (48). Accordingly, the present study has the potential to transform our understanding of population health by opening novel investigations of the pathways through which intra-uterine experiences are biologically embedded in the individual‘s constitution, and might be reflected in risk factors for disease which emerge in childhood and evolve thereafter.

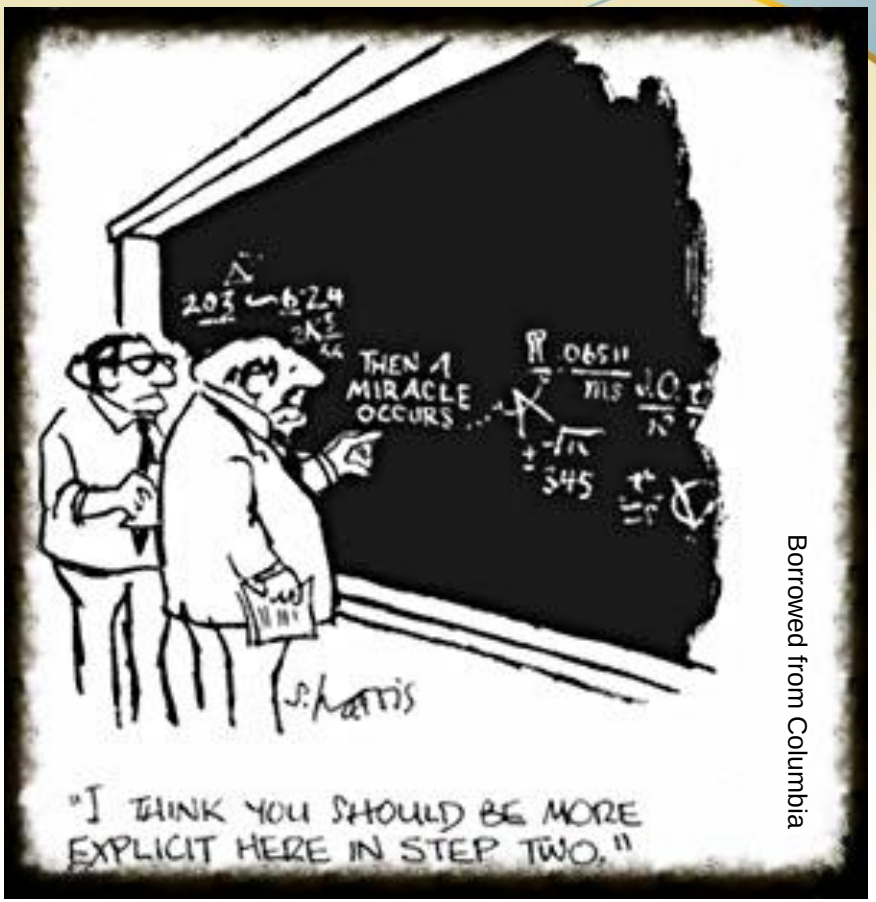
Borrowed from Columbia

Approach section

- **Describe the experimental design and methods proposed and how they will achieve *robust and unbiased* results.**
  - Variables such as age, sex, weight, height and underlying health conditions often are associated with health and disease.
  - NIH *expects* that these will be factored into research designs, analyses, and reporting in vertebrate animal and human studies.
    - ESPECIALLY SEX: Need justification for studies that only include one sex.
  - Explain how relevant biological variables are factored into research designs and analyses for studies in vertebrate animals and humans. (e.g. strong justification from the literature or preliminary data must be provided for studying only female mice.)

Approach Section

- **Again have specific sections in your grant titled**
  - Scientific Rigor
  - Consideration of Sex and Other Biological Variables



Borrowed from Columbia

Concern over biomarker reliability. We have revised our study design to restrict to the biomarkers with greatest reliability (interclass correlation coefficients [ICCS] from 0.49-0.55). It should be noted that an ICC of 0.40 has been proposed as a cutoff for sufficient reproducibility in a biomarker to justify its use in an epidemiologic analysis (Rosner). This cutoff has been cited in previous reports on biomarker reliability and epidemiologic studies using biomarkers with ICCs within this range are routinely conducted. However, we acknowledge that measurement error in our dosimeter could attenuate study power. We have therefore adjusted our power calculations to include correction for measurement error. As seen in Section XX, with the increase in sample size we have excellent power to test study hypotheses after this correction.

Authentication of Key Biological and/or Chemical Resources

- **Briefly describe methods to ensure the identity and validity of key biological and chemical reagents used in the proposed studies.**
  - What is a key biological resource?
    - Resources that might differ from lab to lab over time
    - Have qualities or qualifications that could influence the results
    - Are integral to the proposed research
    - These include, but are not limited to cell lines, specialty chemicals, antibodies and other biologics
  - Briefly describe the methods you will use to authenticate your key resources.
  - Information in this section must focus only on authentication/validation of key resources used in the study; all other methods and preliminary data must be included within the page limits of the research strategy. *Applications identified as non-compliant with this limitation will be withdrawn from the review process*

Authentication of Key Biological and/or Chemical Resources

- **Researchers should transparently report on what they have done to authenticate key resources, so that NIH can develop understanding of consensus approaches.**
- **You can use one description for multiple different resources in the same category (example: authenticating cell lines)**
- **Actual data demonstrating that authenticated resources exist is not necessary**
- **If a key resource is being made as part of the project or is under development, that should be in your research strategy, not this document.**
- **Save this information in a single PDF file named "Authentication of Key Resources Plan," and attach it on the R&R Other Project Information page of the application package**

PHS 398 Research Plan

OMB Number: 0925-0001

Introduction				
1. Introduction to Application (Resubmission and Revision)		Add Attachment	Delete Attachment	View Attachment
Research Plan Section				
2. Specific Aims		Add Attachment	Delete Attachment	View Attachment
3. *Research Strategy		Add Attachment	Delete Attachment	View Attachment
4. Progress Report Publication List		Add Attachment	Delete Attachment	View Attachment
Human Subjects Section				
5. Protection of Human Subjects		Add Attachment	Delete Attachment	View Attachment
6. Data Safefty Monitoring Plan		Add Attachment	Delete Attachment	View Attachment
7. Inclusion of Women and Minorities		Add Attachment	Delete Attachment	View Attachment
8. Inclusion of Children		Add Attachment	Delete Attachment	View Attachment
Other Research Plan Section				
9. Vertebrate Animals		Add Attachment	Delete Attachment	View Attachment
10. Select Agent Research		Add Attachment	Delete Attachment	View Attachment
11. Multiple PD/PI Leadership Plan		Add Attachment	Delete Attachment	View Attachment
12. Consortium/Contractual Arrangements		Add Attachment	Delete Attachment	View Attachment
13. Letters of Support		Add Attachment	Delete Attachment	View Attachment
14. Resource Sharing Plan(s)		Add Attachment	Delete Attachment	View Attachment
15. Authentication of Key Biological and/or Chemical Resources		Add Attachment	Delete Attachment	View Attachment
Appendix				
16. Appendix	Add Attachments	Delete Attachments	View Attachments	☐

Review Criteria

Element of Rigor	Section of Application	Criterion Score	Additional Review Consideration	Contribute to Overall Impact?
Scientific Premise	Research Strategy	Significance	NA	Yes
Scientific Rigor		Approach	NA	Yes
Consideration of Sex and Other Relevant Biological Variables		Approach	NA	Yes
Authentication of Key Biological and/or Chemical Resources	New Attachment	NA	Acceptable or unacceptable	No

Adapted from San Diego

Review criteria

- **Reviewers will be asked to consider additional review questions in order to assess rigor and transparency**
  - Scored Review Criteria
    - Significance
      - Is there a strong scientific premise for the project?
      - The scientific premise will be reviewed as part of the **Significance** criterion, i.e., the importance of the problem, critical barriers to progress, how the proposed project will improve scientific knowledge, and how the field will change if the aims are achieved
    - Approach
      - Have the investigators presented strategies to ensure a robust and unbiased approach, as appropriate for the work proposed?
      - Have the investigators presented adequate plans to address relevant biological variables, such as sex, for studies in vertebrate animals or human subjects?

Additional Review Considerations

- **Authentication of Key Biological and/or Chemical Resources**
  - For projects involving key biological and/or chemical resources, reviewers will comment on the brief plans proposed for identifying and ensuring the validity of those resources.

More information

- **NIH has published notices outlining updates to instructions for applications:**
  - Overview (NOT-OD-16-004),
  - Implementing Rigor and Transparency in NIH & AHRQ Research Grant Applications (NOT-OD-16-011)
  - Implementing Rigor and Transparency in NIH & AHRQ Career Development Award Applications (NOT-OD-16-012).
- **Other Resources**
  - NIH Video "NIH Policy: Enhancing Reproducibility through Rigor and Transparency"
  - UCSF Video "Data Reproducibility in Preclinical Discovery Is it a Real Problem?"
  - UAB presentation "New Year, New NIH Expectations: Are You Ready?"

More, more information

Nature article on quality science

Help

ResearchHelp@urmc.rochester.edu

More examples

REPRODUCIBILITY AND RIGOR: Background for the scientific premise of this project is described above. The literature contains conflicting reports on the use of ultrasound for soft tissue and bone healing. A weakness in some investigations is the lack of critical calibrations of acoustic fields. Additionally, many studies focus on a narrow range of acoustic exposure parameters, thereby limiting understanding of underlying mechanisms and optimization potential. Our proposed project addresses these concerns and others in regards to scientific rigor. Ultrasound fields will be thoroughly calibrated before and after each experiment, and we have proposed investigating how different acoustic parameters (e.g. frequency, intensity, pulsing parameters, exposure duration) influence efficacy. Furthermore, we have incorporated blinding and randomization to reduce bias, have clear laboratory practices for data collection and analyses and transparency in reporting results. We have a quality system of operation in our laboratories, and ensure regular and proper training of investigators involved with experiments. In this project, we have incorporated testing of two important biological variables: we include experiments comparing responses in normal and genetically-diabetic mice, and between male and female mice. An important biological resource is the genetically-diabetic mouse model. This strain will be purchased from Jackson Laboratories, and glucose levels will be monitored as metrics of diabetes for each mouse. The response of diabetic mice will be compared to their strain-matched, non-diabetic controls. At the initiation of a protocol, the treatment site (i.e., left or right dorsal ulcer) will also be randomly chosen; the contralateral ulcer will serve as an untreated control. During daily exposures of individual mice, the treatment order (including sham exposures) will be randomized using a random number generator to avoid grouping identical ultrasound protocols in time. Separate investigators will be responsible for assigning treatment protocols, performing ultrasound exposures, and collecting data. For protocols involving data acquisition, the investigator will be blinded to treatment conditions and investigators will not be made aware of the treatment allocations until all data have been collected and analyzed. Based on our earlier studies using diabetic mice, and our other studies using normal mice to evaluate bioeffects of ultrasound, we anticipate that 9-10 mice per group will be required to evaluate significance. Dose response models of the various acoustic exposure parameters are utilized and threshold dependency will be assessed. Statistical analysis will be performed by one-way ANOVA followed by Tukey's post-test. Results will be considered significant when $p<0.05$.

Denise Hocking U01

Strengths and limitations: Unlike previous studies of the PAC effect, we will have accurate measures of LTL at birth and of parental variables which are potential confounders, such as maternal age at conception, SES and race. We will also have measures of age-adjusted paternal and maternal LTLs, making it possible to examine the PAC effect through mechanisms that are not mediated by the traditional modes of heritability. Since the PAC effect is evident throughout the age range of PAC (1-5) that we will consider (20-44 yrs), and we will oversample trios in which the father is > 35 yrs, we will be able to examine both the magnitude and the shape (e.g. linear or not) of the relationship between PAC (20-44 yrs) and LTL at birth. Aim 1 also has limitations; for instance, we will not explore specific genetic/epigenetic factors (e.g., DNA methylation) and telomere length in the fathers‘ sperm, which might provide further insight into the PAC effect. We note that we have a track record of studying sperm (1) and plan to apply for funding of a project that will explore the impact of genetics /epigenetics on sperm telomere length.

Borrowed from Columbia

Questionnaire and medical record data.

Repeat questionnaires are administered by trained bi-lingual research workers to the child's mother or other guardian during pregnancy and every year until the child is age 3 and then every other year until the child is age 9-11 years. Questionnaires gather information on demographics, maternal marital status, maternal education and employment history, family income, access to basic necessities (food, shelter, clothing, home characteristics, residential history, history of active and passive smoking, maternal stress, maternal pre-pregnancy weight and height, history of prior births and medical history of the child. Data are also extracted from the maternal prenatal and newborn medical records, including newborn gender, gestational age, birth weight, length and head circumference, illicit drug and alcohol use during pregnancy, maternal prenatal medical conditions, delivery conditions, and weight gain during pregnancy. Data from questionnaires and medical records will be used to identify potential confounding as described below.

Borrowed from Columbia

Limitations and Strengths.

The competitive renewal is responsive to recent epidemiologic and experimental evidence indicating that phthalates modulate thyroid function and reduce circulating thyroid hormone levels. These findings have important implications for child cognitive and behavioral function, as thyroid hormones during pregnancy and early childhood are critical to brain development. Even modest reduction may impact child mental, motor and neuropsychological function. Our preliminary research has shown a significant inverse association between maternal prenatal phthalate exposures and child mental development at age 3 years. However, limitations in the study design need to be recognized. Phthalates are ubiquitous contaminants, and measuring exposures is always a challenge given the potential for contamination and the fact that biologic half-lives are short. To address this, we will use phthalate monoester levels in urine samples from the mother during pregnancy and the child between ages 3-11 as our primary dosimeter of exposure. As urinary lipase activity is negligible, monoester levels provide a reliable internal dosimeter for exposure. Further, based on the ICCs that we and others have documented in phthalate concentrations in repeat urine samples, urinary measures appear to provide reasonably reliable biomarkers of exposure.

Borrowed from Columbia

Limitations and Strengths. (continued)

This may be due to the fact that, even though biologic half-lives are short, exposure levels remain relatively constant in the home. Given the role of maternal thyroid hormone levels early in pregnancy on fetal brain development, it would have been ideal to have had first trimester maternal sera samples for thyroid hormone measures. This was not possible in the current study because most women were enrolled after the first trimester and also because our Community Advisory Board to the CCCEH had warned that the collection of a blood samples during pregnancy for research purposes would not be acceptable among the Dominican populations and could adversely affect our enrollment. **We did collect and store sera samples from the umbilical cord blood and, based on the reviewer recommendations, have now added funds to analyze thyroid hormone in these samples.** A major strength of the proposal is that it builds on an existing well-established cohort, and many of the required elements for testing study hypotheses are already being gathered within the parent study design. This results in a rich dataset and enables hypotheses to be tested in an efficient and cost-effective manor.

Borrowed from Columbia