

V. Description of Specific Tasks

While the technology that would be used for linking the records from these three programs has been proven to be effective in a number of prior studies, there are a number of tasks that must take place that will require the cooperation of both Federal and State agencies. Below is an overview of each of the tasks, including a listing of proposed subtasks. In addition, at the end of each task is a summary of the cost information for that task. In order to be consistent with Kenyon, et. al (2001) and Wittenburg and Alderson (2001) we assume that each agency has three labor categories; Senior Manager, Senior Analyst, and Research Assistant. The Senior Manager category includes experts who have at least ten years of research and /or program experience. The Senior Analyst category includes researchers and technical staff that have between three to nine years of experience. The Research Assistant category includes individuals with limited experience.

Estimates for labor hours were derived from examining prior efforts to conduct data merges at the State level, prior experience of the principle investigator in implementing the WIC minimum dataset, and experience from other federal agencies conducting research with data merging. In particular, information provided by the Maternal and Child Health Information Resource Center, which is funded by the Bureau of Maternal and Child Health (HHS) and administered by HSR, was very helpful in estimating labor hours.

The following tasks are proposed for creating the database.

Task 1. Conduct Initial Meeting Among Federal Agencies

One of the key issues that will need to be addressed by Federal and State officials is the need to develop data sharing agreements between the WIC Program, the State Office of Vital records, and the State Medicaid program. In order to facilitate this process, it is proposed that an initial meeting be conducted between representatives from USDA and those from offices within the Department of Health and Human Services responsible for the state-level programs included in this initiative. The meeting should include representatives from the National Center for Health Statistics (vital records), the Medicaid program, and possibly representatives from the Maternal and Child Health Bureau.

The purpose of this initial meeting would be to brief the representatives from these agencies on the proposed initiative and request their involvement and cooperation. A number of products could come from this meeting, including a joint letter of agreement supporting the initiative that could be shared with State agencies, a working “steering committee” to help oversee the implementation, and a sample data sharing agreement that could be used by States to meet both federal and state data confidentiality requirements.

Subtasks:

- Identify potential participants to be included in initial meeting.
- Prepare summary of project and distribute to invited participants.
- Conduct meeting.
- Identify committee to prepare model data sharing agreement.
- Prepare and distribute model data sharing agreement

Federal Costs:	\$12,056
State Costs:	0
Total Task Cost:	\$12,056

Task 2. Conduct a Series of Meetings with State Program Representatives and steering committee members.

Once there is agreement among the Federal Agencies, the next task would be for USDA to conduct a meeting with representatives from each of the three State programs involved in this initiative. This meeting would be with representatives from the National Association of WIC Directors, The National Association of State Registrars, and the National Association of State Medicaid Directors. The purpose of the meeting would be to provide information about the initiative, gain their support and cooperation with the initiative, and develop a workgroup to help oversee the implementation process and solve problems as they occur. It will be critical to have support and input from these agencies as the initiative progresses.

At this meeting, it will be important to identify a core group of State and Federal staff who can serve as a steering committee for the project. The committee should represent all of the major players in the initiative. This committee will have responsibility to oversee both the developmental and testing phase of the project, as well as the implementation of the national database.

Subtasks:

- Identify and invite participants to initial meeting.
- Prepare and distribute background materials to meeting participants.
- Conduct meeting and develop Steering Committee for ongoing oversight of the initiative.

- Prepare joint letter to State Health Departments and Medicaid Agencies from committee describing the initiative and requesting State agency support.
- Conduct additional Steering Committee meetings as the project progresses.

Federal Costs:	\$39,772
State Costs:	\$17,627
Total Task Cost:	\$57,399

Task 3. Develop a Nationally Representative Sample Plan

Deciding on how to construct the sample frame for the database will require considerable thought on the part of USDA and other federal officials. While developing sample frames for national studies are not new, the fact that the ultimate goal is to link the data with other participant databases may provide for opportunities not commonly considered in developing samples. In the case of this initiative, the sample will be designed to allow for both state-level analysis of data as well as national research. This means that a larger number of WIC participants will be included in the sample than if the purposed was to examine national issues alone.

Another key issue that will need to be decided is at what point during WIC participation the record should be entered into the database. While the database should be representative of women, infants and children enrolled in WIC, there may need to be some consideration given to the age of the child and length of participation prior to inclusion. For example, if a four-year old is added to the file, and then data are linked, there will be more data available for that child than a two-year old. However, if the four-year old is only recently enrolled in WIC, then there will be less WIC data available than for a two-year old that has participated since birth. This may mean that the sample may need to be controlled for the child’s age and length of participation.

Once these decisions are made, the sample can be drawn. The sample frame will need to be constructed using data currently available in the participant characteristics study, but actual data to complete the sample will need to be obtained from the State WIC program participant database.

Subtasks:

- Identify model and structure for national sample that includes the ability to examine state-level data.
- Develop sampling plan for national data collection.

- Develop State-by-State sample requirements to assure validity of State-level data.
- Prepare and submit sampling plan to oversight committee for review.

Federal Costs:	\$7,599
State Costs:	\$2,820
Total Task Cost:	\$10,419

Task 4. Create the Database Structure for Data to be Linked.

This next step will require that the specific data elements to be contained in the database be identified, and a database be constructed in a form that allows for analysis. The database structure would be hierarchical, creating a header record that links all data back to a family unit tied to the mother. Then separate detail records are created containing information about the mother and child, including the health status of the mother and pregnancy information, birth outcome data, and information about the child. The database should be large enough to capture the essential demographic components of the WIC participant, as well as health outcome information contained in the records. For example, the WIC record for a pregnant woman will contain both demographic information as well as health outcome information such as weight gain during pregnancy and nutritional risk factors. The birth certificate information will also contain numerous demographic data for the baby’s parents, as well as information about the pregnancy. The Medicaid record will also contain information about health care utilization, such as prenatal care visits.

As a result, the database structure will need to be carefully constructed in order to recognize duplicate information that will be available from each of the three files. For example, in the case of a pregnant woman, it is likely that all three files will contain information regarding her age, racial or ethnic identification, weight gain during pregnancy, and other demographic and service-related factors. The database will need to be structured in such a way that duplications can be recognized and inconsistencies resolved.

Additional consideration will need to be made when linking the child’s file to the mother’s. The database will need to be structured in such a way that individual information about the parent and child can be extracted for research purposes. At the same time the data file must be able to create associations between the mother’s pregnancy, the birth outcome, medical participation and ongoing WIC participation.

In implementing this task, it would be advisable to develop a participatory process that would include representatives from the federal and state agencies involved, as well as independent researchers. Having this type of input may make it easier to construct a workable database as well as avoid pitfalls that other researchers may have experienced.

Subtasks:

- Identify key data elements to be included in the merged database, including the development of consistent data definitions and acceptable ranges of values;
- Identify potential duplications of data that will need to be resolved through edits;
- Identify an appropriate structure for the database, including the structure of the hierarchical database, how to display family data within the hierarchical structure, and how the overall structure of the database can be made user-friendly
- Develop security and access protocols for the database; and
- Develop methods by which client identifiers are stripped from the records prior to submission to USDA.

Federal Costs:	\$19,514
State Costs:	\$0
Total Task Cost:	\$19,514

Task 5. Create the Protocols for Linking the Three Databases.

Once the sampling method is developed, the next step is to develop the protocols that will be used to link the WIC record with the Medicaid and vital record information. This can be done using a number of methods. Because the data files contain a significant amount of demographic data about the clients, matches can be created by identifying common demographic characteristics. As has been noted, matching WIC records with birth records has been successfully accomplished in the past, and would likely be the first two databases to be linked. For example, the WIC, Medicaid, vital records links that occurred in the Missouri studies accomplished a 93.6% linkage in 1994 and a 95% linkage rate in 1997. In some cases, records can be linked if the mother's social security number is available. However, in most cases, WIC programs do not collect social security numbers, so matching will need to take place using demographic data such as name, birthday, date of delivery, and other client demographic data.

After a combined WIC/birth record match has been established, the next step is to determine whether or not the individual WIC participant also is enrolled in Medicaid. This can be done by matching the newly linked database to the state's Medicaid enrollment file. The newly created database will then have information as to whether the participant was enrolled in Medicaid or not. This will allow for outcome comparisons of WIC clients who are enrolled in Medicaid as compared to those who are not. In addition,

some States will have information on services provided through the Medicaid program, such as the immunization status of children or the prenatal visits of the mother, can be captured and analyzed.

Subtasks:

- Examine prior efforts to link the three databases and develop a range of specific protocols and algorithms that will be considered for developing model methods for data linkages;
- Develop a technical panel to review the prior efforts and make recommendations as to the best protocols to recommend and implement; and
- Prepare a document that provides guidance to State agencies on the best methods by which data can be linked.

Federal Costs:	\$16,667
State Costs:	\$17,627
Total Task Cost:	\$34,294

Task 6. Develop Access Protocols and Guidelines for Removing Client Identifying Information

The next step will be to develop recommended protocols for remove identifying information about the client and develop methods by which researchers will be able to gain access to the database. In addition, decisions will need to be made about how often to update the database. To provide quality information, it is recommended that the database be updated at least every two years with new data. The updating procedure should be able to be linked with the prior database in order to update client records, and allow for tracking of certain clients over time. By using the birth record number as an identifier, information about the delivery of pregnant women included in the prior database can be updated. The resulting file will then be available for research purposes. It is recommended that the database be maintained by and made available through USDA.

Subtasks:

- Identify key client identifiers that must be removed from the database in order to ensure client confidentiality;

- Review methodologies used in prior studies linking data from these sources to develop the best method by which to remove client identification information; and
- Develop a method by which dummy client identifiers can be inserted so comparisons can be run once the data are linked.

Federal Costs:	\$5,558
State Costs:	\$4,230
Total Task Cost:	\$9,789

Task 7. Test the Protocols for Data Linkage Integrity and Database Functionality.

One of the key elements of this initiative will be to test the protocols that have been developed by asking a limited number of States to create a linked database using the developed protocols. In order to avoid having to obtain clearance from the Office of Management and Budget during this test phase, it is recommended that fewer than nine states be selected to test the protocols. The States selected for the test should be ones that have had experience linking their WIC files with Medicaid and vital records data. Each State would be asked to develop their data sharing agreements and then follow the linkage protocols developed by USDA. These data would be submitted to USDA electronically and placed into the newly designed database. Once the data were in place, tests should be run on the data to assess the functionality of the database and query software.

Subtasks:

- Identify testing methods and protocols;
- Identify States that have linked records in the past, and would be willing to test the data merge protocols;
- Develop and implement data sharing agreements in the selected States;
- Run test data linkages in each State and submit data to trial database;
- Develop and test query software on the sample database; and
- Review results of the testing and make appropriate adjustments to the data collection and merger protocols and/or query software.

Federal Costs:	\$32,861
State Costs:	\$48,448

Total Task Cost:	\$81,309
-------------------------	-----------------

Task 8. Revise Protocols Based Upon Test Results.

After completion of the testing, the protocols may need to be revised to incorporate lessons learned. Once the protocols are revised, it is recommended that representatives from the nine test states review the protocols to be sure they are still functional and have not created any new problems.

Federal Costs:	\$8,521
State Costs:	\$4,230
Total Task Cost:	\$12,715

Task 9. Prepare OMB Clearance Package.

It is not known whether or not this initiative will require clearance from the Office of Management and Budget (OMB), but because data are being merged from different Federal programs, it is assumed that such clearance will be required. Upon completion of the testing phase, the next step will be to request clearance from OMB to collect data from all the States that will be included in the sample. This package should emphasize that these data are already being collected by State agencies, and the linking of these existing database will lessen State-level burdens when conducting future research on WIC outcomes.

Subtasks:

- Prepare Federal Register notice of data collection;
- Prepare and submit OMB clearance package; and
- Make revisions to clearance package as required by OMB.

Federal Costs:	\$6,742
State Costs:	\$0
Total Task Cost:	\$6,742

Task 10. Implement the Initiative on a National Level.

Once OMB clearance is granted, the next step will be to implement the initiative on the national level. It will be important for there to be technical assistance available to States during the first year in order to help them with implementing the protocols. Each

State would complete the data linkage, and then submit the sample data to USDA. Once the file is complete, USDA will need to clean and edit the file, and then can prepare a public access file for use by researchers.

Subtasks:

- Develop time table for full implementation of the initiative;
- Notify States of timetable and submission requirements;
- Ensure that proper data sharing agreements are in place within each State;
- Ensure that appropriate data collection and merge protocols are in place;
- States WIC program obtain Medicaid and Vital Record data and complete data merge;
- Run State-level edits to remove duplicate information and clean data;
- Submit Data to USDA;
- Edit and clean data at Federal level; and
- Implement query software and make data available to researchers

Federal Costs:	\$44338
State Costs:	\$175,552
Total Task Cost:	\$219,860