

**OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP**

**REQUEST FOR APPLICATIONS TO PROVIDE A SIMULATED DATASET FOR THE  
OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP**

**RELEASED FEBRUARY 13, 2009**

**PROJECT TITLE: OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP (OMOP)**

**PROPOSAL DUE DATE:**

Friday, March 6, 2009 at 4:30 p.m. Eastern time in Washington, DC.

**EXPECTED TIME PERIOD FOR CONTRACT:**

April 2009 to July 2009

**ELIGIBILITY:** This procurement is open to those organizations that satisfy the qualifications stated herein and that are available for work.

**CONTENTS OF THE REQUEST FOR APPLICATION:**

1. Introduction
2. General Information
3. Proposal Contents
4. Evaluation

## 1.INTRODUCTION

### 1.1 PURPOSE AND BACKGROUND

The Observational Medical Outcomes Project hereafter called "OMOP," is initiating this Request for Application (RFA) to solicit proposals from organizations interested in participating in designing and developing a simulated dataset to support its methodological research.

#### **Background:**

The Observational Medical Outcomes Partnership (<http://omop.fnih.org>) is a public-private partnership designed to protect human health by improving the monitoring of drugs for safety and effectiveness. The partnership began in Q4-2008 and will conduct a two-year research initiative to determine the contribution and utility of using existing healthcare databases to identify and evaluate safety issues of drugs already on the market.

OMOP draws on the expertise and resources of the pharmaceutical industry, academic institutions, non-profit organizations, the Food and Drug Administration (FDA), and other federal agencies. It is funded and managed through the Foundation for the National Institutes of Health. In addition to sponsoring specific research efforts, OMOP will create a set of tools—such as data models, experimental protocols, and database evaluation tools—that will be placed in the public domain to encourage research by a broad community of scientific investigators. All project results will be made public in accordance with the public health mission of the partnership. These will include comprehensive reports on scientific and technical findings, lessons learned, and peer-reviewed articles on the experimental findings by our sponsored investigators.

To facilitate its methodological research, the Partnership will evaluate the performance of various analytical methods identifying drug-outcome associations across multiple disparate observational data sources (administrative claims and electronic health records). OMOP will partner with a modest number of organizations with observational data to collaborate in this research. Methodological research typically requires some benchmark to measure performance, so that the analysis results can be compared against a 'gold standard'. In this context, a desired gold standard would be a true causal relationship between a drug and a health outcome. Unfortunately, most observational data sources are poorly characterized and clinical observations may be insufficiently recorded or poorly validated. It can be difficult to obtain 'true' relationships between drugs and outcomes as these 'known associations' may not be observed due to issues including sample size, adequacy of data capture, and confounding.

As such, the Partnership is soliciting support to design and develop an automated procedure to construct simulated datasets to supplement the methods evaluation. The simulated datasets will be modeled after real observational data sources, but will be comprised of hypothetical persons with fictional drug exposure and health outcomes occurrence. The datasets will be constructed such that the relationships between the fictional drugs and fictional outcomes are well-characterized as 'true' and 'false' associations. That is, hypothetical persons will be created and assigned fictional drug exposure periods and instances of health outcomes based on random sampling from probability distributions that define the relationships between the fictional drugs and outcomes. The relationships created within the simulated datasets will be contrived, but will be representative of the types of relationships expected to be observed within real observational data sources. The Partnership will identify and define the dimensions, categories, and associated probability distributions that are required to characterize the types of scenarios necessary to be included within the simulated dataset. Example dimensions and categories are provided in Section 1.3 as an illustration.

The simulated datasets will only be used to perform statistical evaluations of the analytical methods offered to identify drug-outcome associations. The performance of the analytical methods can be empirically measured in terms of sensitivity, specificity, positive and negative predictive value; the known characteristics of the data will enable the classification of the drug-outcome relationships as 'true' or 'false' and methods will be executed to classify the drug-outcome pairs as 'positive' or 'negatives'. Because the simulated data will represent hypothetical patients, fictional drug classes and artificial outcomes types, there can be no clinical interpretations drawn from the data.

It is expected that the procedure will be used to create two publicly available simulated datasets. The first simulated dataset will be made available along with the document outlining the expected relationships contained within the data. This dataset will be used to encourage methods development from the broader research community, and enable researchers to evaluate the performance of their methods for themselves. The second simulated dataset will be made available without publication of the expected associations, and will be used to enable independent replication and validation of methods findings. That is, researchers will be encouraged to provide the methods implementation and summary results to the Research Core, so they can be compared to the 'true relationships' to measure method performance. The simulated data structure, including table names, field names and data types, will conform to the Partnership's Common Data Model. The simulated populations may be at least 100 million hypothetical persons. The number of fictional drugs and outcomes required is to be determined based on the final set of dimensions and categories determined by the OMOP Principal Investigators and Advisory Boards.

## **1.2 SCOPE OF WORK**

The solicitation is for the design and development of a simulated observational database that mimics known relationships between fictional drugs and health outcomes.

The Applicant will:

1. Design and develop a procedure for constructing simulated datasets of hypothetical patients with varying drug exposure and outcomes occurrence that follow defined specifications based on agreed dimensions and categories that mimic relationships expected to be observed in real observational data sources:
  - a. The procedure should allow for configurable parameters for population size and the dimension probability distributions
  - b. The program should be developed in SQL, SAS or R
2. Assist the research team in executing the procedure within the OMOP research lab to construct two simulated datasets of 100 million hypothetical patients
3. Develop a final report that describes the process used to create the dataset, issues that arose and how they were addressed
4. Participate in OMOP Advisory Board meetings as appropriate

Deliverables include the source code and the final report. Both deliverables will be made publicly available through the OMOP website.

### 1.3. POTENTIAL OBSERVATIONAL DATA DIMENSIONS TO CONSIDER

The OMOP Research team is actively engaged in a process to define the dimensions and categories that are representative of the types of relationships expected to be observed in real observational databases. These dimensions will serve as the basis for the simulated dataset design and will inform the selection of the probability distributions required to sample patient characteristics, periods of drug exposure, and outcomes occurrence. Examples of dimensions under consideration are provided below as illustrations of the types of characteristics that may be modeled within the simulated dataset construction procedure. While these examples only serve as illustrations, it can be reasonably expected that the number and type of dimensions will be comparable. There will likely be less than 10 dimensions, each with 3-5 categories that would require definition of probability distributions to sample from. The proposal should demonstrate the Applicants' understanding of key issues with observational data that require simulation by describing the sources of information that would be considered for use in defining the dimensions, categories, and associated probability distribution parameters.

Prevalence of drug use: the proportion of the overall population have at least one period of exposure for each drug category

- High use drug (~10% of population)
- Medium use drug (~1% of population)
- Low use drug (~.001% of population)

Length and variability of drug exposure: among persons who have been exposed to a particular drug, the characterization of the exposures in terms of the periods of time of use (start date, end date)

- Acute use (one exposure, with constant 7 day treatment)
- Adherent chronic use (one period of persistent exposure, variable duration up to 5 years)
- Non-continuous chronic use (multiple periods of exposure over time, variable cumulative exposure up to 5 years)

Incidence of outcome in background population: the rate of occurrence of each outcome of interest across all persons.

- Commonly occurring (~20% of overall population)
- Rare (~.001% of population)
- Does not occur

Strength of drug-outcome association: the magnitude of relationship between the drug and outcome can be used to infer the rate of outcome within the exposed population, relative to the incidence rate of the outcome from across the background population. Note, to mirror reality, the large majority of drug-outcome pairs should have 'no association'

- Strong positive association
- Weak positive association
- No association
- Negative association

Timing of drug-condition relation onset: Among those patients who experience the outcome following drug exposure, it is the natural of the temporal relationship of when the outcome occurs relative to drug initiation

- Acute onset (outcome occurs immediately following drug exposure)
- Insidious onset (outcome occurs any time following exposure)
- Delayed onset (outcome occurs after 1 yr of cumulative exposure, or some time following last exposure)

Degree of confounding: For each drug-outcome relationship, how much is the magnitude of the association modified by additional variables related to both drug and exposure. Note: if confounding is introduced into the simulated data, the confounding covariates need to be modeled accordingly

- No confounding variables
- Confounding by age, gender only
- Weak confounding by prior drugs/conditions
- Strong confounding by prior drugs/conditions

Adequacy of outcome coding capture: the confidence that the observation of an outcome in the dataset truly represents the outcome having occurred to the patient at the time specified. Note: in real observational data sources, outcomes are commonly validated to determine performance characteristics.

- 100% positive predictive value (PPV), negative predictive value (NPV)
- good PPV (e.g. >80%), poor NPV (e.g. <20%)
- poor PPV (e.g. <20%), good NPV (e.g. >80%)
- poor PPV (e.g. <20%), poor NPV (e.g. <20%)

Outcome type: the nature of how each outcome would be recorded within the dataset

- Chronic condition, with multiple instances of outcome definition over time related to existence of one common disease (e.g. something like 'diabetes' where a condition may be recorded repeatedly over time, but the primary entry of interest is the first onset diagnosis)
- Acute condition, with multiple instances of outcome over time related to unique episodes of care (e.g. something like 'anaphylactic shock' where sequential diagnoses over time are more likely be new incident events)
- Other outcomes from other data elements beyond diagnosis codes (e.g. hospitalizations, death)

#### **1.4 REQUIRED QUALIFICATIONS**

The ideal candidate will be principally responsible for the development of an algorithm that generates a database containing hypothetical patients with fictitious drug exposure and condition occurrence that are related in predefined ways.

##### Qualifications: The Applicant must:

- Have 10 years experience in managing healthcare claims or electronic healthcare data and experience with patient data research
- Knowledge of medical informatics and / or biostatistics
- Knowledge of creating datasets for relational database management systems
- Proficient use of standard office computer software and statistical tools

- Proven ability to meet deadlines and produce quality work in this area
- Ability to interact and verbally communicate with groups of managers, clients, customers, and the general public
- Review and coordinate the preparation of reports, papers, and presentations for the OMOP Research Core and Advisory Boards

### **1.5 PERIOD OF PERFORMANCE**

The period of performance of any contract resulting from this RFA is tentatively scheduled to begin in April 2009 and to end on July 31, 2009. Amendments extending the period of performance, if any, shall be at the sole discretion of OMOP.

## 2. GENERAL INFORMATION

### 2.1 RFA COORDINATOR

The RFA Coordinator is the sole point of contact for this procurement. All communication between the Applicant and OMOP upon receipt of this RFA shall be with the RFA Coordinator, as follows:

Name	Emily Welebob
Address	818 Connecticut Avenue, N.W., Suite 500
City, State, Zip Code	Washington, D.C. 20006
Phone Number	703-508-6225
E-Mail Address	ewelebob@fnih.org

Communication directed to parties other than the RFP Coordinator may result in disqualification of the Consultant.

### 2.2 ESTIMATED SCHEDULE OF PROCUREMENT ACTIVITIES

OMOP reserves the right to revise the schedule.

Issue Request for Application on OMOP Website	February 13, 2009
Question and answer period	February 13, 2009 to February 27, 2009
Proposals due	March 6, 2009
Evaluate proposals	March 9 – 20, 2009
Announce “Apparent Successful Applicant” and send notification via e-mail to unsuccessful proposers	March 27, 2009
Negotiate contract & begin work	April

### 2.3 SUBMISSION OF PROPOSALS

Proposals shall be submitted in hard copy (1 original) and sent electronically. Proposals must be received by OMOP no later than 4:30 p.m. Eastern, on March 6, 2009. The proposal is to be sent to the RFA Coordinator at the address and email noted in Section 2.1.

Late proposals will not be accepted and will be automatically disqualified from further consideration. The proposals must respond to the procurement requirements. Do not respond by referring to material presented elsewhere. The proposal must be complete and must stand on its own merits.

Failure to respond to any portion of the procurement document may result in rejection of the proposal as non-responsive. All proposals and any accompanying documentation become the property of OMOP and will not be returned.

### 2.4 PROPRIETARY INFORMATION/PUBLIC DISCLOSURE

Materials submitted in response to this competitive procurement shall become the property of OMOP.

**2.5 COSTS TO PROPOSE**

OMOP will not be liable for any costs incurred by the Applicant in preparation of a proposal submitted in response to this RFA, in conduct of a presentation, or any other activities related to responding to this RFA.

**2.6 REJECTION OF PROPOSALS**

OMOP reserves the right at its sole discretion to reject any and all proposals received without penalty and not to issue a contract as a result of this RFA.

### 3. PROPOSAL CONTENTS

Proposals must be submitted as Microsoft Word documents. Supporting materials should also be provided electronically. The major sections of the proposal are to be submitted in the order noted below:

1. Project Approach / Methodology
2. Work Plan and Schedule
3. Deliverables
4. Staff and Organizational Qualifications / Experience
5. Cost Proposal

Proposals must provide information in the same order as presented in this document with the same headings.

#### 3.1 PROPOSAL CONTENTS

The Technical Proposal must contain a comprehensive description of services that address the scope of work described in Section 1.2. The proposal should including the following elements:

- A. Project Approach / Methodology** – Include a complete description of the proposed approach and methodology for the project as explained in the Scope of Work Section 1.2. This section should convey the Applicant’s understanding of the proposed project, as well as describe any experience on similar projects.
- B. Work Plan and Schedule**– Include all project requirements and the proposed tasks, services, activities, etc. necessary to accomplish the scope of the project defined in this RFA. This section of the technical proposal must contain sufficient detail to convey to members of the evaluation team the Applicant’s knowledge of the subjects and skills necessary to successfully complete the project. Include a project schedule indicating when the elements of the work will be completed and when deliverables, if any, will be provided.
- C. Deliverables** – Fully describe how the dataset and source code to be submitted under the proposed contract will be delivered.
- D. Staff and Organizational Qualifications / Experience** – Identify staff who will be assigned to the potential contract, indicating the responsibilities and qualifications of such personnel, and include the amount of time each will be assigned to the project. Provide résumés for the named staff, which include information on the individual’s particular skills related to this project, education, experience, significant accomplishments and any other pertinent information. The Applicant must commit that staff identified in its proposal will actually perform the assigned work. Any staff substitution must have the prior approval of OMOP.
- E. Cost Proposal**  
Identify all costs including expenses to be charged for performing the services necessary to accomplish the scope of work of the contract. The Applicant is to submit a fully detailed budget including staff costs, administrative costs, travel costs, and any other expenses necessary to accomplish the tasks and to produce the deliverables under the contract. Costs for subcontractors are to be broken out separately.

## 4. EVALUATION

### 4.1 EVALUATION PROCEDURE

Responsive proposals will be evaluated strictly in accordance with the requirements stated in this solicitation and any addenda issued. The evaluation of proposals shall be accomplished by an evaluation team, to be designated by OMOP, which will determine the ranking of the proposals.

### 4.2 CLARIFICATION OF PROPOSAL

The RFA Coordinator may contact the Applicant for clarification of any portion of the proposal.

### 4.3 EVALUATION WEIGHTING AND SCORING

The following weighting and points will be assigned to the proposal for evaluation purposes:

<b>Proposal Section</b>	<b>Totals</b>
a. Project Approach/Methodology	40
b. Work Plan and Schedule	10
c. Project Deliverables	5
d. Staff Qualifications/Experience	30
e. Costs	15
<b>GRAND TOTAL FOR WRITTEN PROPOSAL</b>	<b>100 points</b>