T-3537

THE GENERATION OF DISEASE OCCURRENCE AND POTENTIAL ENVIRONMENTAL EXPOSURE VALUES FOR AN ECOLOGIC STUDY OF CRANIOSYNOSTOSIS IN COLORADO

by

James F. Uhlinger

T-3537

A thesis submitted to the faculty and the Board of Trustees of the Colorado School of Mines in partial fulfillment of the requirements for the degree of Master of Sciences (Engineering Ecology).

Golden, Colorado
Date: Nov. 11, 1988

Signed: _____
James F. Uhlinger

Approved: _____
Dr. R. R. Cohen
Thesis Advisor

Golden, Colorado
Date: Nov. 11, 1988

_____
Dr. John C. Emerick
Professor and Acting Head
Engineering Ecology Department

ABSTRACT

Ecologic studies of craniosynostosis have been initiated by the Colorado Craniosynostosis Project to evaluate possible environmental causes of the birth defect. To best utilize the ecologic study design, disease occurrence and potential environmental exposure values were derived for environmentally significant geographic areas. This required the development of geographic mapping and plotting procedures to create Universal Transverse Mercator (UTM) coordinates for each of 1353 study subjects. Computerized data capture techniques were utilized to create and validate a disease incidence data set comprised of UTM coordinates and study subject identification codes. This data set was loaded to a geographic information system (GIS) which automatically assigned United States Geological Survey (USGS) Hydrologic Unit codes to each address location. Environmental data was acquired and formatted for integration with the disease incidence data on the GIS. USGS Hydrologic unit codes were automatically assigned to mineral industry locations derived from the US Bureau of Mines MASNC data set for Colorado. Finally, these hydrologic unit codes were used to sort and count the disease incidence data to create disease occurrence values, and the mineral industry location data to create potential environmental exposure values. The result was the creation of a data set consisting of frequency counts of cases,

controls and mineral industry locations by type of commodity for hydrologic units throughout the state. Thus, data were generated that would support ecological analyses of the relationship between the occurrence of craniosynostosis and number of particular types of mineral industry locations in Colorado. The generation of these values as well as intermediate products generated as a result of this process are identified as results with application to other epidemiologic studies. Secondly, GIS was shown to be an effective tool for maintaining, manipulating, and displaying geographically referenced disease occurence and potential environmental exposure data for use in epidemiologic studies.

TABLE OF CONTENTS

LIST OF FIGURES

Figures

LIST OF TABLES

Tables

ACKNOWLEDGEMENTS

This study was performed with cooperation from numerous individuals and organizations. Specific thanks go to Dr. Beth Alderman and the entire staff of the Colorado Craniosynostosis Project, for guidance, support, and advice on epidemiologic issues; Dr. Ellen Mangione and the entire staff of the Disesase Control and Environmental Epidemiology Division of the Colorado Department of Health, for flexibility in the use of time and resources that facilitated the completion of this study; Harold Bennett of the U. S. Bureau of Mines Minerals Availability Division, for the cooperation of his staff in providing the MASNC data; Dave Litke of the USGS Water Resources Division, for providing Hydrologic Unit data and advise; Alex Goetz, Leeann Ayers, and the other CIRES personnel, for their patience and expertise in the utilization the ARC/INFO GIS; my thesis committee members at the Colorado School of Mines for their patience and interest in this project; and finally, my wife Diane, for her tolerance and patience throughout this endevour.

Chapter 1

INTRODUCTION

Craniosynostosis is a birth defect that has been reported in high incidence in Colorado from 1978 to the present. A literature review reveals that documentation pertaining to the etiology of craniosynostosis is limited to genetic and clinical causes. The incidence rates presented by this documentaion of known causes do not account for the rates documented in Colorado (PHS CDC Correspondence EPI-83-56-2, 1987; PHS CDC Correspondence EPI-86-42-2, 1988). Thus, potential environmental causes need to be evaluated.

The only identified documentation that addresses potential environmental causes of craniosynostosis is a masters thesis done by M. K. Anderson (1984). It is a survey of environmental agents and exposure pathways based on the environment and population distribution in Colorado. Various techniques are used in this survey for determining relationships between the identified environmental agents and disease incidence data. Anderson graphically compares case distributions to mineral belt, coal mining, and hazardous waste site distribution maps to show potential environmental exposures. This is done by plotting cases on base maps depicting the areal extent of Colorado mineral belt and coal mining regions and locations of hazardous waste sites in Colorado. Environmental agents such as

heavy metals are implicated in discussions of physiological mechanisms. Air quality measures are graphed as a function of time and compared to incidence reported as a function of time.

Along with Anderson, concerned parents, public health workers, medical and scientific experts, and others have presented numerous conjectural hypotheses pertaining to environmental causes of craniosynostosis. These conjectures include (but are not limited to) implication of mine operations (Anderson, 1984), elevation related oxygen stress (Anderson, 1984; Ruenner, 1986), chemical water quality (Anderson, 1984; Craft, 1987; Coillouette, 1987), and effects from use of silver iodide in cloud seeding (Marshall, 1986) as examples. These presented hypotheses are interesting and often clever, but provide little or no substantial scientific evidence on which to justify a research proposal that requires funding or other expense.

The Colorado Craniosynostosis Project was created in April 1986 to address the public health concerns posed by the reported high incidence of the birth defect in Colorado. The work plan for the Colorado Craniosynostosis Project states that "[t]he mission of the project is to promote the better understanding of the causes, prevention, natural history, and the treatment of [the birth defect] craniosynostosis in Colorado." To assist in achieving the goals of this mission, environmental epidemiologic studies

have been initiated to statistically evaluate the relationship of disease occurrence to potential environmental exposures (Alderman, 1986). These studies are area based in design and thus, called ecologic studies.

Ecologic studies have been documented as a means of generating hypotheses by identifying statistical relationships between characteristics of disease occurrence within human populations and characteristics of the environments in which those populations reside. For this type of study, geographical area boundaries define study populations (Morgenstern, 1982). For these populations, some measure of disease occurrence (relative or discrete) and potential environmental exposure (direct or indirect) must be obtained (Mausner and Bahn, 1974; Rothman, 1985; Kelsey et al., 1984). These measures are compared to those of other geographic areas using various analysis techniques and relationships between the two values, if any, defined statistically (Morgenstern, 1982; Gesler, 1985; Mausner and Kramer, 1985). If a specific relationship is discovered, it can be used to support the generation of etiologic hypotheses for the disease (Gesler, 1985; Rothman, 1985; Kelsey et al., 1984), then specific exposure studies can be designed and conducted to test these hypotheses (Stavraky, 1976; Gesler, 1985; Rothman, 1985; Kelsey et al., 1984).

The Colorado Craniosynostosis Project has initiated ecologic studies of craniosynostosis to systematically study

hypotheses pertaining to environmental factors influencing the occurrence of the birth defect. These studies are expected to statistically characterize relationships between disease occurrence and potential environmental exposure, prioritize and support presented conjectures, and assist in the generation of causal hypotheses (Alderman, 1986).

Statewide disease outcome data (i.e. craniosynostosis incidence data) for Colorado have been compiled and were identified for use in a pilot ecologic study. This data set consists of birth certificate records for cases identified by the Federal Centers of Disease Control (CDC) and controls selected from all births occurring in Colorado for the period of 1978 through 1985. These data include locational parameters as presented in addresses and codes that refer to county of residence. Both sets of locational parameters refer to the mother's residence when the afflicted child (case) or the unafflicted child (control) was born.

A survey was conducted to identify applicable environmental data, geographic data, and data processing resources for use in a pilot study. Two appropriate data resources were identified. These are the United States Geological Survey (USGS) Hydrologic Units for Colorado which can serve as geographic boundaries and United States Bureau of Mines Mineral Availibility Survey (MAS) Database as a source from which potential environmental exposure could be developed. These resources were selected because they

contain statewide locational data components and allow for the generation of several types of potential environmental exposure variables.

Thus, data exist to characterize disease occurrence and potential environmental exposure within delineated geographic areas in Colorado. It is the intention of the Colorado Craniosynostosis Project to develop procedures and an environment in which to utilize these disease outcome data in their most basic format and iteratively test for relationships with potential environmental exposure data derived from mine site data using the ecologic study design. This requires the development of procedures to integrate the existing data using manual and computerized geographic techniques to generate disease occurrence and potential environmental exposure values for specific areas within the state. Once identified and utilized in the pilot study, it is expected that these procedures, equipment resources, and data can serve as a framework and example to conduct additional studies utilizing other existing disease incidence data or environmental data.

The general objective of this thesis is to document the procedures I have designed to develop geographically referenced disease occurrence and potential environmental exposure values used in the pilot study. This includes the presentation of issues and criteria that were considered to generate these pertinent geographically based values.

Specific tasks were required to implement the data integration and the creation of values for disease occurrence and potential environmental exposure for the designated geographic areas. These are identified as follows:

1) evaluation of the disease occurrence data and their transformation into an appropriate format;

2) evaluation of the environmental data and its utility as a measure of potential environmental exposure; and

3) creation of disease occurrence and potential environmental exposure values to be utilized in epidemiologic analyses.

The general objective is discussed in the following background section. The background section describes in greater detail the integration of concepts from the various disciplines of Environmental Epidemiology, Geography, Environmental Sciences, and Data Management. The specific tasks are described in greater detail in the methods and in the results section.

Chapter 2
BACKGROUND

Craniosynostosis

Craniosynostosis is a rare birth defect (see Appendix A) that was detected in Colorado at elevated rates compared to the Federal Centers For Disease Control (CDC) Metropolitan Atlanta comparison group studied during the same time period (1978-1985) (see Figure 1).

Even though the comparison of the Colorado and Atlanta study results is difficult due to confounding factors such as differences in diagnostic criteria, reporting, time scale, and demographics, the results can serve as an initial indication of an increased incidence rate (epidemic) within areas of Colorado. Other proof of the existence of an epidemic of craniosynostosis reported in Colorado includes documented clustering in Steamboat Springs (PHS CDC Correspondence EPI-83-56-2, 1987; PHS CDC Correspondence EPI-86-42-2, 1988).

The high incidence of craniosynostosis in Colorado presents a public health concern. It is a poorly understood disfiguring birth defect that affects the entire family of the afflicted. The developmental implications are not understood and the current remedy is expensive and invasive surgery involving the skull of the afflicted child. The need for sugery is left to the judgement of the physician

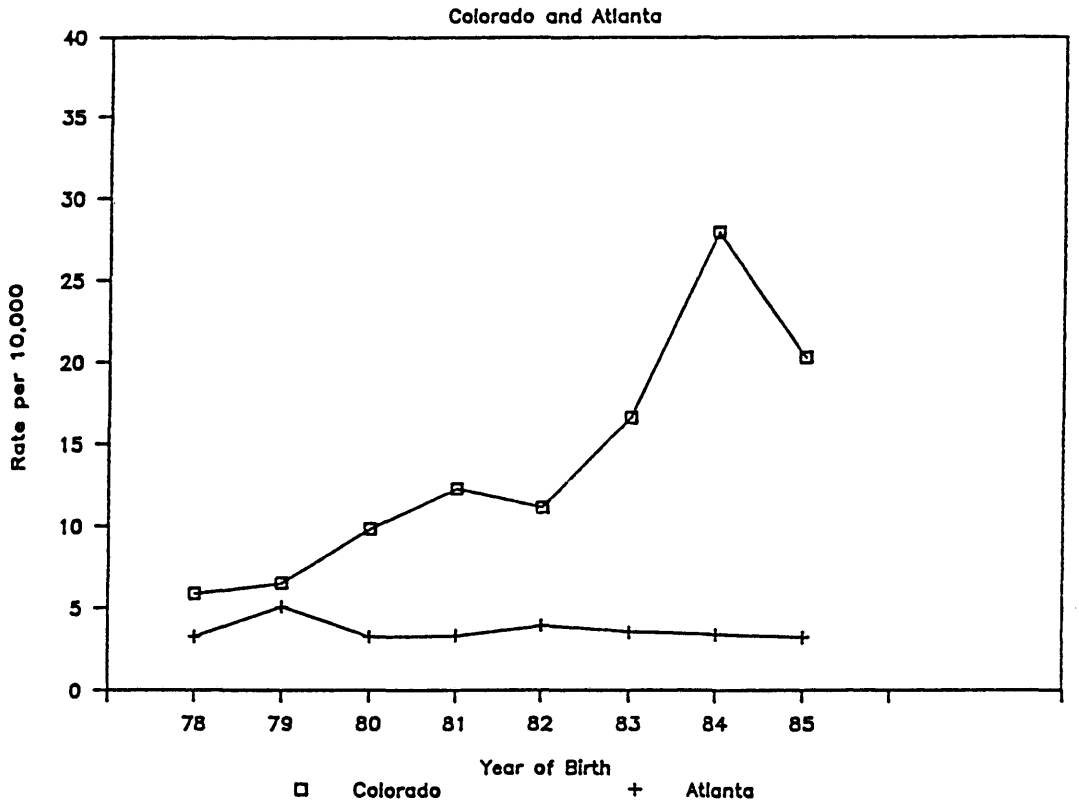


Figure 1. Craniosysnostosis by Year of Birth - Comparison of Rates for Colorado and Atlanta from CDC Correspondence 86-42-2 1988.

and the expense can be an excessive burden to young families. As stated before, the known causes of the birth defect do not account for the rates reported in Colorado (see Figure 2). Thus, the high incidence rates of craniosynostosis in Colorado are a public health concern and interest has developed to conduct research into potential environmental causes using data available for Colorado.

The Colorado Craniosynostosis Project (CCP) was created to manage public health investigations associated with the birth defect. The CCP is a joint program involving the Colorado Department of Health, Office of Health and Environmental Protection, Disease Control and Environmental Epidemiology Division and the University of Colorado Health Sciences Center, Department of Preventative Medicine. As previously stated, the CCP was established in 1985 "to promote the better understanding of the causes, prevention, natural history, and treatment of [the birth defect] craniosynostosis in Colorado" (Alderman, 1986).

There are three objectives described in the workplan of the CCP that contribute to the attainment of the above stated mission. The first is to obtain an accurate count and description of the number of diagnosed cases of craniosynostosis throughout the state since April, 1985. A statewide prospective registry was established to obtain occurrence and outcome data to be analysed and reported on a quarterly basis. The CDC has also supplied the CCP with a

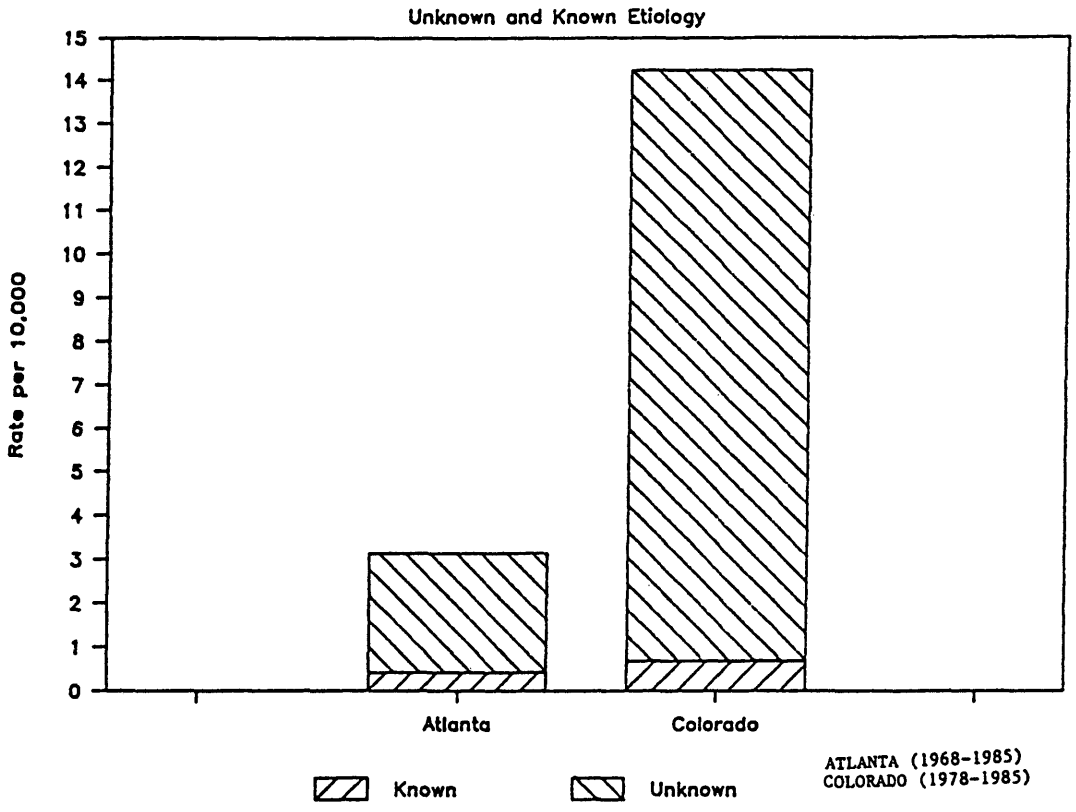


Figure 2. Craniosynostosis in Colorado and Atlanta - Unknown and Known Etiology from CDC Corresspondence EPI-86-42-2 1988.

data set of identified cases and randomly selected controls for the time period from 1978 through 1985 that can be incorporated into epidemiologic studies.

The second objective is to promote the dissemination of information regarding craniosynostosis amoung parents, community members, and health care providers. Newsletters, parent support groups, and public education seminars have been initiated to achieve this goal.

The third objective of the CCP, as stated in the workplan, is to conduct environmental epidemiologic studies of craniosynostosis using retrospective study data received from the CDC in Atlanta, Georgia. These studies are ecologic in design (area based) and are expected to result in the statistical definition of relationships between disease occurrence as derived from the CDC study data and potential environmental exposure values as derived from existing data resources for specified areas in Colorado. These defined relationships are expected to lead to or support the generation of etiological hypotheses to be investigated in future studies.

Ecologic Studies

As with any study design there are limitations and benefits. Limitations exist for the ecologic study primarily as the "ecologic fallacy". This results from drawing conclusions about individuals from information

pertaining to populations (Morgenstern, 1982; Mausner and Bahn, 1985). Thus, any etiologic implications derived from ecological studies must be further tested using exposure studies based on individuals before proof of causality can be presented (Stavraky, 1976; Rothman, 1985; Gesler, 1986; Kelsey et al., 1984).

There are also some limitations in the use of ecologic studies imposed by the uncertainties in characterizing areas by a value for either disease occurrence or potential environmental exposure of the study population. The two most important of these are generally termed underascertainment and misclassification errors. These errors reflect assumptions regarding the completeness of disease reporting or the ability to measure or identify a potential environmental exposure factor among numerous confounding factors (Morgenstern, 1982; Manton et al., 1982; Janerich and Polednak, 1983). For instance, misdiagnoses or misreporting of cases are factors that contribute to underascertainment error. Misclassification error can occur through missassignment of characteristcs to geographic areas.

For ecologic studies there are numerous benefits for their use in development of etiological hypotheses. The first is the potential for integrating information from various existing data sources to assess relationships. This allows for studies to be conducted for the cost of use of

existing data as opposed to spending vast amounts of money designing and carring out expensive sampling procedures.

The well designed ecologic design allows for a focusing approach to be used. Disease occurrence and potential environmental exposures can be determined for populations defined by large areas on a gross scale and systematically focused to smaller areas at a smaller scale (Gesler, 1986). Also, disease occurrence and potential environmental exposure data may be initially assigned from general characteristics and focused on specific factors as deemed appropriate.

Another benefit is that even if no statistically significant relationships are identified in an ecologic study, intermediate products of the study may have utility. For example, the characterization of the disease incidence distribution may be an important result that can be incorporated into other disease incidence studies such as cluster analysis (Grimson et al., 1981; Gesler, 1986) or diagnostic bias studies (PHS CDC Correspondence EPI-86-42-2, 1988). Gesler (1986) identifies the use of expert opinion to evaluate mapped disease distributions. Presentation of the lack of significant results may help to ease tensions of the public until significant results can be generated in different studies. Finally, identification, acquisition and use of environmental data in public health studies can be invaluable for other or future studies of

other diseases. Thus, as long as the limitations are understood, ecologic studies can be used as an effective tool to address public health concerns.

The ecologic studies proposed in the CCP workplan require the creation of area based disease occurrence values and area based potential environmental exposure values relevent to craniosynostosis in Colorado that can be analyzed using accepted epidemiological techniques. These techniques include Spearman rank correlations as used by Berg and Burbank (1972), and bivariate and multivariate regression (Gesler, 1986). Because statistical analysis techniques are defined, the development of procedures to generate values for use in analysis is the specific task that needed to be addressed to conduct ecologic studies within the Colorado Craniosynostosis Project.

Project Development

At the inception of the CCP, only the general goal of conducting ecologic studies of craniosynostosis was defined. As previously stated this goal is to iteratively compare disease occurrence data in its most basic form to potential environmental exposure data derived from existing thematic maps and computerized databases using the ecologic study design. Thus, a pilot study was designed to utilize existing data resources to generate several types of disease occurrence and potential environmental exposure values for

different types of geographic areas throughout Colorado.

Several interrelated study design issues required resolution to design a meaningful ecologic study. The first issue was the identification and selection of environmentally significant geographical areas within Colorado that could be used to define populations for the study. Another issue was the identification of procedures to utilize the locational parameters in the CDC craniosynostosis study data set to generate and maintain disease occurrence values for the selected geographic areas. The final issue was the identification of data and procedures to generate and maintain potential environmental exposure values for the selected geographical areas.

To adequately address these study design issues, available environmental data, geographic techniques, and data processing tools had to be identified and evaluated for applicability in the ecologic study. A systematic survey was conducted to identify environmental data that could be utilized to define geographic areas and derive potential environmental exposure values. A literature search was simultaneously conducted to identify existing studies, geographic techniques, and data processing tools that could be utilized.

To initiate the survey of data resources, a list of all environmental agents and processes that were presented or inferred by informal hypotheses was compiled. Next, a list

of agencies and organizations that could serve as sources of such data was compiled. These lists were used as a starting point to identify types of data that could be useful and where they could be acquired.

Each agency and organization on the list was contacted and inquiries made requesting availability and format information pertaining to any of the listed data. Administrative personnel at the agencies and organizations were contacted first, and then technical personnel for more descriptive information. When possible, documentation and variable listings of specific resources were requested and personal meetings to discuss specific issues conducted.

Simultaneously, a literature review of documented ecologic studies, existing medical geographical methods, and available data processing resources was conducted. After approximately six months, the results of the survey and literature review were evaluated against perceived requirements and a specific pilot study was designed.

Envisioned Pilot Study

The pilot ecologic study was designed to compare disease occurrence values derived from the CDC craniosynostosis study data to potential environmental exposure values derived from the US Bureau of Mines Mineral Availibility System (MAS) database for USGS Hydrologic Units in Colorado. This approach and resource utilization was

chosen to broaden and enhance the approach of Anderson (1986) which addressed the relationship of mining activity to the incidence of craniosynostosis reported in Colorado.

The newly designed pilot study utilizes actual mineral industry locations to define the extent of mining activity. It also incorporates a control group that can be utilized to control for background birth distributions and, potentially, birth rates.

Specifically, for each of three types of hydrologic units, frequency of occurrence for cases, controls, and number of mineral industry locations by commodity produced were to be calculated. These data were intended to be used in accepted ecologic statistical analyses such as previously mentioned and support the the generation of hypotheses as to the cause of craniosynostosis in Colorado.

Data Maintanence and Processing

The first study design issue addressed was that of data processing. Armstrong (1972) describes the benefits of computerization of disease distribution maps and environmental maps for easy maintenance and spatial analysis. Gesler (1986) identifies computerized geographical information systems as a tool providing graphics and data integration capabilities that could support environmental epidemiological studies. Thus, it was determined that these tools should be utilized for the pilot

study if they were available.

A geographic information system (GIS) is a computer system designed to store, manipulate, and display mapped information as well as maintain and access databases of information related to mapped features. The use of GIS allows for the maintanence and manipulation of various types of data sets. Point, linear, and area based geographic features can be stored and referenced in related attribute files (Burrough, 1986).

Because this ecologic study design utilizes location as the common parameter with which to compare disease occurrence and potential environmental exposure, geographical map reading and map comparison techniques were identified for application. Specifically, these techniques include plotting of point locations, overlaying of point location maps with maps depicting area boundaries, and assigning the appropriate area codes to points that reside within the area boundaries. These procedures can be performed manually on hard copy maps, but they can also be accomplished with added data management benefits on GIS if geographic coordinates exist for the mapped features.

Access to ARC/INFO was obtained. This system is a GIS developed by Environmental Systems Research Institute (ESRI) Inc. in Redlands, California (ESRI Inc., 1984). It was identified during evaluation for this study as the industry standard of GIS and capable of performing the required

geographic functions by USGS and US Environmental Protection Agency (EPA). It was made available to the CCP through a usage and support contract with the Cooperative Institute of Research in the Environmental Sciences (CIRES) in Boulder, Colorado.

The ARC/INFO GIS consists of two aspects; a computerized mapping and graphics aspect and a database management aspect. The computerized mapping and graphics aspect is responsible for the graphic generation and spatial transformation of geographic features. The database management aspect maintains two types of data; the geographic feature data utilized by the computerized mapping system and the related attribute data from which a user can make specific queries and qualified requests of the system (Burrough, 1986; ESRI Inc., 1984).

Once coordinates for mapped features are provided to the database management aspect of the system, maps or coverages can be created, maintained, and even automatically analyzed as was done by Openshaw et al. (1987). Manipulation and transformation of the maps and mapped features is accomplished using appropriate functions built into the system. These are especially helpful in comparing or overlaying maps of different features created at different scales or based on different projections. These computerized transformation functions minimize human error in reading and writing of mapped data between maps (Walsh et

al., 1987). Thus, by using ARC/INFO to perform map reading functions, human error could be reduced. This human error would result from misreading or misreporting of codes or locations associated with the value generation process. These human errors may also affect accuracy of classification of points within areas in the ecologic study.

The database management benefits of using a GIS for ecologic studies are associated with the maintanence of created maps and attribute data for further application. Additional attributes can be added easily to the database, additional mapped features can be added for further investigation, and different techniques can be used to generate new values such as distance and direction between points for statistical analysis.

Geographic Boundaries

The study design issue of identifying and choosing geographic boundaries to define populations within Colorado was addressed by reversing the standard ecologic study investigator's approach. Typically, ecologic studies utilize geographical area boundaries that are convenient for the implementation of disease occurrence data as reported by states, counties, municipalities, census tracts, or census blocks (Mausner and Bahn, 1974; Stavraky, 1976; Morgenstern, 1982). The use of these administrative boundaries to define populations dictates that potential environmental exposure

values be obtained for these areas (Stavraky, 1976). This can be accomplished if the potential environmental exposure is known and reported for the same administrative area, such as is done for occupational or behavioral studies performed by Blot (1977) and Robinson (1950).

For the ecologic study of craniosynostosis, the use of administrative area boundaries was thought inappropriate for three reasons. First, the role of the CCP's ecologic studies is to evaluate exposures to unique environmental agents present in Colorado's physical environment such as elements in water, soils, and air. Administrative boundaries are derived from human population characteristics independent of environmental characteristics and often drastically vary in areal extent and ecologic or physical homogeneity. Mausner and Bahn (1974) conclude that "[n]atural boundaries are likely to be more useful than political lines for understanding the etiology of disease".

Secondly, disease occurrence data received from the CDC in Atlanta for use in the CCP ecologic study of craniosynostosis contains records of mother's residence for cases and controls in the form of county-city codes as well as street addresses. Thus, disease incidence and background birth distributions can be plotted in their most basic form as address locations on environmentally based maps. This results in the transformation of the typical ecologic study design from an administrative boundary approach to an

environmental boundary approach which allows for the use of geographic area boundaries that are environmentally significant for the ecologic study.

Finally, the initial results of the environmental data resource survey indicated that potential environmental data existed in computerized forms that would allow for the generation of potential environmental exposures for environmentally significant areas using the ARC/INFO GIS.

The choice to utilize environmentally significant geographic area boundaries was the first of several criteria used to evaluate the results of environmental data survey and literature review to identify usable data resources. Environmentally significant areas were those defined by environmental or physical criteria that were associated with environmental agents or pathways of exposure that could be linked to etiological hypotheses of craniosynostosis. Other defined criteria were based on implementation requirements such as quality, scale, format, and availability. Statewide coverage at consistent scale and quality was required. A digital or computerized format was also required. And finally, the data had to be available to CCP for minimal expense. Table 1 lists the identified area boundary data sets that were evaluated for use in the study.

USGS Hydrologic Units were chosen from this set because of the nested nature of the smaller units within the larger. This allows for the investigation of the larger areas as

Table 1. Identified Geographically-Based Environmental Data Resources: Geographic Boundary Data

DATA RESOURCE NAME	BRIEF DESCRIPTION OF RESOURCE	SOURCE AGENCY OR ORGANIZATION	COMMENTS AND RECOMMENDATIONS
Colorado administrative boundaries	state, county and city boundaries for Colorado	Colorado Dept. of Highways	potentially useful for locational reference on output maps displaying results of ecologic studies
census tracts, blocks, and enumeration districts	boundaries delineated for census purposes	U.S. Bureau of Census	would require digitizing of boundaries to utilize area attribute (population) based statistics not recommended for use
Bailey's Ecoregions	ecoregions defined by ecology and surface morphology	U.S. Fish and Wildlife Service	computerized on national scale good attributes for study potentially useful
EPA Ecoregions	ecoregions defined by ecology and elevation	U.S. EPA	computerized from 7 1/2 minute quads good spatial and attribute resolution recommended for use
USGS Hydrologic Units	surface water drainage basins defined by topography and hierachically coded	USGS Division of Water Resources	digitized from 7 1/2 minute quad sheets good quality and areas related hierarchically - recommended for use
Air Quality Control Regions (AQCR)	regions defined for monitoring and regulating air quality	Colorado Dept. of Health Air Quality Division	county boundaries are used to delineate regions and not computerized, thus not recommended for use
geological surface units	geologic map of surface units for Colorado	USGS Branch of Distribution	potentially useful, but would require digitizing
water producing aquifers	map identifying extent of major water producing aquifers and related water quality data	Colorado Dept. of Health Water Quality Division	potentially useful if source of drinking water can be discerned for the study subjects

well as nested smaller areas with minimal added effort. Also, the statewide digital map of the hydrologic unit boundaries was made available by the USGS Division of Water Resources for no cost.

These hydrologic units are environmentally significant for use in the ecologic study of craniosynostosis. Berg and Burbank (1972) show application of environmentally significant USGS Hydrologic Unit Boundaries in their ecologic study of trace metals and cancer mortality in the United States. Many of the proposed hypotheses presented by the public implicate water quality related mechanisms and environmental agents. These mechanisms and agents are potentially dependent upon the distribution of surface water and groundwater, and therefore potentially dependent on the topography of the land surface.

The USGS Hydrologic Units are defined by topographic relief (Seaber et al., 1987) and represent drainage areas within which surface water influences should be identifiable and distinguishable. Also, the topography could potentially aid in isolating ecological, geological, or climatological influences that could be related to the etiology of craniosynostosis.

Each of the smallest identified units are indexed to an eight digit Hydrologic Unit Code which numerically relates these smaller units to the larger units. The first two digits of the code represent regions, these are the largest

units and consist of drainage basins for major rivers. The next two digits represent subregions, which consist of drainage areas for smaller river systems. The next two digits represent the smaller accounting units. And finally, the last two digits represent cataloging units. These units consist of all or part of a surface drainage basin and are all larger than 700 square miles (Seaber et al., 1987).

In Colorado, there are 4 regions, 17 subregions, 4 accounting units, and 94 cataloging units. Because unit boundaries represent topological divides, there are divides defined between regions, subregions, accounting units, and cataloging units. These divides can be used to deliniate four heirarchically related areas of decreasing extent. Because there are only four accounting units and they are similar in extent to subregions in Colorado, they were considered to be analagous to subregions for purposes of the pilot study. Thus, disease occurrence and potential environmental exposure values were derived for the three heirarchically related hydrologic unit areas defined as regions, subregions and accounting units, and cataloging units.

Disease Occurrence Data

The next presented study design issue, how to utilize locational attributes of the CDC craniosynostosis study data, is partially addressed in the choice to utilize

environmentally significant areas instead of administrative areas to delineate populations for the study. Because of this decision, mother's residence at time of the study subject's birth was utilized as a proxy measure of potential exposure to environmental agents residing within topographically defined hydrologic unit areas.

An identified weakness of using mother's residence location as a proxy measure for potential environmental exposure is that presented by Janerich and Polednak (1983). Namely, exposures of the mother or father to environmental agents that may cause craniosynostosis may occur at any time prior to birth. The residence location utilized for the CCP study was identified as the residence reported on the birth certificate; mother's residence at the time of the identified case's or control's birth. Thus, no information exists regarding prior exposures as indicated by previous residence locations unless an assumption of continuous residency is made as was done by Knox (1963) in his study of cleft lip. This problem may best be overcome by contacting parents of cases and ascertaining exposures or previous residences. Additional expense would be required due to accomedate additional considerations involved in this type of sampling. Thus, because of the need to initiate ecologic studies to address existing public health concerns as soon as possible and at minimal cost, mother's residence is to serve as a proxy measure for exposure with consideration of

the potential weaknesses involved.

The utilized CDC craniosynostosis study data set contains birth certificate records for 553 cases and 800 controls. Appendix B lists the variable fields for this data set. Cases are those reported in Colorado from 1978 through 1985. Controls are randomly selected from statewide records for the same period as the cases. These controls serve to represent a background distribution of births for the study period; where children were born in the state for that period.

Procedures for geographically plotting the mother's residence for the study subjects and counting them within hydrologic units were developed to generate frequency of ocurrence values for cases and controls to serve as disease occurrence values for hydrologic units. Specifically, procedures were designed to minimize potential misclassification of locations within geographic areas. By plotting the mother's residence location on large scale topographic maps as accurately as possible, the assignment of cases and controls to hydrologic unit areas was as exact and reproducible as possible. Thus, future calculations of case and control ratios based on frequency of occurrence for residence locations within these area boundaries were also as reproducible as possible.

A craniosynostosis type parameter was related to the birth certificate records for cases in the CCP disease

incidence data set. This type parameter was derived by L. Freisleben, for a study in which diagnosis of suture involvement was discerned and assigned to each case of the CDC craniosynostosis study data. This suture involvement or craniosynostosis type parameter is important because it delineates between sagittal craniosynostosis and other types. Sagittal craniosynostosis has been shown to be more accurately diagnosed probably because it is more distinct and easier to recognize. This type parameter was related to the cases so that eventually three separate case control ratios could be derived for each geographic area. These types include a sagittal only case control ratio, a nonsagittal only case control ratio, and a total reported case and control ratio. The use of the different types of case control ratios allows for the investigation of disease ascertainment in the ecologic study.

Potential Environmental Exposure Data

The last study design issue, identification and use of environmental data to generate potential environmental exposure values, was addressed by evaluating the data survey and literature review results. As with the geographic boundary data evaluation, criteria were set to choose appropriate data. Although the choice of this element of an ecologic study can be arbitrary, Stavraky (1976) maintains that "observations which emerge from an environmental study

are only as good as the quality and scope of the data" utilized. McGlashan (1972) states that relevency of the choice of this aspect of a study depends largely on luck to prevent the omission of critical factors. Thus, to help minimize the dependence on luck as described by McGlashan, it was determined that environmental agents selected for the study were to be associated with hypothesized pathways of exposure and disease causation. Consistent and complete statewide coverage was also a requirement. The data had to be available in a computerized format containing geographically referenced paramenters. And finally, the data had to be available to the CCP at minimal expense.

The United States Bureau of Mines Mineral Availability System (MAS) computerized database was selected from a list of environmental data resources (see Table 2) that met the previously defined criteria because of its potential use in evaluating the effects of mine operations on the distribution of craniosysnostosis as initially presented by Anderson (1984). Also, it was made available at no cost.

The Minerals Availability System (MAS), is a national database from which the Bureau of Mines can generate various economic analyses and engineering reports pertaining to mineral resource reserves, availability, and use. Within the MAS is the nonconfidential data set (MASNC) which consists of a national inventory of mineral industry

Table 2. Identified Geographically-Based Environmental Data Resources Potential Environmental Exposure Source Data

DATA RESOURCE NAME	DESCRIPTION OF RESOURCE	SOURCE AGENCY OR ORGANIZATION	COMMENTS AND RECOMMENDATIONS
Minerals Availability System (MAS) MASNC tape	Mineral Industry Locations and related site attributes	U.S. Bureau of Mines	locational coordinates for plotting attribute data for each location can potentially determine loadings
Staionary Source Air Pollution Site Inventory	criteria air pollution emissions and loadings by type of pollutant and location	Colorado Dept. of Health Air Quality Division	location coordinates and loadings by criteria pollutant, limited time parameters
CERCLIS Database	site characteristics and locations for CERCLA hazardous waste sites	U.S. EPA	location coordinates registered at county centroids, good attribute data and time parameters
RCRIS database developed from HWDMS Database	site characteristics and locations for RCRA hazardous waste sites	U.S. EPA	under development available April 1989 location coordinates registered at county centroids, attributes good
WATSTORE Database	gauging station flow, location and water quality data	USGS	location and water quality data is variable in regards to completeness would require link with drinking water
STORET Database	water quality and location data for sampling stations	U.S. EPA	same as for WATSTORE
Geological Surface Units	geologic map of surface units for Colorado	USGS Branch of Distribution	potentially useful, but would require digitizing of unit boundaries
Water Producing Aquifers	map identifying extent of major water producing aquifers in Colorado	Colorado Dept. of Health Water Quality Division	Potentially useful if source of drinking water can be discerned for study subjects: digitized statewide coverage

(continued)

Table 2 (continued). Identified Geographically-Based Environmental Data Resources - Potential Environmental Exposure Source Data

DATA RESOURCE NAME	DESCRIPTION OF RESOURCE	SOURCE AGENCY OR ORGANIZATION	COMMENTS AND RECOMMENDATIONS
Bailey's Ecoregions[1]	ecoregions defined by ecology and surface morphology	U.S. Fish and Wildlife Service	digitized at national scale, attributes are well defined
EPA Ecoregions	ecoregions defined by ecology and elevation	U.S. EPA	digitized from 7 1/2 minute quads themes well defined, elevation attribute is incorporated into thematic areas
Nonpoint Source Pollution Assessment Report Data[2]	characterized impacts of nonpoint source pollution for drainages in Colorado	Colorado Dept. of Health Water Quality Division	provides resolution of 3 levels of severity for 5 types of nonpoint source pollution impacts in digital format

1 (Bailey 1980)
2 (WQCD CDH 1988)

locations (MILS) and attributes associated with geological description of the deposit (Berg and Carrillo, 1980; Babitzke et al., 1982). The MILS sites are defined "as metallic or nonmetallic occurrence, prospect, mine (including past producers), geothermal well, or mineral processing plant such as a mill, smelter, or refinery" (Babitzke et al., 1982).

Each mineral industry site has associated data records pertaining to site characteristics . Appendix C lists these site characteristics (attributes) and their format. From this national inventory of mineral industry locations, only those in Colorado were requested. For Colorado, the MASNC database as queried April, 1987 consists of roughly 14,000 mineral industry locations with associated site characteristic records.

The location and type characteristics of mineral industry locations were utilized to derive potential environmental exposure values for use in the pilot ecologic study. These values were derived as frequency counts of mineral industry locations delineated by types of commodities produced within hydrologic unit area. These counts serve as rough indicators of potential exposures to environmental agents associated with environmental impacts of specific types of mining operations within hydrologic unit areas.

The choice to utilize these counts as potential

environmental exposure values was based on exposure pathways identified in standard toxicological texts. Environmental exposures can occur as a result of breathing contaminated air, ingestion of contaminated soil material, ingestion of contaminated water, ingestion of food in which bioaccumulation of contaminants has occurred, or dermal contact with contaminated air, water, and soils. These exposure pathways exist as a result of occupational, residential or recreational behaviors of individuals and are dependent on the presence of environmental agents or processes that contaminate air, soil, or water (Lu, 1985; Klaassen, 1986).

Different environmental agents are made available for potential exposures by different types of mining operations. For instance, coal mining operations have been identified as responsible for loading large amounts of sulfur oxides and metals to the air, water, and soils (Schoenberger, 1980; Guthrie, 1983; Stern et al., 1984; Sanders, 1986). Heavy metals, nonmetals, and radioactive minerals are brought to the surface and may exist in tailings piles, soils, water, and air in localized areas associated with mineral extraction operations (Smith, 1978; Guthrie, 1983; Martin, 1983; Stern et al., 1984). Gravel operations are often responsible for increases in suspended solids in streams (Guthrie, 1983). Thus, the perturbation of the environment due to the mineral industry may present opportunity for the

exposure of humans to potential disease causing agents.

Characterization of the human reproductive toxicological effects of exposures to mineral industry wastes requires toxicological expertise that is rare or nonexistent. Therefore, it is well beyond the scope of this discussion to evaluate the toxicological effects of all the specific agents presented by mineral industry perturbations of the environment. In fact, it is the purpose of ecologic studies to evaluate a broad range of general or unknown effects on populations instead of individuals for which analysis would be too labor, time, and resource intensive. Thus, categories of mining operation type as delineated by commodities present at mineral industry locations were created from the MASNC data.

Categories consist of coal mining operations (both surface and underground), metal mining operations by commodity, nonmetal mining operations by commodity, radioactive element mining operations, gravel and clay operations, and all others. The use of these categories is expected to add some resolution of responsible environmental agents to the ecologic study, without requiring intensive study of each potentially responsible environmental agent.

There are several weaknesses in using these categories and mineral industry locations in the ecologic study. These are based on the lack of resolution of information pertaining to exposure pathways. For example, mineral

industry locations may be remote from populations, thus, minimizing exposures to the general population. Mineral industry locations may be on or near topologic hydrologic unit boundaries and thus, influence two or more areas. Another weakness is that intuitively, water quality is affected downstream of the mineral industry location, and the population may reside upsteam of the mineral industry location. Likewise, water treatment facilitates may remove contaminants that may be involved in an ingestion exposure pathway. Also, environmental impacts may vary widely between mineral industry locations that extract similar commodities due to the use of different mineral extraction techniques. Another weakness is that the number of mineral industry locations gives only a relative estimate of impact on the environment because locations will have various and unknown degrees of environmental impacts.

Finally, there are no dates of operation for the mineral industry locations. The only parameter that addresses time is the operating status (current or past producer). Thus, assumptions must be made that the presence of a mineral industry location regardless of operative status does present environmental impacts. Sites not listed between 1978 and 1985 (the date limits of the disease data) must also be assumed to impact the environment during the time period of the disease data.

Thus, because the use of the potential environmental

exposure values derived from the MASNC data set for Colorado addresses an indicated public health concern in a timely manner and on a general scale, it is therefore appropriate for this ecologic study. as long as the weaknesses pertaining to pathways of exposure are considered in the analysis of results.

The generation of the potential environmental exposure values was accomplished using a methodology similar to that of the disease occurrence value generation. A distribution map of the mineral industry locations was created. This map was overlaid with the hydrologic unit map and hydrologic unit code assigned to each site. The assigned hydrolgic unit codes were related to the maintained MASNC data. Computerized sort and count functions were employed to count the number of mineral industry locations by commodity type for the different hydrologic unit codes.

As with other computerized systems the actual use of the GIS capabilities and specific functions is almost instantaneous. The bulk of time and effort must go into data preparation phases such as structuring, formatting, and quality assurance of the data to be used (Johnson, 1987).

Several procedural steps were devised to format the disease occurrence data for computerized use, assure the quality of the data, and load it into the GIS. Likewise, the mineral industry location data had to be loaded into the GIS. These steps, based on manual and computerized

geographic techniques and tools, are described in detail in the following methods section.

Chapter 3

METHODS

Prepatation of Birth Certificate Data

In order to generate disease occurrence values using ARC/INFO, mother's residence location map coordinates and identifiers had to be derived from the birth certificate records for each study subject.

Birth certificates records for study subjects were received from CDC in the format of a computerized tape. These records were read at the University of Colorado Health Science Center Medical Computing Center (UCHSC-MCC) by writing the tape records to an American Standard Computer Information Interchange (ASCII) format file using SAS software on the VAX 11/780. The ASCII data were transferred from the VAX onto a 5 1/4 inch floppy diskette using KERMIT software. This floppy was copied for backup purposes and utilized to transfer the data to the CCP's personal computer (PC) on which CCP data processing occurs.

Once loaded on the CCP PC, records relevant to the plotting of mother's addresses were selected and put into a plot information file using PCSAS software. These records included study subject identification code, birth certificate year and number, mother's street address, mother's mailing address, and county-city of residence code for each study subject. This plot information file was

sorted by the county-city of residence code to assist in workflow procedures. Unique map identification codes were assigned to the records of each study subject. This sorted and updated plot information file was printed out for use in the plotting procedures.

Procedures and resources for relating residence address data contained in the plot information file to locations on USGS Topographic maps for Universal Transverse Mercator (UTM) coordinate data capture were developed. These procedures and resources are described in Appendix D, which was written to document such procedures in the CCP Protocol Book. The residence locations were plotted using these procedures and guidelines. The result was a series of USGS topographic maps with residence locations represented as plotted points with study subject identification codes assigned to each point (plot maps).

To read the UTM coordinates (northing and easting) for each point, digitizing equipment at the Colorado School of Mines Computing Center (CSMCC) was utilized. For each plot map a computer file was constructed using the DIGITZ program resident on the CSMCC VAX. These files were named according to a convention that incorporates the plot map name and sheet number (first four characters), UTM zone of the source plot map (next two characters), and scale (last two characters). For example, locations digitized from sheet one of Eagle County 1:50,000 scale plot map were

incorporated into the file EAG11350.DAT; those points digitized from the Montrose 1:100,000 plot map were incorporated into the MONT1210.DAT.

Each plot map was registered on the digitizing tablet according to the program requirements and residence location points selected using the digitizing mouse. For each residence location point, X and Y coordinates (easting and northing) values were assigned by the digitizing program and the related study subject identification code assigned by the user. Thus, for each map, a file was created containing X and Y coordinate values and subject identification codes.

Once files were created containing digitized X and Y coordinates and subject identification codes for each plot map, these files were copied to 5 1/4 inch floppy diskettes for storage and transport back to the CCP PC. On the CCP PC, PCSAS was used to add attributes to each file according to the information contained in the file name convention. Thus, updated files contained X and Y coordinates, subject identification codes, plot map scale and UTM zone attributes.

The digitized plot map files were merged by UTM zone. This resulted in the creation of two files containing residence locations for subjects plotted in UTM zone 12 and UTM zone 13. Finally, these two files were written to tape at the UCHSC MCC and transported to CIRES for uploading to the ARC/INFO GIS.

Quality Assurance. The subject identification codes in the digital files were compared to the subject identification codes in the plot information file listing to verify that each location point had been digitized. The location values were also compared to estimates made from the plot maps to identify obvious outlying values that would indicate human error in the digitizing procedure. Any completeness or digitizing errors were immediately rectified by redigitizing the identified problem residence location points while the plot map was still registered on the digitizing tablet.

The digitizing procedure using the equipment at the CSM CC is precise to 0.003 inches or approximately 0.08 millimeters. On 1:50,000 scale plot maps this results in the ability to disciminate between points that are separated by at least 4000 millimeters or 4 meters of true distance. On 1:100,000 scale plot maps this precision is reduced by a factor of two to 8 meters.

By representing the residence location points by a 5/64 inch diameter circle, points separated by 10 meters true distance on 1:50,000 scale plot maps and 20 meters true distance on 1:100,000 scale plot maps distance can be resolved. In either case, the precision involved in the digitizing procedure was appropriate for the measurement of location coordinates.

A quality control step was implemented that tested the

accuracy of the residence plotting procedures; the expected source of inaccuracy. Eighteen randomly selected residence locations from sheet one of the Jefferson County 1:50,000 scale plot map (file name JEFF0150.DAT) were plotted to overlay with the original plot map. The residence locations from this map were selected because this map contains urban, suburban and rural residences and represented the results of the various plotting procedures referenced in Appendix D.

The plotted overlay map was scaled so that three assigned reference points would overlay exactly with those of the plot map. A two millimeter plot symbol was chosen to represent the selected points on the computer-generated plot. The overlay was compared to the plot map on a light table. All the computer-generated points overlaid with corresponding points on the plot map indicating that the digitizing (UTM coordinate data capture) procedures were accurate.

Next, the documented plotting procedures (Appendix D), reference maps, and a new plot map were given to a staff member with little geographic expertise. After, a short orientation, this staff member was asked to plot the eighteen selected residence locations on the new plot map. When finished, the overlay was compared to this new plot map.

Of the eighteen points, fifteen overlayed exactly or overlayed such that the circles representing the locations

were overlapping to some degree. The distances between the overlay locations and the plot map locations for the remaining three points were measured. These distances were one millimeter or less and occurred in different directions. The original plot map locations for these three points were rechecked and the new locations shown to be valid for two of the three. Thus, the quality control measures identified that two of the original eighteen points were initially misplotted and one in the second attempt.

The accuracy involved in plotting the residence locations was approximated as the maximum observed distance of 50 meters. The maximum discrepancy between overlay and plot map locations would result if it occurs in either of the easting or northing directions. Thus, the accuracy to which the points are located is determined by an area 50 meters by 50 meters.

These limits can be doubled for the 1:100,000 scale maps to set the outside limits of the accuracy of the all the coordinate values to a 100 meter by 100 meter area. This variance was determined to be acceptable for the study due to the implementation of map reading procedures to be performed on a statewide scale of 1:500,000.

The merge process was verified by comparing the the original plot information listing to listings of the merged files. Because the plotting and digitizing procedures were judged to be acceptable in the previous quality assurance

step, only the completeness of these files was verified. Figure 3 and Figure 4 illustrate the procedures utilized to prepare and verify the disease incidence data.

Preparation of Hydrologic Unit Boundary Data

The use of the hydrologic unit boundary data required no preparation besides acquiring the data. The data were recieved from the USGS Water Resources Division as an ARC/INFO formatted map written to tape using the EXPORT command resident on the USGS Water Resources Division ARC/INFO system. Thus, the digital map was loaded to the CIRES ARC/INFO system using the IMPORT command to read the tape.

Quality Assurance. The hydrologic unit boundaries for regions, subregions, accounting units and cataloging units were digitized as cataloging units. These cataloging unit boundaries were digitized on the USGS ARC/INFO system as polygons from seven and one half minute quadrangle maps (approximate scale 1:24,000) of Colorado. The digitized boundaries from the quadrangle maps were merged and cleaned to alleviate digitizing errors using ARC/INFO software.

These procedures resulted in the generation of a 1:500,000 scale statewide hydrologic unit map digitized from smaller scale maps that is more accurate than that produced from digitizing a 1:500,000 scale map. Thus, the boundaries

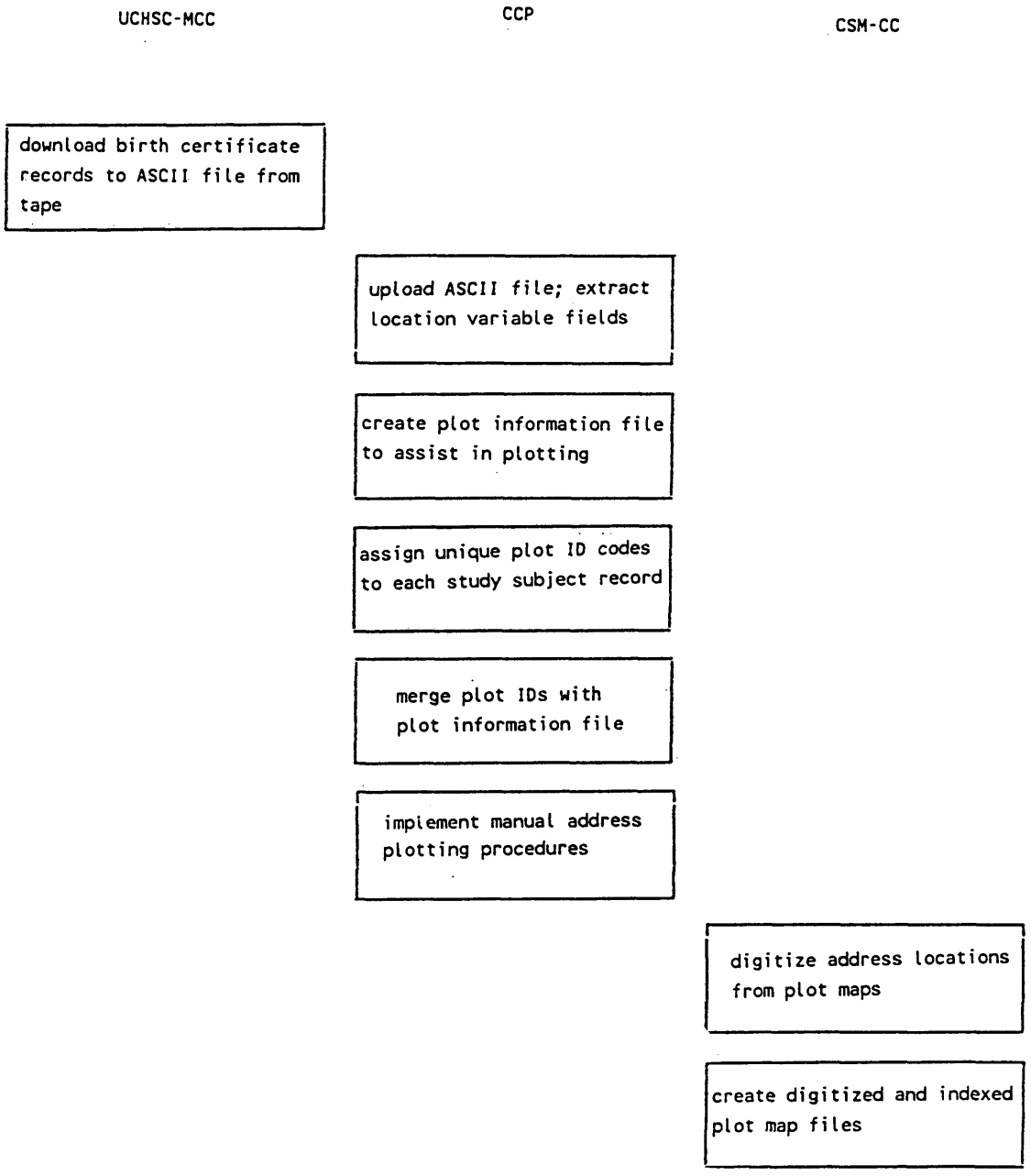


Figure 3. Schematic Diagram of Procedural Steps to Prepare Study Subject Residence Data - Birth Certificate Records to Digitizing of Plot Map Points

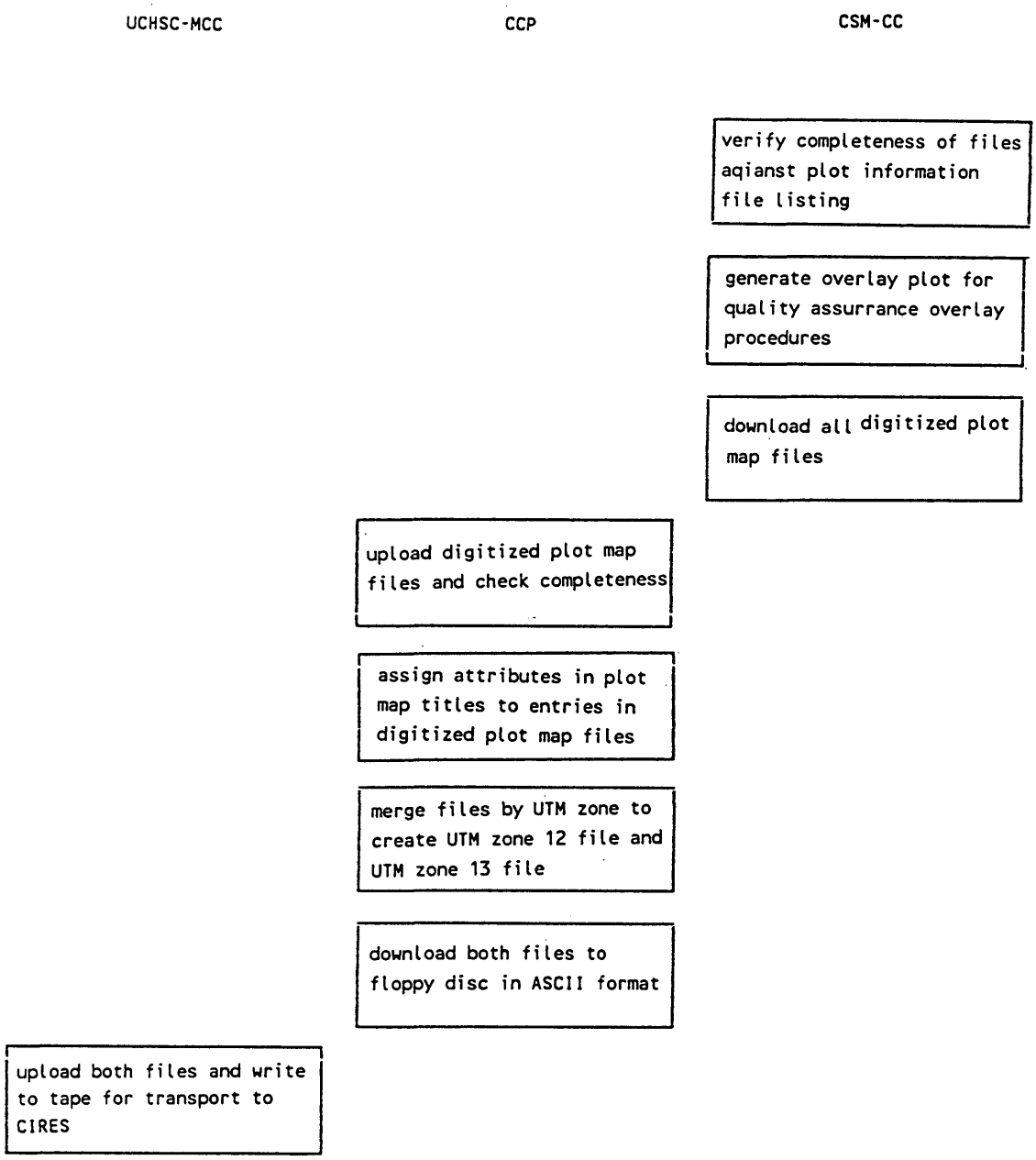


Figure 4. Schematic Diagram of Procedural Steps to Prepare Study Subject Residence Data - Digitizing of Plot Map Points Through the Creation of UTM Zone 12 and 13 Files

are digitized as accurately as possible and appropriate for the ecologic study.

Preparation of the MAS Environmental Data

The data requested from the MAS database was in the format of the MASNC tape containing only mineral industry sites for Colorado identified as of April, 1987. This MASNC data was written to tape in an ASCII format with a known record length of 120 characters. This allowed for FORTRAN programs to be written to extract UTM coordinates and attributes for each mineral industry location in the data set. The tape was copied onto the CIRES GIS allocated disc in the form of a large data file. From this file containing all the variable fields as the tape (see Appendix C) only UTM coordinates, site identification codes, operation name, operation type, status, and commodities present were extracted using FORTRAN programs written by Leeanne Ayers at CIRES (Appendix E). This smaller data file was checked for completeness against the file specification listing provided by Bureau of Mines.

This smaller file was sorted by UTM zone and two files created containing UTM zone 12 coordinates and location attributes and UTM zone 13 coordinates and location attributes using additional FORTRAN programs (see Apendix E). Finally, these files were loaded to ARC/INFO for geographic value assignment. The procedures for preparing

the MASNC data for loading on to the the ARC/INFO system are illustrated in Figure 5.

Quality Assurance. The location coordinates and mineral industry location attributes are verified through the Minerals Availablity Field Office of the US Bureau of Mines. This verification process includes identification of anomalies in the data as identified by mining engineers, metaleurgists, mineral economists, and geologists using existing published and unpublished resources such as that provided by the US Bureau of Mines, the USGS, and Mine Safety and Health Administration (Babitzke et al., 1982).

Location information such as UTM coordinates are derived from published sources and field verification procedures. Each entry is assigned a description of the accuracy and a point of reference at the mineral industry location (Berg and Carrillo, 1980).

Upon an initial review of the data, the UTM coordinate values are present at each of 13727 mineral industry locations. These values reflect measurement to and accuracy to meters (seven digits for eastings and six digits for northing). There are very few missing entries. Because the data are collected to evaluate economic issues, quality is expected to be well maintained. According to Bureau of Mines personnel the database is updated more than once a year. Also, if discrepancies are identified while the CCP

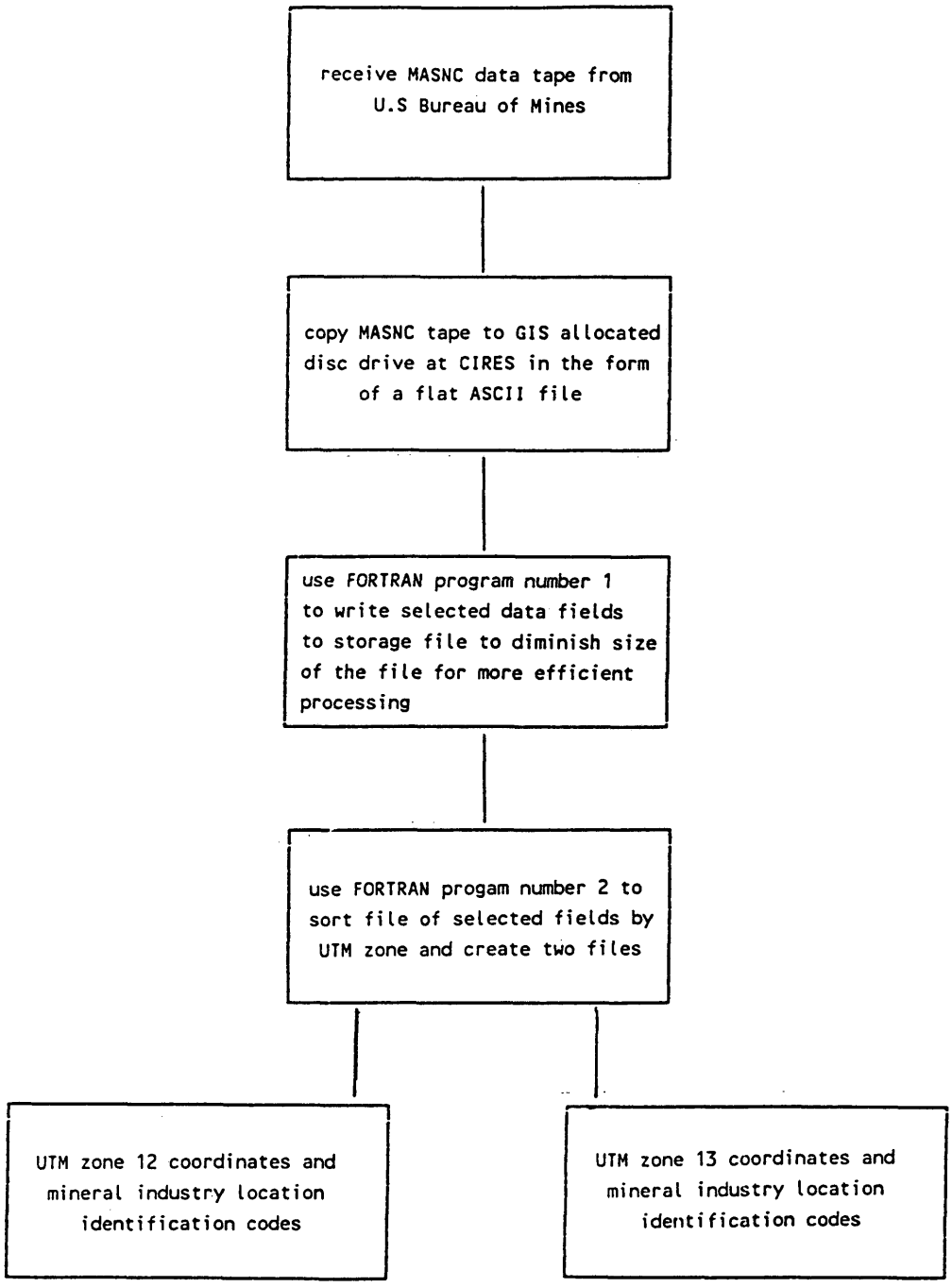


Figure 5. Schematic Diagram of Preparation Procedures for the MASNC Tape Environmental Data

is using the MASNC data, it can be updated by acquiring a listing of data and system updates from the Bureau of Mines.

Assignment of Hydrologic Unit Codes to Point Locations

The same general procedure was used to assign hydrologic unit codes to residence locations and to mineral industry locations. This process involved two steps. The first step was the transformation of UTM zone 12 coordinates contained in the MASNC UTM zone 12 file and the study subject UTM zone 12 file to UTM zone 13 coordinates. The point coordinates contained in these files were transformed into zone 13 coordinates using the PROJECT function of ARC/INFO. Transformed coordinate values and associated point identification codes were merged with the existing UTM zone 13 file for the MASNC and study subject data. Finally, distribution plots of case and control residences and mineral industry locations by commodity types were generated as UTM zone 13 coverages at a scale of 1:500,000 using the GENERATE and BUILD functions of ARC/INFO.

The IDENTITY function was used to overlay the study subject residence and mineral industry location coverages with the hydrologic unit polygon coverage at the 1:500,000 scale. A polygon identification code was automatically assigned to each point (residence or mineral industry site) according to the polygon within which the point was located. Finally, this polygon identification code for each point was

related to the appropriate eight digit cataloging unit code. Thus, hydrologic unit codes were assigned to each point location contained in the study subject digital plot file and mineral industry location digital plot file. Figure 6 illustrates the procedures used to assign hydrologic unit identification codes to each study subject residence and each mineral industry location.

Quality Assurance. The accuracy of this code assignment procedure is dependent on the precision and accuracy of the point locations contained in the different coverages and the dimensions of the boundaries used to delineate areas. The accuracy of the disease occurrence values has been described as limited by the inaccuracy in the address plotting and digitizing procedure. These limits were determined to be roughly 100 meters in the easting or northing directions even though coordinate values reflect a degree of precision measured to meters.

The format of the mineral industry location coordinates implies a one meter precision, but this is in fact variable depending on the souce of the measurements. The identification of the precision and accuracy of location for each of 13727 points would be a very time consuming proposition. Thus, inspection of values and accuracy descriptions suffices to identify a reasonable estimate of accuracy comparable to that of the residence location

coordinates.

ARC/INFO imposes a six digit limit to any coordinates input to the system. This reduces the precision of the northing values (input at seven digits) to tens of meters, but does not affect the easting values. Finally, the dimensions of the boundary lines were ascertained through conversations with ESRI, Inc. personnel to be that of a point digitized. This dictates that the maximum thickness of the boundary line would be 10 meters if the line ran parallel to the easting direction.

Thus, points within 10 meters of boundaries may be misclassified. The number of these points for the disease occurrence data and the MAS data could be quantified using buffer functions of ARC/INFO. This was not done because the size of the hydrologic unit areas is so large compared to the size of any point, the large numbers of points contained in each data set, the precision and accuracy to which the coordinate values were generated and input, and the small thickness of the boundaries at this scale. Thus, the effect of misclassification would be small and does not warrant the additional expense to quantify.

Dataset Creation

The INFO database management system of ARC/INFO was utilized to maintain geographic and attribute data for the study subject data, the MASNC data, and hydrologic units.

INFO files were created for each. For the study subject INFO file, subject identification codes and UTM coordinates were read from the digital files used to generate the point coverage.

Selected records from the MASNC tape were also read into a the MASNC INFO file. These records include identification codes, operation name, operation type, operation status, and commodities mined for each mineral industry location. UTM coordinates were read from the digital files used to create the UTM zone 13 point distribution coverage.

The last INFO file was created for hydrologic units. Hydrologic unit identification codes and automatically generated polygon area values were read from the digital boundary file used to create the boundary plot map.

After hydrologic unit codes were assigned to each point in the digital plot files for study subjects and mineral industry locations, the eight digit hydrologic unit codes for each point were read into the appropriate INFO file.

Using the REPORT function of INFO, disease occurrence values were generated from the disease occurrence INFO file. The frequency of total cases and controls were determined for each eight digit hydrologic unit code representing cataloging units. The same type of REPORT was used to count the frequency of the cases and controls for the two, four and six digit regions, subregions and accounting units

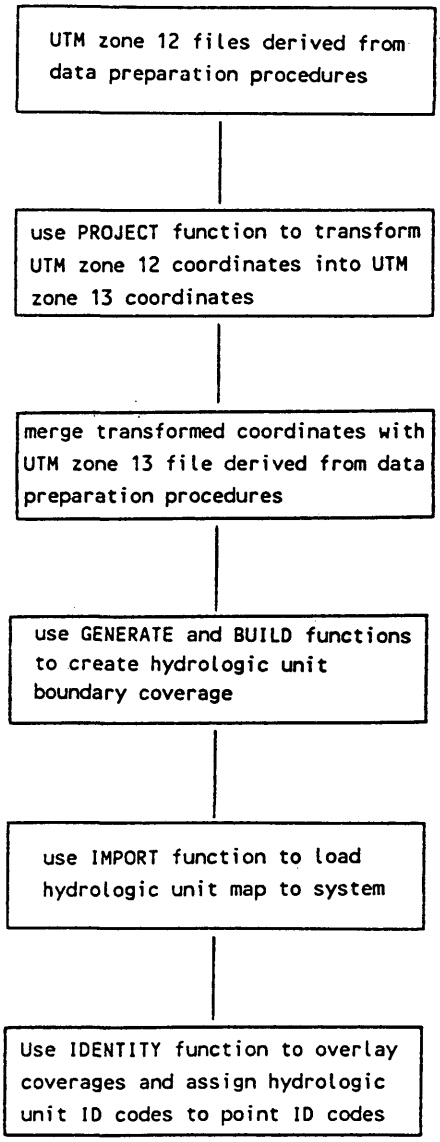


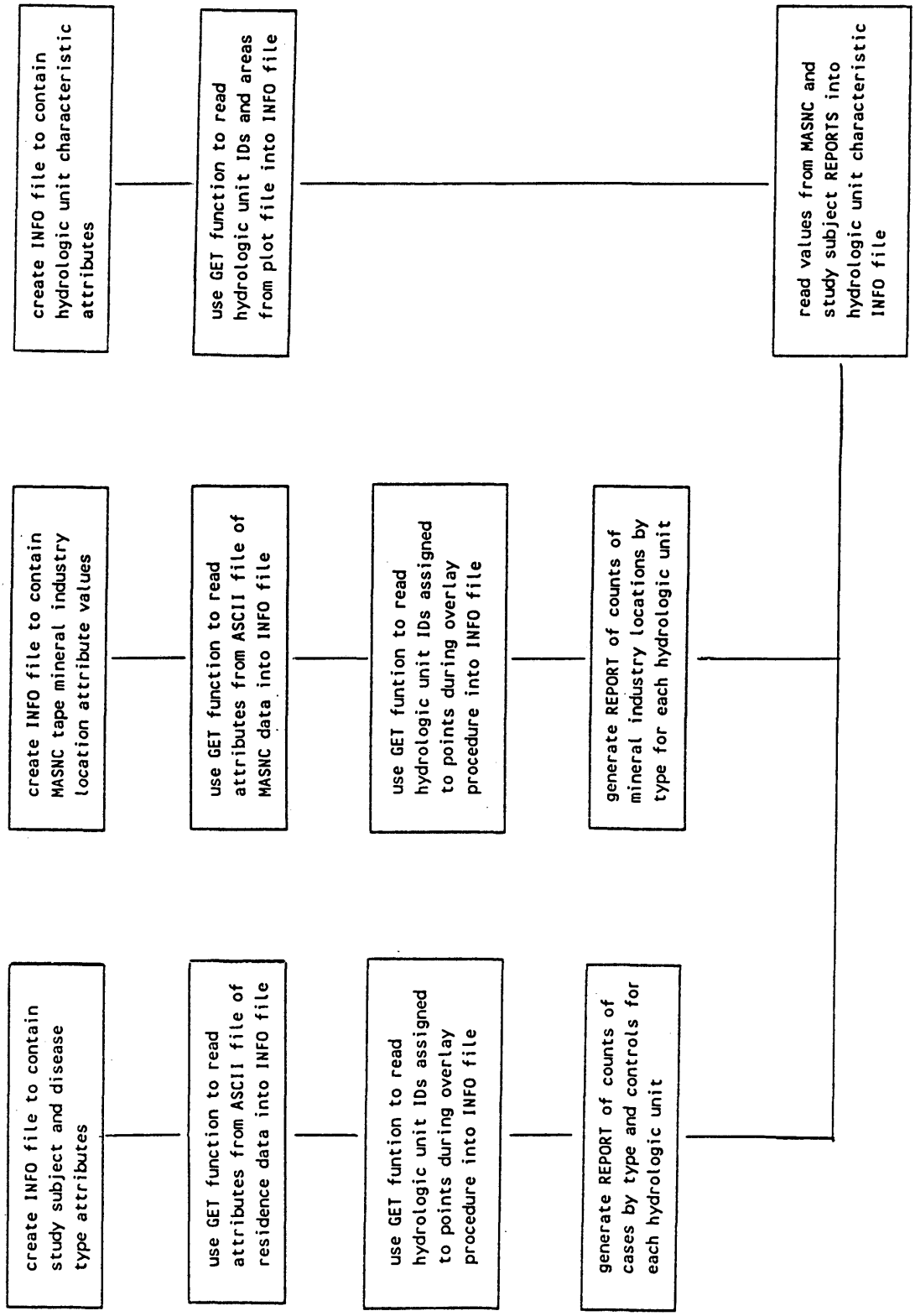
Figure 6. Schematic Diagram of Procedures to Assign Hydrologic Unit Codes to Each Study Subject Residence and Mineral Industry Location

respectively. The frequency counts were then output to a temporary disease occurrence summary file using the REPORT function.

Frequency of occurrence values were similarly generated for mineral industry locations from the MASNC INFO file. These values were delineated further by nine selected commodity types. These nine commodity types were selected because the MASNC data includes roughly 250 specific commodity type categories of which a sample of nine would suffice to initiate ecologic analyses as well as show application of this study without adding to the cost of data processing. The selected categories include sand and gravel, coal, uranium, radium, lithium, lead, copper, cadmium, and oil shale.

Each mineral industry location has up to eight commodity fields, thus an occurrence of a commodity type in any one of these fields constituted an occurrence for the frequency of mineral industry location by type within a specified hydrologic unit. These frequency values were also read into a temporary potential environmental exposure summary file for maintenance and output. Figure 7 illustrates the flow of data used to generate the hydrologic unit summary file.

Finally, the REPORT function was used to consolidate the appropriate information contained in the hydrologic unit, disease occurrence, and potential environmental


create INFO file to contain study subject and disease type attributes
use GET function to read attributes from ASCII file of residence data into INFO file
use GET funtion to read hydrologic unit IDs assigned to points during overlay procedure into INFO file
generate REPORT of counts of cases by type and controls for each hydrologic unit
create INFO file to contain MASNC tape mineral industry location attribute values
use GET function to read attributes from ASCII file of MASNC data into INFO file
use GET funtion to read hydrologic unit IDs assigned to points during overlay procedure into INFO file
generate REPORT of counts of mineral industry locations by type for each hydrologic unit
create INFO file to contain hydrologic unit characteristic attributes
use GET function to read hydrologic unit IDs and areas from plot file into INFO file
read values from MASNC and study subject REPORTS into hydrologic unit characteristic INFO file

exposure INFO files. The resulting file consists of summary information by hydrologic unit regions, subregions and accounting units, and cataloging units delineated by number of digits in the identification code. Each hydrologic unit code is assigned an polygon area value, disease occurrence values of frequencies of cases and controls, and frequencies of relevant commodities mined within each hydrologic unit.

This final form of the data was downloaded in ASCII format to tape for transport to UCHSC where it was downloaded to a floppy diskette and transported to the CCP PC for analyses of the geographically derived values.

Chapter 4

RESULTS AND FINDINGS

Disease Occurrence and Potential Environmental Exposure Dataset

The outcome of the procedures was the creation of disease occurrence and potential environmental exposure values for the three types of hydrologic units described. To summarize these procedures, values were derived from overlaying case, control, and mineral industry location distribution maps (see Figures 8, 9, and 10) with the USGS Hydrologic Unit Boundary map (Figure 11) to produce composite maps (Figure 12, 13, and 14). The IDENTITY function of ARC/INFO was used to automatically assign the appropriate eight digit USGS Hydrologic Unit (cataloging unit) code to each point.

Table 3 is a print out of summary information pertaining to hydrologic units. The values in this table are to be used in the pilot ecologic study as previously described in Chapter 2. The case control counts can be ratioed to produce case control ratios for any of the hydrologic units.

The potential environmental exposure values are frequency counts of occurrence of mineral industry locations by commodity type. This allows for ecologic comparisons to be made between any one type or combination of types of commodities mined at mineral industry locations and the

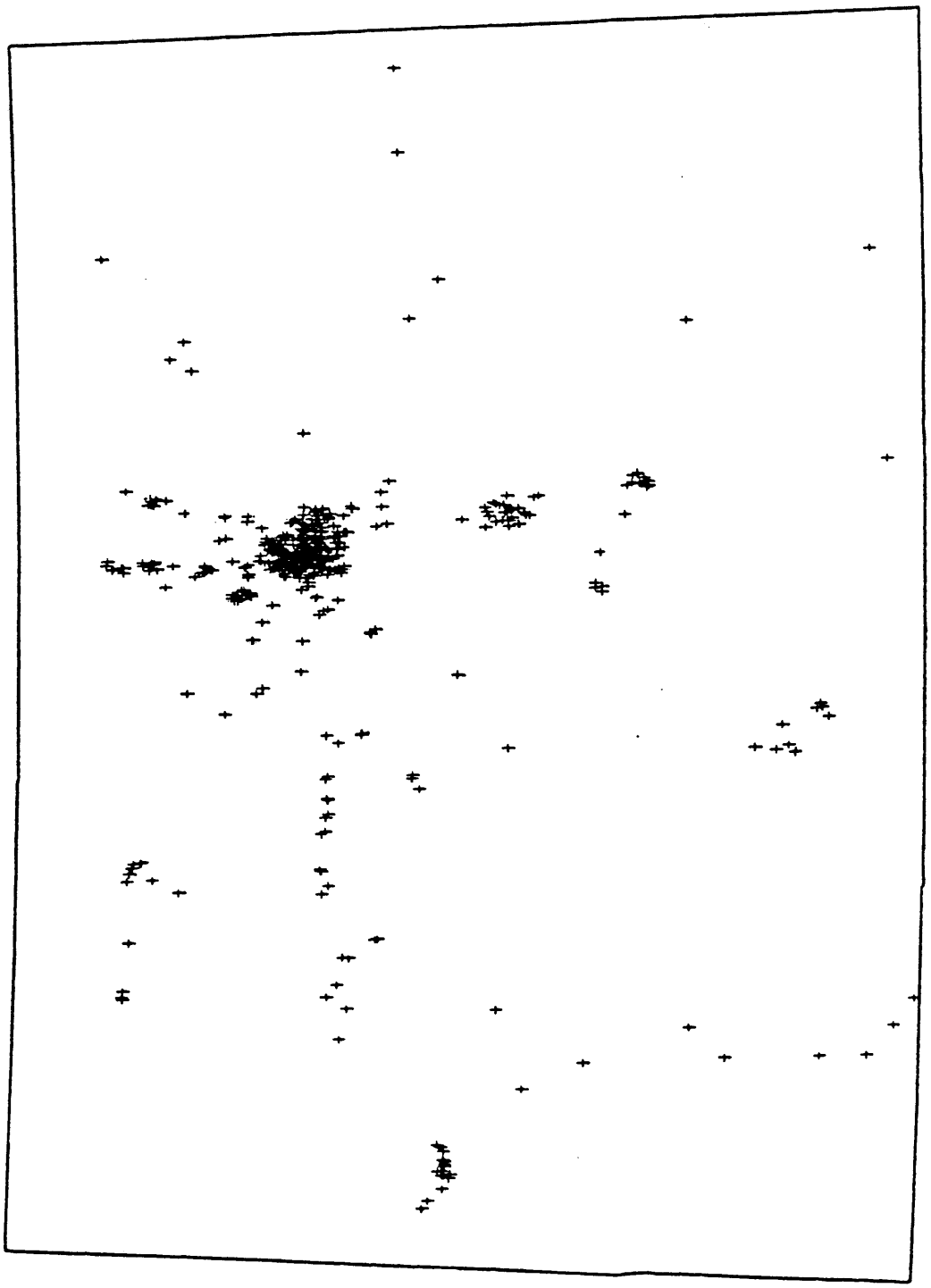


Figure 8. Distribution of Mother's Residence for Cases (+) in Colorado: 1978 - 1985. Reduced from 1:500,000 scale, north at top of page.

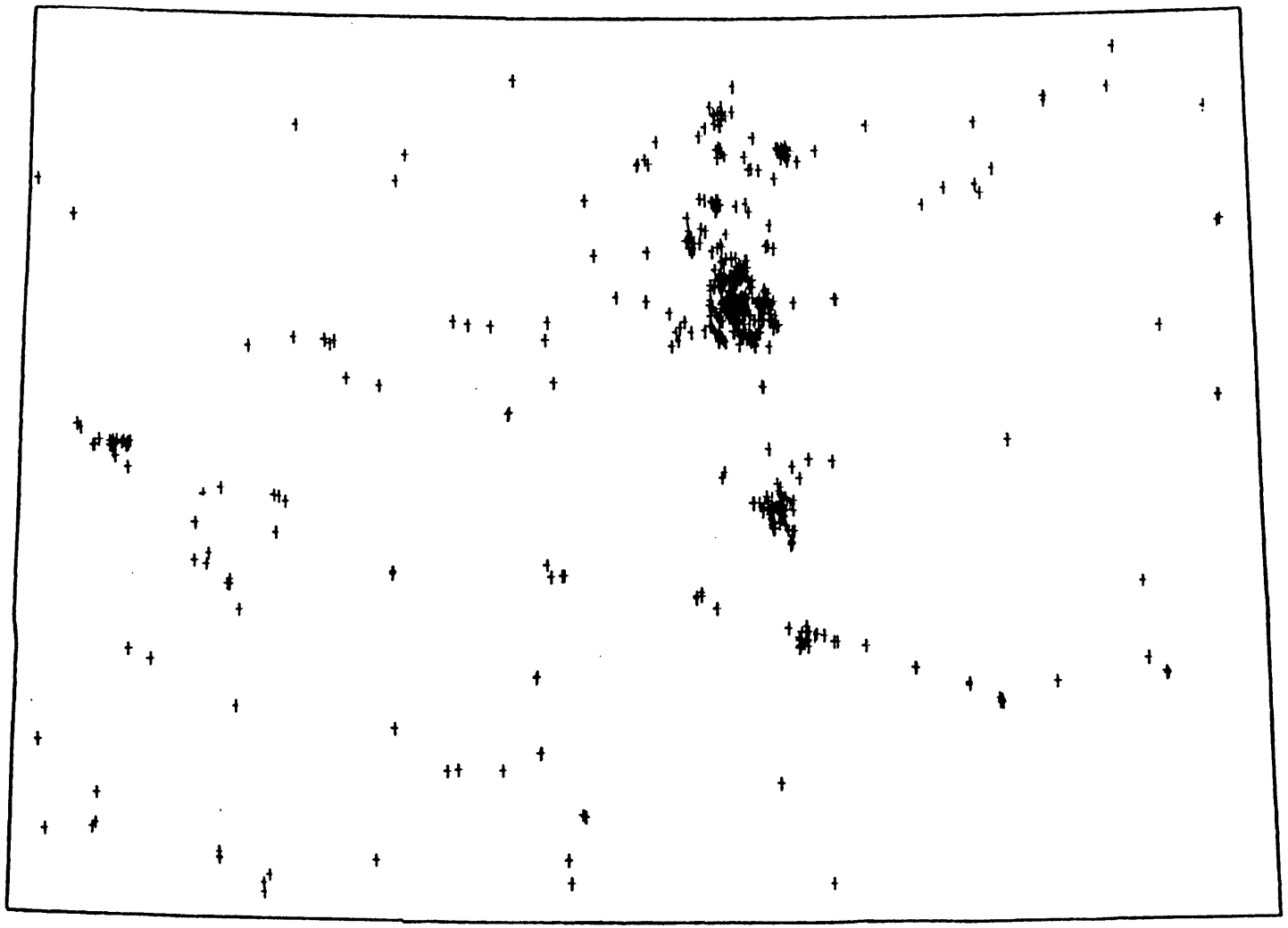


Figure 9. Distribution of Mother's Residence for Controls (+) in Colorado: 1978 - 1985. Reduced from 1:500,000 scale, north at top of page.

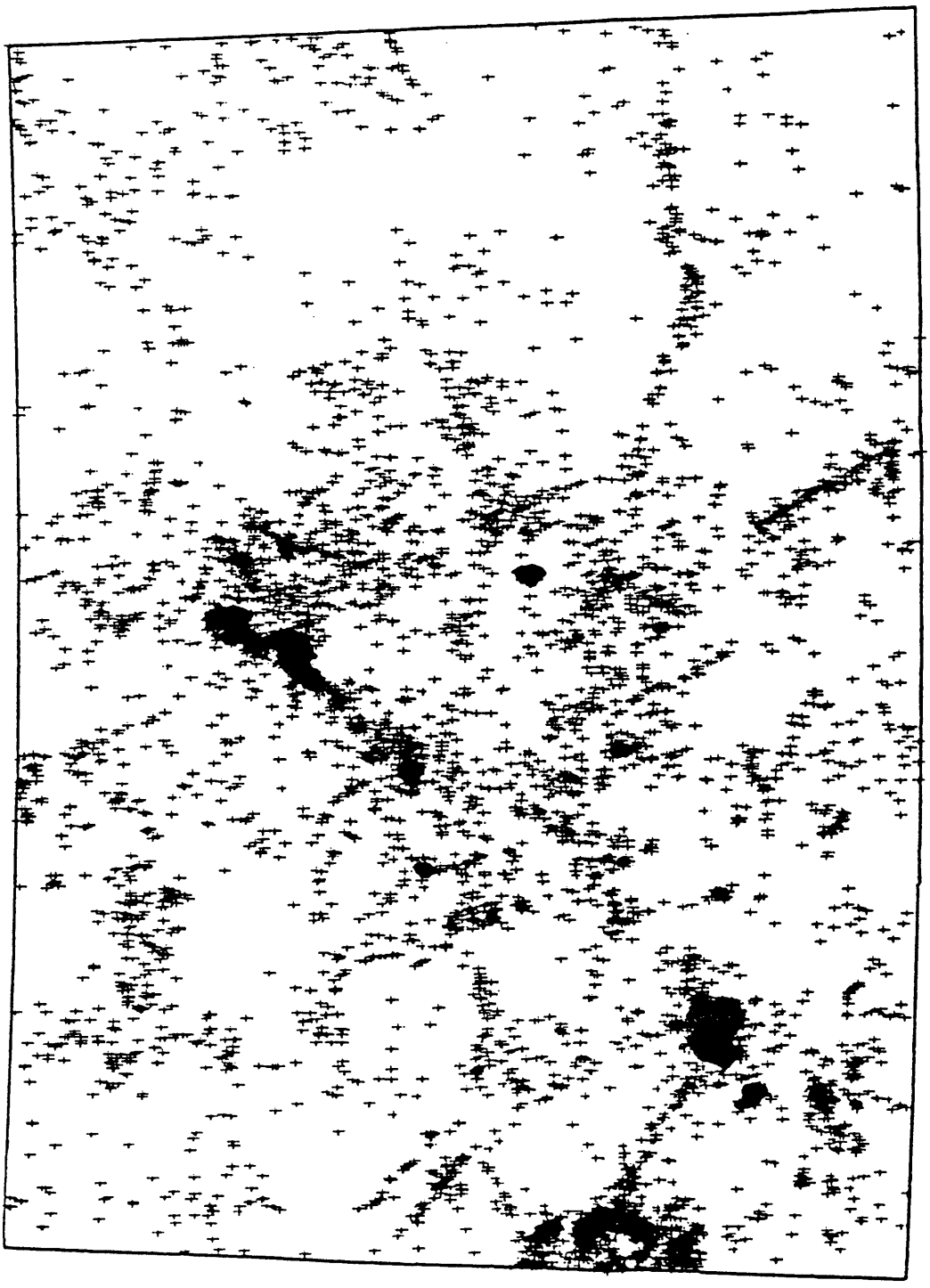


Figure 10. Distribution of All Mineral Industry Locations (+) from MASNC Dataset for Colorado. Reduced from 1:500,000 scale; north at top of page. (apparent shaded areas are areas of high density)

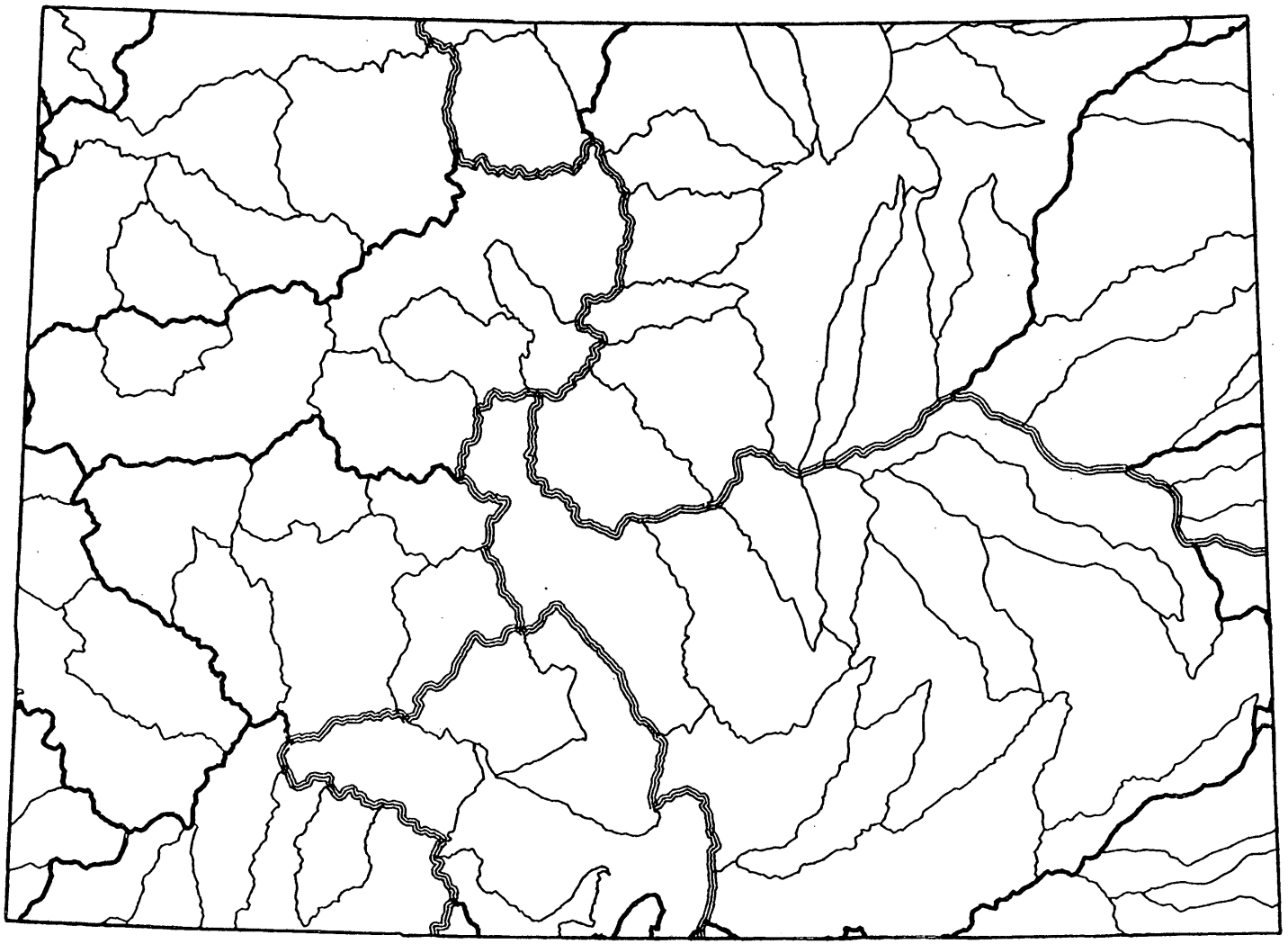


Figure 11. USGS Hydrologic Unit Boundaries for Colorado. Reduced from 1:500,000 scale; north at top of page. (Triple line denotes region boundaries, thick line denotes subregions and accounting unit boundaries and thin line denotes cataloging unit boundaries.)

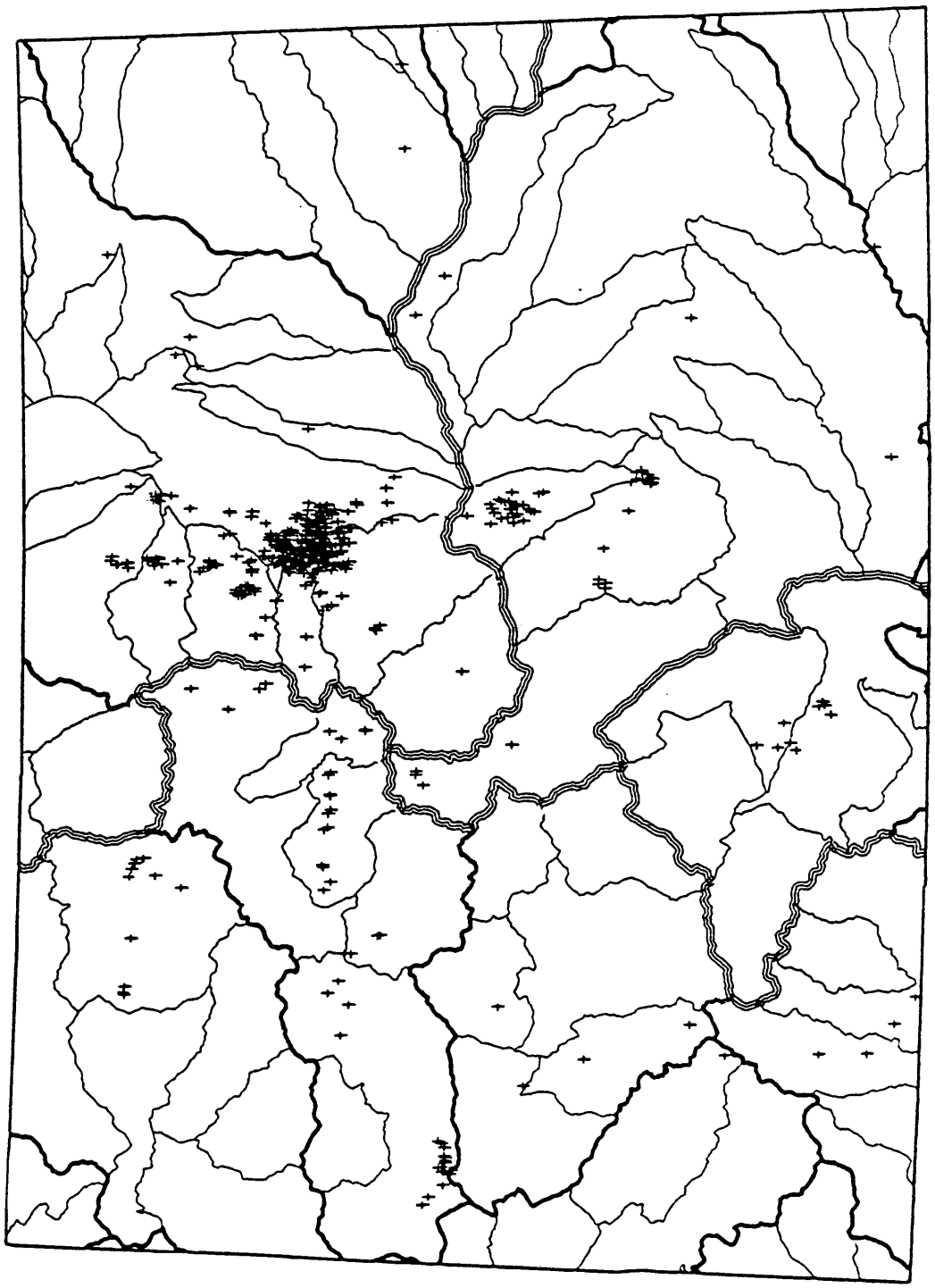


Figure 12. Composite Map Consisting of Mother's Residence for Cases (+) and USGS Hydrologic Units. Reduced from 1:500,000 scale; north at top of page.

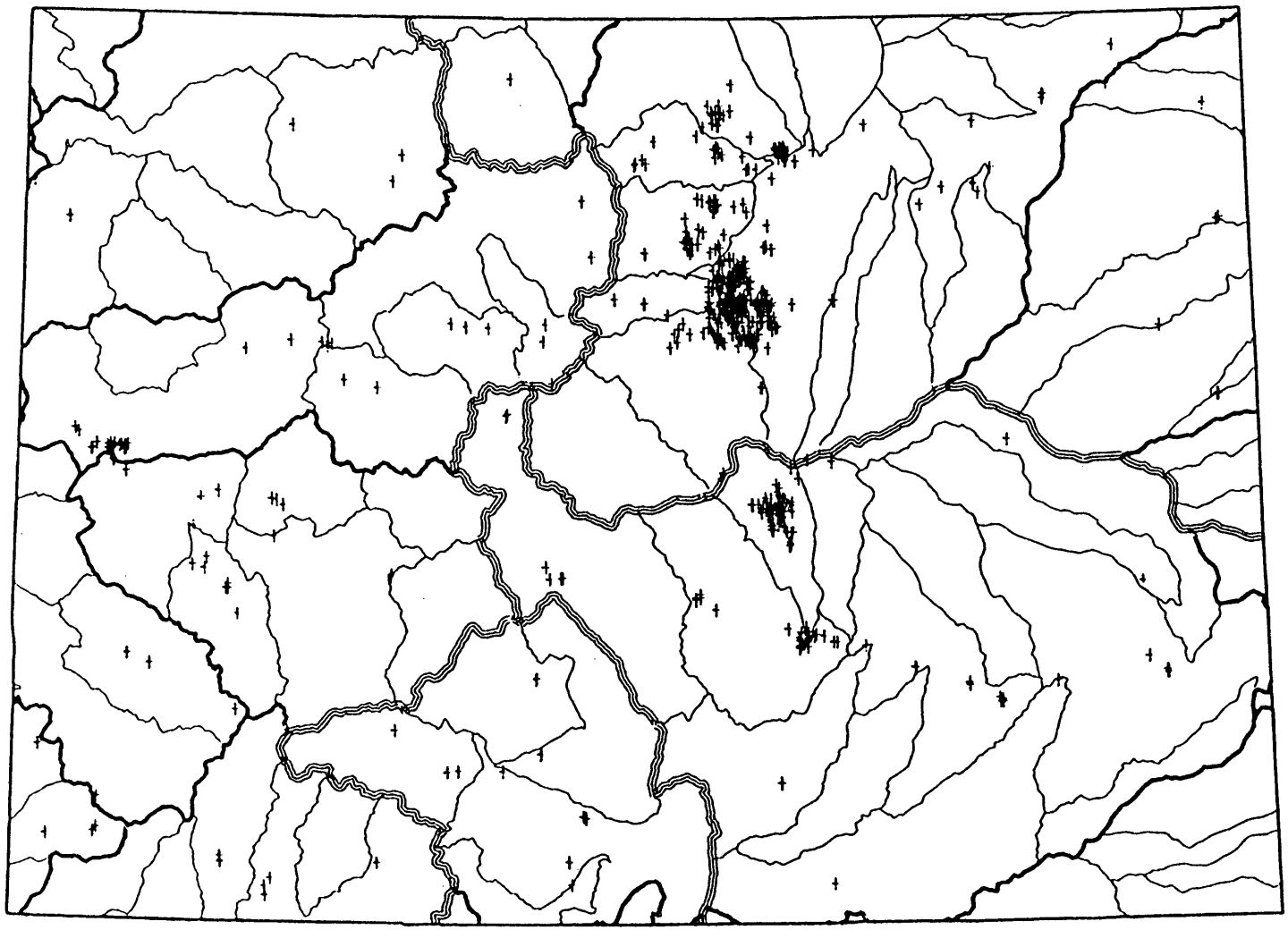


Figure 13. Composite Map Consisting of Mother's Residence for Controls (+) and USGS Hydrologic Units. Reduced from 1:500,000 scale; north at top of page.

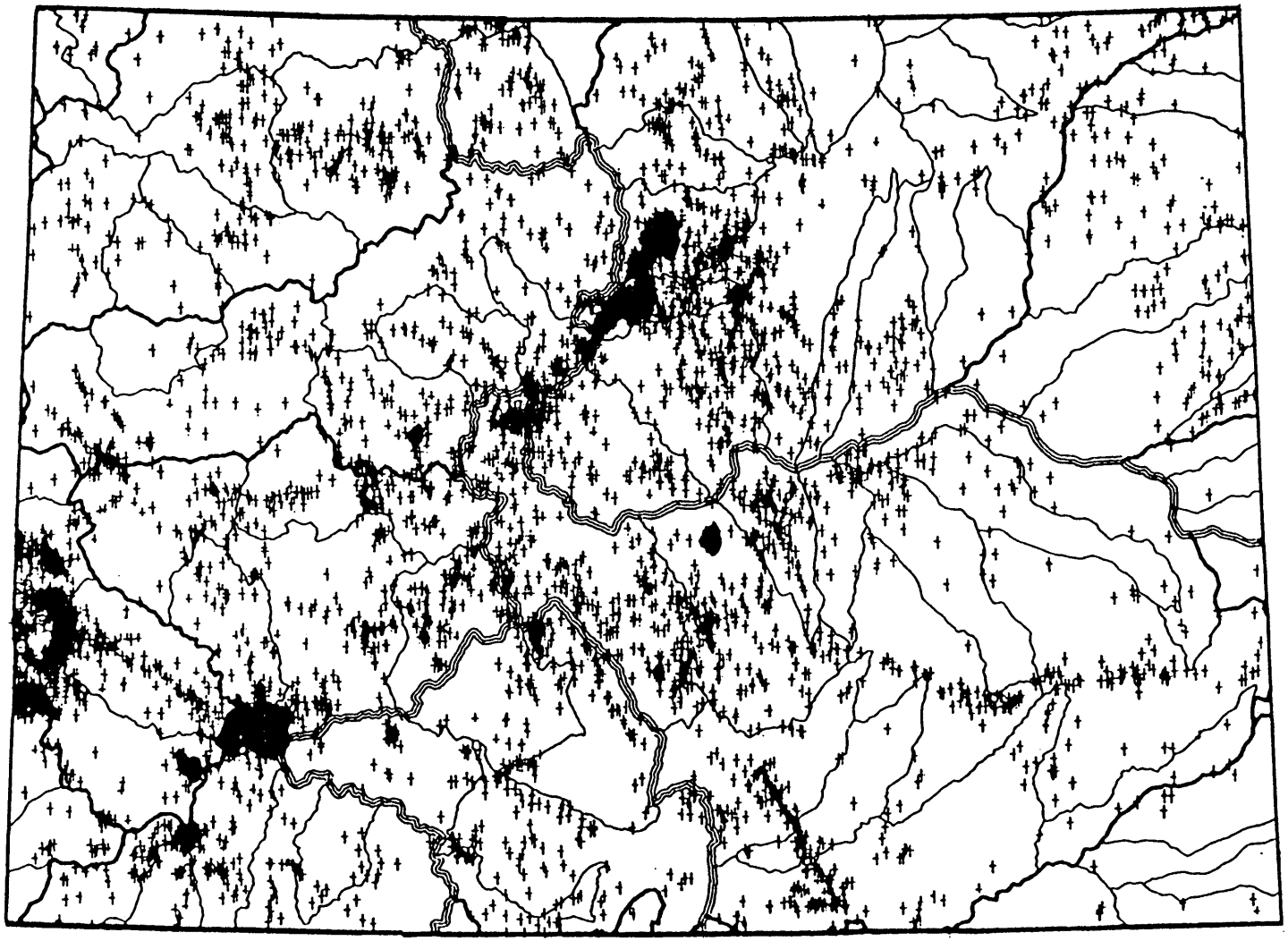


Figure 14. Composite Map Consisting of Mineral Industry Locations and USGS Hydrologic Units. Reduced from 1:500,000 scale; north at top of page. (apparent shaded areas are areas of high density)

Table 3. Summary Report of Values Generated

CAT.UNIT	S&G	COAL	URAN.	RADI.	LITH.	COPP.	LEAD	CADM.	OILS.	CASES	CNTLS	AREA (M*M)
10180001	28	49	2	0	0	9	3	0	0	0	1	3.715120E+09
10180002	3	1	1	0	0	8	2	0	0	0	0	4.856963E+08
10180010	0	0	0	0	0	0	0	0	0	0	0	1.020665E+09
SUBREGION TOTAL	31	50	3	0	0	17	5	0	0	0	1	5.221481E+09
10190001	0	8	110	0	0	11	78	0	0	1	0	4.164976E+09
10190002	42	2	14	2	0	10	15	1	0	113	163	4.797437E+09
10190003	136	28	0	0	0	1	0	0	0	154	185	7.523843E+09
10190004	60	13	83	0	0	257	394	1	1	66	63	1.466300E+09
10190005	42	186	27	0	0	76	97	0	0	44	48	2.562153E+09
10190006	15	1	2	0	0	0	0	0	0	12	18	2.145927E+09
10190007	32	1	4	0	0	4	2	0	0	13	41	4.624964E+09
10190008	23	1	1	0	0	0	0	0	0	1	0	1.122157E+09
10190009	6	0	1	0	0	0	0	0	0	0	0	2.184190E+09
10190010	18	1	0	0	0	0	0	0	0	1	2	1.891003E+09
10190011	51	5	0	0	0	0	0	0	0	2	0	3.581947E+09
10190012	72	0	0	0	0	1	0	0	0	2	9	7.360873E+09
10190013	9	0	0	0	0	0	0	0	0	0	3	2.835308E+09
10190014	7	0	1	0	0	0	0	0	0	0	0	1.908893E+09
10190015	0	0	0	0	0	0	0	0	0	0	0	20765746.000
10190016	3	0	0	0	0	0	0	0	0	0	0	1.688201E+08
10190017	0	0	0	0	0	0	0	0	0	0	0	3.961156E+08
10190018	12	2	0	0	0	0	0	0	0	0	0	2.738088E+08
SUBREGION TOTAL	528	248	243	2	0	360	586	2	1	409	532	4.902948E+10
10250001	25	0	0	0	0	0	0	0	0	0	1	4.351992E+09
10250002	48	0	0	0	0	0	0	0	0	0	5	7.258744E+09
10250003	39	0	0	0	0	0	0	0	0	1	0	5.365380E+09
10250004	0	0	0	0	0	0	0	0	0	0	0	6055378.500
10250005	9	0	0	0	0	0	0	0	0	0	2	1.549460E+09
10250006	6	0	0	0	0	0	0	0	0	0	0	9.752451E+08
10250012	16	0	0	0	0	0	0	0	0	0	2	5.433839E+08
10250013	4	0	0	0	0	0	0	0	0	1	1	2.397469E+08
SUBREGION TOTAL	147	0	0	0	0	0	0	0	0	2	11	2.029001E+10
10260001	1	0	0	0	0	0	0	0	0	0	0	8.704445E+08
10260002	1	0	0	0	0	0	0	0	0	0	0	9.577536E+08
10260004	1	0	0	0	0	0	0	0	0	0	0	6.991710E+08
SUBREGION TOTAL	3	0	0	0	0	0	0	0	0	0	0	2.527369E+09
REGION TOTAL	709	298	246	2	0	377	591	2	1	411	544	7.706834E+10

(continued)

Table 3. (continued) Summary Report of Values Generated

CAT.UNIT	S&G	COAL	URAN.	RADI.	LITH.	COPP.	LEAD	CADM.	OILS.	CASES	CNTLS	AREA (M*M)
11020001	9	2	55	0	0	63	262	0	0	4	6	7.899923E+09
11020002	51	47	22	0	0	2	7	0	0	19	27	5.990332E+09
11020003	108	43	4	0	0	1	2	0	0	34	100	2.392748E+09
11020004	29	1	0	0	0	0	0	0	0	0	1	1.902592E+09
11020005	78	0	0	0	0	0	0	0	0	1	11	5.686022E+09
11020006	39	89	1	0	0	1	1	0	0	0	3	4.794097E+09
11020007	1	50	0	0	0	0	0	0	0	0	0	2.783461E+09
11020008	26	0	0	0	0	0	0	0	0	0	0	3.672296E+09
11020009	74	0	2	0	0	0	0	0	0	0	5	9.841107E+09
11020010	8	116	0	0	0	0	0	0	0	1	2	8.532639E+09
11020011	49	3	1	0	0	0	0	0	0	2	1	4.902953E+09
11020012	21	0	0	0	0	0	0	0	0	0	1	3.547051E+09
11020013	7	0	0	0	0	0	0	0	0	0	0	2.091463E+09
SUBREGION TOTAL	500	351	85	0	0	67	272	0	0	61	157	6.403668E+10
11030001	0	0	0	0	0	0	0	0	0	0	0	96533406.000
11030002	1	0	0	0	0	0	0	0	0	0	0	9.235179E+08
SUBREGION TOTAL	1	0	0	0	0	0	0	0	0	0	0	9.235179E+08
11040001	0	1	0	0	0	4	0	0	0	1	0	1.701602E+09
11040002	3	0	0	0	0	0	0	0	0	0	0	1.281670E+09
11040003	1	0	0	0	0	0	0	0	0	0	0	9.449351E+08
11040004	0	0	0	0	0	3	0	0	0	0	0	1.360989E+09
11040005	4	0	0	0	0	0	0	0	0	0	0	2.537312E+09
SUBREGION TOTAL	8	1	0	0	0	7	0	0	0	1	0	7.826507E+09
11080001	0	0	0	0	0	0	0	0	0	0	0	1.265729E+08
SUBREGION TOTAL	0	0	0	0	0	0	0	0	0	0	0	1.265729E+08
REGION TOTAL	509	352	85	0	0	74	272	0	0	62	157	7.300982E+10
13010001	6	0	0	0	0	30	64	1	0	0	3	3.447607E+09
13010002	53	0	2	0	0	15	6	0	0	7	9	6.521510E+09
13010003	34	0	4	0	0	42	78	0	0	3	2	4.159182E+09
13010004	29	0	1	0	0	3	23	0	0	0	2	3.442967E+09
13010005	12	0	0	0	0	4	3	0	0	0	0	1.384232E+09
SUBREGION TOTAL	134	0	7	0	0	94	174	1	0	10	16	1.895550E+10
13020101	1	0	0	0	0	0	0	0	0	0	0	3.513483E+08
13020102	0	0	0	0	0	0	0	0	0	0	0	2.192706E+08
SUBREGION TOTAL	1	0	0	0	0	0	0	0	0	0	0	5.716189E+08
REGION TOTAL	135	0	7	0	0	94	174	1	0	10	16	1.952612E+10

(continued)

Table 3. (continued) Summary Report of Values Generated

CAT.UNIT	S&G	COAL	URAN.	RADI.	LITH.	COPP.	LEAD	CADM.	OILS.	CASES	CNTLS	AREA (M*M)
14010001	1	8	4	0	0	15	12	0	0	5	4	7.502070E+09
14010002	4	0	2	0	0	228	229	0	0	4	3	1.766051E+09
14010003	1	1	6	0	0	4	9	0	0	14	3	2.528890E+09
14010004	4	22	2	0	0	14	160	0	0	3	2	3.763496E+09
14010005	59	56	30	0	0	2	5	0	11	21	26	8.041909E+09
14010006	3	0	0	0	0	0	0	0	3	0	0	1.827604E+09
SUBREGION TOTAL	72	87	44	0	0	263	415	0	14	47	38	2.543002E+10
14020001	3	6	2	0	0	10	71	0	0	0	0	1.989986E+09
14020002	17	3	8	0	0	45	98	0	0	0	3	6.236740E+09
14020003	15	0	33	0	2	17	55	0	0	0	0	2.866794E+09
14020004	10	24	0	0	0	5	20	0	0	1	3	2.520300E+09
14020005	53	22	10	0	0	0	0	0	0	2	5	4.246456E+09
14020006	41	11	6	0	0	49	202	0	0	2	11	2.930726E+09
SUBREGION TOTAL	139	66	59	0	2	126	446	0	0	5	22	2.079100E+10
14030001	0	0	2	0	0	0	0	0	0	0	0	7.318794E+08
14030002	24	5	1,095	5	0	55	53	0	0	0	1	4.932387E+09
14030003	51	12	391	8	0	69	114	0	0	1	3	4.025530E+09
14030004	2	3	451	0	0	12	0	0	0	0	0	1.645090E+09
14030005	0	0	0	0	0	0	0	0	0	0	0	7248096.000
SUBREGION TOTAL	77	20	1939	13	0	136	167	0	0	1	4	1.134213E+10
14040106	0	0	2	0	0	0	0	0	0	0	0	8.955209E+08
14040109	8	0	5	0	0	0	0	0	0	0	0	1.167831E+09
SUBREGION TOTAL	8	0	7	0	0	0	0	0	0	0	0	2.063352E+09
14050001	54	122	5	0	0	2	3	0	0	13	3	6.780440E+09
14050002	18	10	51	0	0	4	3	0	0	0	0	4.058159E+09
14050003	4	2	19	0	0	0	1	0	0	0	0	4.477538E+09
14050005	16	10	5	0	0	0	1	0	3	0	0	3.563611E+09
14050006	4	4	1	0	0	0	0	0	3	0	0	2.381789E+09
14050007	17	5	3	0	0	1	0	0	1	0	2	3.802540E+09
SUBREGION TOTAL	113	153	84	0	0	7	8	0	7	13	5	2.506408E+10

(continued)

Table 3. (continued) Summary Report of Values Generated

CAT.UNIT	S&G	COAL	URAN.	RADI.	LITH.	COPP.	LEAD	CADM.	OILS.	CASES	CNTRLS	AREA (M*M)
14060001	1	0	0	0	0	0	0	0	0	0	1	1.996235E+08
SUBREGION TOTAL	1	0	0	0	0	0	0	0	0	0	1	1.996235E+08
14080101	15	15	0	0	0	6	4	0	0	2	5	4.225002E+09
14080102	2	6	1	0	0	0	0	0	0	0	0	1.727772E+09
14080104	25	17	13	0	0	191	278	0	0	2	3	2.982447E+09
14080105	4	23	4	0	0	17	18	0	0	0	0	1.134660E+09
14080107	12	27	1	0	0	2	2	0	0	0	0	1.902154E+09
14080201	0	0	0	0	0	0	0	0	0	0	0	4.205174E+08
14080202	5	11	8	0	0	1	0	0	0	0	3	1.652189E+09
14080203	4	0	31	0	0	0	0	0	0	0	2	9.659474E+08
SUBREGION TOTAL	67	99	58	0	0	217	302	0	0	4	13	1.501069E+10
REGION TOTAL	477	425	2191	13	2	749	1338	0	21	70	93	9.990090E+10
STATEWIDE TOTAL	1830	1075	2529	15	2	1294	2375	3	22	553	800	2.695052E+11

disease occurrence values.

Finally, these values have been summarized for each of three heirarchically related hydrologic units (regions, subregions and accounting units, and cataloging units). This allows for the systematic use of each of these different environmentally significant areas in this exploratory application of the ecologic study design.

The format of the disease incidence data and MASNC data used to derive these values is maintained in its most basic form of UTM point location coordinates and associated attributes. This allows for flexibility to iteratively test the disease incidence distribution against the potential environmental exposure values as derived from the MASNC tape data or other types of environmental data. The database management capabilities of ARC/INFO also allow for the addition of more attributes from birth certificate records (date of birth, age of mother, race, disease type, etc...) and loadings from MAS records to be added as deemed necessary. For instance, a disease type indicator could be incorporated into the disease occurrence INFO file. This would allow for resolution between the different types of craniosynostosis and thus, the ability to discern any differences in ascertainment of the disease between areas for the ecologic study.

Furthermore, mineral industry specific attributes from the MASNC tape data can be incorporated into the mineral

industry location INFO file and used to generate different potential environmental exposure values that may resolve some of the limitations imposed by just using frequency counts of commodity types. These additional values could include volumes extracted at mineral industry locations to derive a measure of relative degree of environmental impacts. A concentrator-type attribute could be use to categorize mine sites by particular milling processes types to more specifically categorize specific environmental impacts.

The documentation of the value generation and quality assurance procedures is also a valuable result that can be incorporated into the ecologic analysis of these values. Specifically, these procedures are the first and probably most time consuming step in this study. CCP epidemiologists can utilize this documentation in their analysis of any statistical relationships potentially discovered in the ecological analysis. Finally, the quality assurance steps should be important in evaluating the degree to which the geographical techniques affect misclassification errors of the individual points to an appropriate area.

Thus, this data set was created to meet the needs of the ecologic studies proposed by the CCP. Namely, the disease occurrence data has been transformed into its most basic form to allow for iterative comparison with the numerous potential environmental exposure values derived

from inexpensive environmental data resources for three heirarchically related environmentally significant geographic areas: USGS Hydrologic Units.

Other products important to the study of disease incidence were generated as intermediate products of the value generation process. These include disease distribution maps, documentation of geographic techniques to transform birth certificate location records to UTM coordinates, identification of pertinent environmental data resources for use in ecologic or other environmental epidemiologic studies, and finally, the documentation of the applicability of GIS as a powerful tool in supporting such studies.

Distribution Maps

Disease distribution maps were generated from the transformed birth certificate records as a ARC/INFO coverages of cases and controls (see Figure 8 and Figure 9). From these maps, the CCP plans to elicit expert opinion from environmental scientists and medical specialists that may notice something peculiar in these distributions that may relate to distributions of environmental factors or social factors throughout the state. This process may be helpful in identifying other factors for hypotheses as well as including more people from various backgrounds to assist in the overall mission of understanding craniosynostosis as a

birth defect.

The capabilities of ARC/INFO allow for the creation of maps containing any information contained in the INFO data files. When data such as disease type are related to the disease occurrence INFO file, queried requests can be made to create distribution maps of sagittal or nonsagittal cases. Date of birth can also be incorporated into the disease occurrence INFO file to create case control distributions dependant on specific intervals of time.

Mineral industry location distribution maps can also be generated to define the areal extent of zones in which the mining of particular minerals occurs. This could have application in defining areas of mine impacts due to specific types of operations and be incorporated into future spatial analyses.

Documentation of Plotting Procedures

The documentation of procedures for plotting address derived from birth certificate records is an important result of this study (see Appendix D). The identification of resources that support these procedures is particularly important for the future development of epidemiologic studies in Colorado, in which the residence of the afflicted may be implicated in exposures of the residents to environmental contamination.

Because Colorado is generally a rural state, no single

resource allows for the plotting of addresses on a statewide scale. The development and identification of such procedures may allow for determination of exposures due to the presence of environmental contamination such as hazardous wastes, mine wastes, or water and air quality using spatial analysis techniques available on GIS or other computerized systems.

Identification of Geographically Based Data Resources

The identification and description of geographically based data resources is another important result of this study that can be incorporated into the design of future environmental epidemiologic investigations. By better understanding the resources available, environmental epidemiologic investigations can be better designed for less expense. Also, existing environmental data resources can be adapted to incorporate better locational attributes that can be used for environmental epidemiologic investigations if such applications illustrate the benefits of the additional cost.

Documentation of GIS Application in Ecologic Studies

The last important result of this investigation is the documentation of the application of GIS in ecologic studies. This is important because GIS is a powerful tool that most probably has not realized its potential for application in

numerous fields of study, especially in environmental epidemiology.

The spatial analysis and geographic database management capabilities of such systems are completely appropriate for this type of study as well as many others. GIS has been identified in only a few documented epidemiologic studies and editorials as potentially useful in environmental epidemiologic investigations (Gesler, 1986; Openshaw et al., 1987). It was, however, successfully used in this study to assign area-based codes to point locations, to display geographic distributions, perform sort and count functions, and maintain the data in its most basic form to allow for flexibility in additional or different investigations.

Chapter 5
CONCLUSIONS AND RECOMMENDATIONS

The disease occurrence and potential environmental exposure value generation resulted in the creation of a data set that can be analyzed in an hypothesis generating ecologic investigation of craniosynostosis. Because the intended statistical analysis is beyond the capabilities of this investigator and beyond the scope of this discussion, the final outcome of the use of these values in the ecologic analysis is yet to be determined.

The derivation and quality of these disease occurrence and potential environmental exposure values have been documented for application in the pending ecologic analysis which is based on environmentally significant geographic areas. This provides future investigators with the ability to more effectively evaluate the results of statistical analysis and provide better conclusions. Furthermore, the flexibility of ARC/INFO allows for the addition of more attributes to point locations which may be useful in resolving misclassification and underascertainment errors as well as adding resolution to environmental exposure pathways. Thus, an additional (and time consuming) step towards quantifying the relationship between environmental impacts and disease causation has been made.

Regardless of the outcome of the ecologic analysis,

intermediate products were produced that can support epidemiological investigations of craniosynostosis and other diseases. The first of these products was the documentation of geographic procedures and resources that can be used to plot address information and subsequently capture location data such as UTM coordinates. This allows for the transformation of disease occurrence data typically reported for administrative areas into a format amenable for environmentally based epidemiological analyses.

Secondly, case and control residence distribution maps were generated. These can be compared to distributions of environmental factors as proposed in the pilot ecologic study or they can be used to elicit expert opinion. Either application illustrates to the public that state of the art techniques are being implemented to address possible causes of craniosynostosis.

Associated with the generation of distribution maps, geographically-based data resources have been identified that can be utilized to enhance the proposed ecologic study of craniosynostosis or support the design of new environmental epidemiologic investigations of other diseases.

The last product of this study was the documentation of the application of GIS technology to support this ecologic study. Utilizing a GIS to maintain and perform geographical transformations of geographically-referenced data, value

generation was effectively accomplished for the ecologic study. Additional benefits of utilizing GIS to support ecological studies are provided in the data management and spatial analysis functions included in these systems.

These functions provide the necessary flexibility in adding new data to this type of exploratory study. Additional area based themes such as nested ecoregions or surface geological units can be applied to redefine populations and create new disease occurrence and potential environmental exposure values. Other geographically-referenced environmental data such as the Stationary Source Air Pollution Site Inventory or Nonpoint Source Assesment maps of contaminated surface water can be incorporated into an environmental data set to generate new potential environmental exposures. Lastly, GIS can support application of distance measuring functions potentially applied to identify spatial relationships between mineral industry sites and disease occurrence, if deemed appropriate.

Thus, GIS is an effective tool that can be utilized in exploratory environmental epidemiological investigations. Further research into potential applications of this tool and geographical techniques for studying disease is warranted by present and future public health concerns.

The application of the geographic tools, procedures,

and data resources identified in this study is particularly important in evaluating diseases of known and unknown origin as well as other effects of environmental contamination on our quality of life.

BIBLIOGRAPHY

Alderman, B. A. 1986. Memo re: Craniosynostosis Project Workplan to: B. Arnot, S. Ferguson, J. Emerson, E. Mangione, L. Cook, and A. Lawton; July 1986.

Anderson, M. K. 1984. An Investigation to Identify Potential Environmental Causes of Craniosynostosis. masters thesis. Texas Technological College, Department of Anthropology.

Armstrong, R. W. 1972. Computers and Mapping in Medical Geography. In Medical Geography: Techniques and Field Studies. ed. N. D. McGlashan. London: Methuen and Company, Limited.

Babitzke, H. R., Barsotti, A. F., Coffman, J. S., Thompson, J. G., and Bennett, H. J. 1982. The Bureau of Mines Mineral Availability System: An Update of Information Circular 8654. United States Bureau of Mines Information Circular 8887.

Bailey, R. G. 1980. Description of Ecoregions of the United States. United States Department of Agriculture Miscellaneous Publication no. 1392.

Berg, A. W., and Carrillo, F. V. 1980. MILS: The Mineral Industry Location System of the Federal Bureau of Mines. United States Bureau of Mines Circular 8815.

Berg, J. W., and Burbank, F. 1972. Correlations Between Carcinogenic Trace Metals in Water Supplies and Cancer Mortality. Annals of New York Academy of Sciences. 199: 249-264.

Blot, W. J. and Fraumeni, J. F., Jr. 1977. Geographic Patterns of Oral Cancer in the United States: Etiologic Implications. Journal of Chronic Disease. 30: 745-757.

Borrough, P. A. 1986. Principles of Geographical Information Systems for Land Resources Assessment. Oxford: Oxford University Press.

Craft D. 1987. letter to Dr. B. A. Alderman 6/4/87.

Colorado Water Quality Control Division (CDH WQCD) 1988. Colorado Nonpoint Assessment Report. Prepared to fulfill the requirement of Section 319 of the Clean Water Act.

Coillouette, J. C. M.D. 1987. letter to E. Striegel RNC, 8/3/87.

Environmental Systems Research Institute (ESRI), 1984. Users Guide. Volumes I and II. , Inc. Redlands, California.

Gesler, W. 1986. The Uses of Spatial Analysis in Medical Geography: A Review. Social Sciences in Medicine. 23: 963-973.

Grimson, R. C., Wang, K. C., and Johnson, P. W. C. 1981. Searching for Hierarchical Clusters of Disease: Spatial Patterns of Sudden Infant Death Syndrome. Social Sciences in Medicine. 15d: 287-293.

Guthrie, F. E. 1983. The Energy Crisis. in Introduction to Environmental Toxocology. ed. F. E. Guthrie and J. P. Jerome. New York: Elsevier North Holland, Inc.

Henco Software Inc. 1984. VAX INFO Reference Manual. revision 9, October 1984. Waltham, Maryland.

Janerich, D. T., and Polednak, A. P. 1983. Epidemiology of Birth Defects. Epidemiologic Reviews. 5: 16-37.

Johnson, B. R. 1987. The Design of Computerized Mineral Resource Database Systems, Abstract for COGEODATA meeting presented November 1987.

Kelsey, J. L., Thompson, W. D., and Evans, A. S. 1986. Methods in Observational Epidemiology. New York: Oxford University Press.

Klaassen, C. D. 1986. Principles of Toxicology. in Caserett and Doull's Toxicology. ed C. D. Klaassen, M. O. Amdur, and J. Doull. New York: Macmillan Publishing Company.

Knox, G. 1963. Detection of Low Intensity Epidemicity Applications to Cleft Lip and Palate. Brittish Journal of Preventative Social Medicine. 17: 121-127.

Lu, F. 1985. Basic Toxicology: Fundementals, Target Organs, and Risk Assessment. Washington: Hemisphere Publishing Corporation.

Manton, K. G., Stallard, E., and Riggan, W. 1982. Strategies for Analysing Ecological Health Data: Models of the Biological Risk of Indivduals. Statistics in Medicine. 1: 163-181.

Marshall, W. B. Jr. 1986. Angry? At Who? Letter 8/5/1986.

Martin, D. H. 1983. Nuclear Power. in Introduction to Environmental Toxocology. ed. F. E. Gutherie and J. P. Jerome. New York: Elsevier North Holland, Inc.

Mausner, J. S. and Bahn, A. K. 1974. Epidemiology an Introductory Text. Philadelphia: W. B. Saunders Company.

Mausner, J. S. and Kramer, S. 1985. Mausner and Bahn Epidemiology - An Introductory Text. Philadelphia: W. B. Saunders Company.

McGlashan, N. D. 1972. Medical Geography: An Introduction in Medical Geography. in Medical Geography: Techniques and Field Studies. ed. N. D. McGlashan. London: Methuen and Company, Limited.

Minerals Availability Field Office 1986. Minerals Availability System Nonconfidential (MASNC) User Documentation, November 1986. United States Bureau of Mines.

Morgenstern, H. 1982. Uses of Ecologic Analysis in Epidemiologic Research. American Journal of Public Health. 72: 1336-1344.

Openshaw, S., Charlton, M., Wymer, C., and Craft, A. 1987. A Mark 1 Geographical Analysis Machine for the Automated Analysis of Point Data Sets. International Journal of Geographical Information Systems. 1: 335-358.

Public Health Service (PHS) CDC Atlanta; correspondence coded EPI-83-56-2 dated April 8, 1987. To: Director, CDC, From: Division of Birth Defects and Developmental Disabilities, Center for Environmental Health, Subject: Craniosynostosis-Colorado.

Public Health Service (PHS)-CDC-Atlanta; correspondence coded EPI-86-42-2 dated March 14, 1988. To: Director, CDC, From: Division of Birth Defects and Developmental Disabilities, Center for Environmental Health, Subject: Cluster of Craniosynostosis-Colorado.

Robinson, W. S. 1950. Ecological Correlations and the Behavior of Individuals. American Sociological Review. 15: 351-357.

Rothman, K. J. 1985. Modern Epidemiology. Boston: Little, Brown and Company.

Ruenner, M. 1986. letter to S. Ferguson 5/11/86.

Sanders, C. L. 1986, Toxicological Aspects of Energy Production. New York: Macmillan Publishing Company.

Schoenberger, R. J. 1980. Solid Waste: The Third Pollution in Environmental Health. ed. P. W. Purdom. 2nd ed. Orlando: Acedemic Press, Inc.

Seaber, P. R., Kapinos, F. P., and Knapp, G. L. 1987. Hydrologic Unit Maps. United States Geological Survey Water Supply Paper 2294.

Smith, I. C. 1978. The Geochemical Environment and Its Relationship to Disease. Council of Principal Scientists 1978 Lecture Series. Presented February 17, 1978.

Stavraky, K. M. 1976. The Role of Ecologic Analysis in Studies of the Etiology of Disease: A Discussion With Reference to Large Bowel Cancer. Journal of Chronic Disease. 29: 435-444.

Stern, A. C., Boubel, R. W., Turner, D. B., and Fox, D. L. 1984. Fundementals of Air Pollution. 2nd ed. Orlando: Acedemic Press, Inc.

Walsh, J. S., Lightfoot, D. R., and Butler, D. R. 1987. Recognition and Assessment of Error in Geographic Information Systems. Photgrammetric Engineering and Remote Sensing. 53: 1423-1430.

Appendix A

Facts About Craniosynostosis

COLORADO CRANIOSYNOSTOSIS PROJECT

Colorado Department of Health
4210 East Eleventh Avenue · Denver · Colorado · 80220 · (303) 331-8330

FACTS ABOUT CRANIOSYNOSTOSIS

What is craniosynostosis?

Craniosynostosis is a rare birth defect in which one or more of the sutures in a baby's head are either absent or abnormally developed. This results in the skull bones fusing together before the head growth is completed. Normally, babies have soft sutures (the fibrous tissues between the skull bones) which facilitate the necessary movement of skull bones during delivery and rapid head growth during the first few years of life, gradually turning into bone, by 40-50 years of age. Some children may be diagnosed as having craniosynostosis at birth, or very soon thereafter. Some children do not show signs of this defect until later in infancy or childhood.

What is the result?

The abnormal sutures in the cranium may result in abnormal skull shape and may restrict normal growth and development of the brain due to increased pressure within the skull. Defect severity can vary greatly and usually depends on the number of sutures involved. Craniosynostosis may be associated with multiple birth defects. For most children with craniosynostosis, however, there are no other abnormalities. Some children may have developmental delays associated with their craniosynostosis. It may be important, therefore, for children with craniosynostosis to see experts who can test developmental progress and may suggest intervention if appropriate.

What causes it?

The cause of this birth defect is unknown, nor is it known whether there are single or multiple causes. In a small percentage of cases, craniosynostosis is inherited. (These cases are more frequently associated with other medical problems.) Craniosynostosis is considered to be a symptom of more than 57 different genetic (inherited) illnesses, some of which run in families. Other suspected causes include; lack of adequate oxygen during gestation, fetal head constraint, older maternal age, environmental agents and certain drugs.

How is it diagnosed?

Craniosynostosis should be suspected if the head is not growing at a normal rate or if the head has an abnormal shape such as asymmetry, foreshortening or elongation of the skull or a bony ridging over the suture.

If any suspicion of craniosynostosis arises, the child should be examined by a doctor who will decide if further tests such as skull x-rays or a CT scan of the brain are necessary.

Can it be corrected?

Surgery can improve skull shape and allow room for brain growth by removing abnormal sutures or remodeling around them. The degree to which surgery is successful is often dependent on the severity of the defect, particularly when certain genetic conditions are present. Not all children with craniosynostosis need surgery; certain criteria must be present before surgery is recommended. For some children, an evaluation by a birth defects/ genetics or a child development specialist may help to establish the child's prognosis.

Where does it occur?

Craniosynostosis has been described all over the world, although very little is known about the patterns of occurrence. Because craniosynostosis is a relatively rare birth defect it has been difficult to study on a large scale. Historically, unusual skull shapes indicative of craniosynostosis have been known since ancient times.

How common is it?

The Federal Centers for Disease Control (CDC), which operates a birth defects registry for Atlanta, Georgia, has found 3.4 cases per 10,000 live births. CDC reviewed the 184 Colorado cases which occurred from 1978-1982 and found an annual rate of 7.4 cases per 10,000 live births, more than double the Atlanta rate. In two other studies, however, incidence rates ranged from 6 to 26 cases per 10,000 live births. It is difficult to compare the results of these studies, however, because each used different diagnostic criteria and different ways of ascertaining cases. More study is necessary to resolve these problems and get a better picture of the incidence of craniosynostosis in Colorado.

Where is it most common in Colorado?

Unusual numbers of cases of craniosynostosis have been reported from mountain communities as well as from the Denver Metropolitan area.

The 1978-1982 CDC study reported that six Colorado counties--Routt, Clear Creek, Park, San Miguel, Ouray and Alamosa--had rates of over 30 per 10,000 live births. Three counties - Routt, Clear Creek and Park - had clusters of cases during the 1978-1982 period. Some of these rates are based on small numbers of cases and should be interpreted cautiously. The CDC is currently updating their study from 1983 to 1986.

It appears that geographic trends may be changing. Since January 1986, unusual numbers of cases have been reported to the Health Department from other counties in Colorado (Chaffee, Eagle and Garfield).

What is the Colorado Craniosynostosis Registry?

The Colorado Craniosynostosis Registry includes data on children born in Colorado who have been diagnosed by a physician as having craniosynostosis after April 15, 1986. In addition, the Registry staff is making arrangements to include in the Registry the data collected by the CDC to create a data base from 1978 to the present. The registry has provided physicians with special forms to record basic demographic information, date of birth, county, type of defect, etc.

Parents of registered children will be asked to complete a questionnaire and to authorize release of medical records so that medical information can supplement information given by the parents. This information will be used to investigate the potential risk factors or causes of the defect.

The Registry was established with several purposes in mind. First, it will provide a more accurate statewide count of craniosynostosis cases. Until its establishment, case counts could only be estimated through data provided by specific referral centers. Second, it will allow us to identify a broader spectrum of cases. Previously, we were only able to study those children so severely affected that they required surgery. Since there are many differences in diagnostic criterian and case finding techniques vary from study to study, a uniform system operating within the State would allow us to develop a better picture of the problem in the State and will help identify trends developing over time.

All data will be used to conduct research through the Colorado Craniosynostosis Project's Advisory Council. No names or identifying information will be associated with the computerized data. All information collected in the Registry is kept strictly confidential, and study results will be presented as group statistics.

All health care providers are encouraged to report children newly diagnosed with craniosynostosis to the Registry.

Where can I get more information?

Call or write to: Colorado Department of Health
Disease Control and Epidemiology Division
4210 East 11th Avenue
Denver, Colorado 80220
(303) 331-8330

The Colorado Craniosynostosis Project thanks the following individuals for their contribution to this factsheet: Richard E. Albin, M.D., Robert W. Hendee, Jr., M.D., Pamela McBogg, M.D., Joseph P. Rossi, M.D. and Eva Sujansky, M.D.

Appendix B

Birth Certificate Data Specifications

COLORADO DEPARTMENT OF HEALTH
HEALTH STATISTICS SECTION
BIRTH DATA TAPE FOR CDC/CDH CRANIOSYNOSTOSIS STUDY
CONTROL GROUP

Date: 1-15-87

The enclosed tape has the following characteristics:

Size	1/2 inch magnetic tape
Density	9 track odd parity
	6250 BPI
Labels	non-labeled
Code	~~EBCDIC~~
VOLSER	RD Ø1 Ø6

The tape contains 1 file:

Craniosynostosis study controls from Colorado resident births 1978-85.

LRECL	400
BLKSIZE	6000
# records	800

Please return this tape reel to our library as soon as possible.

Send to: Phoebe Vallejo
Health Statistics Section
Colorado Department of Health
4210 East 11th Ave.
Denver, CO 80220

Record Layout for File # 1: CRANIOSYNOSTOSIS STUDY CONTROLS

Item Name	Type	Position
Certificate Year	2N	1- 2
Certificate Number	6N	3- 8
Blank	60A	9- 68
Blank	30A	69- 98
Blank	30A	99-128
Blank	30A	129-158
Mother's Residence Street Address	30AN	159-188
Mother's Residence Street Zip Code	5N	189-193
Mother's Mailing Address	26AN	194-219
Mother's Mailing City-State	28AN	220-247
Mother's Mailing Zip Code	5N	248-252
Blank	30A	253-282
City of Birth (code)	2N	283-284
Death Certificate Number	11N	285-295
Date of Birth (MMDDYY)	6N	296-301
Sex	1A	302-302
Plurality	1N	303-303
Birth Order	1N	304-304

COLORADO DEPARTMENT OF HEALTH
HEALTH STATISTICS SECTION
BIRTH DATA TAPE FOR CDC/CDH CRANIOSYNOSTOSIS STUDY

Record Layout for File # 1: CRANIOSYNOSTOSIS CASES (continued)

Item Name	Type	Position
Death Certificate Number	11N	285-295
Date of Birth (MMDDYY)	6N	296-301
Sex	1A	302-302
Plurality	1N	303-303
Birth Order	1N	304-304
State of Birth (code)	2N	305-306
County of Birth (code)	2N	307-308
Hospital (code)	3N	309-311
Age of Mother	2N	312-313
Mother's Residence state (code)	2N	314-315
Mother's Residence County/City (code)	4N	316-319
Age of Father	2N	320-321
Attendant Type	2A	322-323
Race of Father (code)	1N	324-324
Education of Father	3AN	325-327
Previous Pregnancies now living	2N	328-329
Previous Pregnancies now dead	2N	330-331
Terminations before 20 weeks	2N	332-333
Terminations 20 weeks and after	2N	334-335
Concurrent Illness	1A	336-336
Race of Mother (code)	1N	337-337
Education of Mother	3AN	338-340
Date of Last Live Birth (MMYY)	4N	341-344
Date of Last Fetal Death (MMYY)	4N	345-348
Date of Last Normal Menses (MMDDYY)	6N	349-354
Month Care Began	1N	355-355
Prenatal Visits	2N	356-357
Legitimacy	1AN	358-358
Birth Weight (lb/oz)	4N	359-362
Complications of Pregnancy (Y/N)	1A	363-363
Congenital Malformations	4N	364-367
Complications of Labor (Y/N)	1A	368-368
APGAR 1 Minute	2N	369-370
APGAR 5 Minutes	2N	371-372
Residence Census Tract	5N	373-377
Race of Child (code)	1N	378-378
Weeks of Pregnancy	2N	379-380
Months since last Delivery	3N	381-383
Type of Delivery	1N	384-384
Spanish Surname (Y/N)	1A	385-385
Low Birth Weight	1A	386-386
CDC Case Flag	3N	387-389
filler	11	390-400

COLORADO DEPARTMENT OF HEALTH
HEALTH STATISTICS SECTION
BIRTH DATA TAPE FOR CDC/CDH CRANIOSYNOSTOSIS STUDY

Date: _____

The enclosed tape has the following characteristics:

Size	1/2 inch magnetic tape
Density	9 track odd parity
	6250 BPI
Labels	non-labeled
Code	EBCDIC
VOLSER	RD0105

The tape contains 3 files:

#1. Craniosynostosis cases. (original file - pre 1986)

LRECL	400
BLKSIZE	6000
# records	552

#2. Craniosynostosis cases. (1986 cases)

LRECL	400
BLKSIZE	6000
# records	261

#3. Craniosynostosis cases. (1987 cases)

LRECL	400
BLKSIZE	6000
# records	63

Please return this tape reel to our library as soon as possible.

Send to: Phoebe Vallejo
Health Statistics Section
Colorado Department of Health
4210 East 11th Ave.
Denver, CO 80220

Record Layout for Files # 1 - 3 : CRANIOSYNOSTOSIS CASES

Item Name	Type	Position
Certificate Year	2N	1- 2
Certificate Number	6N	3- 8
Child's Name and Title	60A	9- 68
AKA 1	30A	69- 98
AKA 2	30A	99-128
Maiden Name of Mother	30A	129-158
Mother's Residence Street Address	30AN	159-188
Mother's Residence Street Zip Code	5N	189-193
Mother's Mailing Address	26AN	194-219
Mother's Mailing City-State	28AN	220-247
Mother's Mailing Zip Code	5N	248-252
Name of Father	30A	253-282
City of Birth (code)	2N	283-284

COLORADO DEPARTMENT OF HEALTH
HEALTH STATISTICS SECTION
BIRTH DATA TAPE FOR CDC/CDH CRANIOSYNOSTOSIS STUDY
CONTROL GROUP

Record Layout for file: CRANIOSYNOSTOSIS STUDY CONTROLS (continued)

Item Name	Type	Position
State of Birth (code)	2N	305-306
County of Birth (code)	2N	307-308
Hospital (code)	3N	309-311
Age of Mother	2N	312-313
Mother's Residence state (code)	2N	314-315
Mother's Residence County/City (code)	4N	316-319
Age of Father	2N	320-321
Attendant Type	2A	322-323
Race of Father (code)	1N	324-324
Education of Father	3AN	325-327
Previous Pregnancies now living	2N	328-329
Previous Pregnancies now dead	2N	330-331
Terminations before 20 weeks	2N	332-333
Terminations 20 weeks and after	2N	334-335
Concurrent Illness	1A	336-336
Race of Mother (code)	1N	337-337
Education of Mother	3AN	338-340
Date of Last Live Birth (MMYY)	4N	341-344
Date of Last Fetal Death (MMYY)	4N	345-348
Date of Last Normal Menses (MMDDYY)	6N	349-354
Month Care Began	1N	355-355
Prenatal Visits	2N	356-357
Legitimacy	1AN	358-358
Birth Weight (lb/oz)	4N	359-362
Complications of Pregnancy (Y/N)	1A	363-363
Congenital Malformations	4N	364-367
Complications of Labor (Y/N)	1A	368-368
APGAR 1 Minute	2N	369-370
APGAR 5 Minutes	2N	371-372
Residence Census Tract	5N	373-377
Race of Child (code)	1N	378-378
Weeks of Pregnancy	2N	379-380
Months since Last Delivery	3N	381-383
Type of Delivery	1N	384-384
Spanish Surname (Y/N)	1A	385-385
Low Birth Weight	1A	386-386
CDC Case Flag	3N	387-389
filler	11	390-400

Appendix C

MASNC Tape Data Specifications

1 INTRODUCTION

THE MINERALS AVAILABILITY SYSTEM NONCONFIDENTIAL (MASNC) INFORMATION, AN INTEGRAL PART OF THE MINERALS AVAILABILITY SYSTEM (MAS) DATA BASE, IS USED FOR THE SYSTEMATIC STORAGE AND RAPID RETRIEVAL OF MINERAL DEPOSIT LOCATION, IDENTIFICATION, AND DEFINITION INFORMATION. MAS IS A U.S. BUREAU OF MINES PROGRAM DESIGNED TO SYSTEMATICALLY MEASURE AND CLASSIFY MINERAL RESOURCES. THE MINERAL INDUSTRY LOCATION SUBSYSTEM (MILS), WHICH IS PART OF MASNC, CURRENTLY CONTAINS DATA ON MOST OF THE SIGNIFICANT MINERAL PROPERTIES THROUGHOUT THE WORLD.

CORRECTIONS, ADDITIONS, AND COMMENTS FOR IMPROVING THIS PRODUCT MAY BE ADDRESSED TO:

CHIEF, MINERALS AVAILABILITY FIELD OFFICE
U.S. BUREAU OF MINES
BLDG. 20, DENVER FEDERAL CENTER
DENVER, COLORADO 80225

1.1 HISTORICAL DEVELOPMENT

THE FOUNDATION FOR COLLECTING AND ASSEMBLING MINERAL DEPOSIT INFORMATION, AND THE STANDARDS FOR CLASSIFYING AND STORING THIS INFORMATION, WERE ESTABLISHED AND DESCRIBED IN THE BUREAU OF MINES INFORMATION CIRCULAR 8654, "THE BUREAU OF MINES MINERALS AVAILABILITY SYSTEM AND RESOURCE CLASSIFICATION MANUAL" (1974). STANDARDS FOR THE MINERAL INDUSTRY LOCATION SUBSYSTEM WERE ESTABLISHED AND DESCRIBED IN 1974.

2. MASNC TAPE DOCUMENTATION

DOCUMENTATION FOR THE MASNC TAPE CONSISTS OF:

A. MASNC TAPE FILE CHARACTERISTICS (PARAGRAPH 2.1)
B. RECORD DESCRIPTIONS (PARAGRAPH 2.2)
C. DATA ELEMENT DESCRIPTIONS (PARAGRAPH 2.3)

2.1 MASNC TAPE FILE CHARACTERISTICS

2.1.1 RECORDING SYSTEM CHARACTERISTICS

A. CHARACTER SET IS EBCDIC
B. PARITY IS ODD
C. EQUIPMENT IS BURROUGHS B-6800
D. DENSITY IS 1600 BPI
E. NUMBER OF TRACKS IS 9
F. UNLABELED

2.1.2 FILE CHARACTERISTICS

A. RECORD LENGTH IS 120 CHARACTERS
B. NUMBER OF RECORDS PER BLOCK IS 180 (21,600 CHARACTERS)
C. RECORD TYPE IS FIXED LENGTH
D. FILE IS MULTI-REEL

2.1.3 FILE SEQUENCE

FILE SEQUENCE IS DETERMINED BY THE DATA ELEMENTS SEQUENCE-NUMBER, RECORD-NAME, AND RECORD-IDENTIFIER. ALL RECORDS COMPRISING A PROPERTY HAVE BEEN ASSIGNED THE SAME SEQUENCE-NUMBER, AND THE MASNC TAPE IS IN ASCENDING SEQUENCE-NUMBER ORDER, POSITIONS 1 THROUGH 10.

ALL RECORDS DEFINING A PROPERTY ARE ORDERED BY THE DATA ELEMENT RECORD-NAME (POSITIONS 11 THROUGH 15) AS FOLLOWS: STATE, IDENT, LOCAT, UTM, TOPOG, BASIN, HOL, REFER, PLANT, PLS, COMMO, N-NAM, O-NAM, OWNER, B-BIB, G-TOB, G-ORI, G-SOB, G-ORE, GEO-1, GEO-2, HISTO, HPDES, HPHIS, HXPLO, LITH1, LITH2, LROCK, MIN-1, MIN-2, M-GEO, R-MEA, R-IND, R-INF, R-UND, RESER, R-REM, AND RASAY.

THE RECORD-IDENTIFIER DATA ELEMENT (POSITIONS 16 THROUGH 19) PROVIDES AN ORDERING FOR THOSE RECORDS WHICH MAY OCCUR MORE THAN ONCE PER PROPERTY. THOSE RECORDS ARE IDENTIFIED BY THE RECORD-NAMES OF COMMO, N-NAM, O-NAM, OWNER, B-BIB, G-TOB, G-ORI, G-SOB, G-ORE, GEO-1, GEO-2, HPDES, HPHIS, HXPLO, LITH1, LITH2, LROCK, MIN-1, MIN-2, M-GEO, R-MEA, R-IND, R-INF, R-UND, RESER, R-REM, AND RASAY. RECORD-IDENTIFIER ORDERING FOR A SPECIFIC RECORD IS DEFINED BY THE PARAGRAPH REFERENCED IN EACH RESPECTIVE RECORD DESCRIPTION.

2.2.1 RECORD DESCRIPTION FOR STATE (STATE/COUNTRY AND COUNTY/POLITICAL SUBDIVISION)

DATA ELEMENT NAME	START POSITION	LAST POSITION	LENGTH	COMMENTS OR PARAGRAPH REFERENCE
SEQUENCE-NUMBER	1	10	10	2.3.1
RECORD-NAME	11	15	5	VALUE IS "STATE"
NOT USED	16	20	5	
STATE-NAME	21	40	20	STATE OR COUNTRY
COUNTY-NAME	41	60	20	COUNTY OR POLITICAL SUBDIVISION
NOT USED	61	120	60	

2.2.2 RECORD DESCRIPTION FOR IDENT (IDENTIFICATION)

DATA ELEMENT NAME	START POSITION	LAST POSITION	LENGTH	COMMENTS OR PARAGRAPH REFERENCE
SEQUENCE-NUMBER	1	10	10	2.3.1
RECORD-NAME	11	15	5	VALUE IS "IDENT"
NOT USED	16	20	5	
PROPERTY-NAME	21	55	35	2.3.2
TYPE-OF-OPERATION	56	67	12	2.3.3
CURRENT-STATUS	68	80	13	2.3.4
NOT USED	81	120	40	

2.2.3 RECORD DESCRIPTION FOR LOCAT (LOCATION)

DATA ELEMENT NAME	START POSITION	LAST POSITION	LENGTH	COMMENTS OR PARAGRAPH REFERENCE
SEQUENCE-NUMBER	1	10	10	2.3.1
RECORD-NAME	11	15	5	VALUE IS "LOCAT"
NOT USED	16	20	5	
LATITUDE-DIRECTION	21	21	1	2.3.5
LATITUDE	22	27	6	2.3.6
LONGITUDE-DIRECTION	28	28	1	2.3.7
LONGITUDE	29	35	7	2.3.8
POINT-OF-REFERENCE	36	48	13	2.3.9
ELEVATION-AND-PRECISION	49	58	10	2.3.10
DATUM	59	60	1	2.3.11
YEAR-FIELD-CHECKED	61	64	4	2.3.12
NOT USED	75	120	57	

2.2.4 RECORD DESCRIPTION FOR UTM (UNIVERSAL TRANSVERSE MERCATOR)

DATA ELEMENT NAME	START POSITION	LAST POSITION	LENGTH	COMMENTS OR PARAGRAPH REFERENCE
SEQUENCE-NUMBER	1	10	10	2.3.1
RECORD-NAME	11	15	5	VALUE IS "UTM "
NOT USED	16	20	5	
ZONE-AND-HEMISPHERE	21	23	3	2.3.13
NORTHING	24	30	7	2.3.14
EASTING	31	36	6	2.3.15
NOT USED	37	120	84	

2.2.5 RECORD DESCRIPTION FOR TOPOG (TOPOGRAPHY)

DATA ELEMENT NAME	START POSITION	LAST POSITION	LENGTH	COMMENTS OR PARAGRAPH REFERENCE
SEQUENCE-NUMBER	1	10	10	2.3.1
RECORD-NAME	11	15	5	VALUE IS "TOPOG"
NOT USED	16	20	5	
QUADRANGLE	21	38	18	2.3.16
MAP-NAME-AND-SCALE	39	62	24	2.3.17
DOMAIN	63	76	14	2.3.18
NOT USED	77	120	44	

2.2.6 RECORD DESCRIPTION FOR BASIN (RIVER BASIN)

DATA ELEMENT NAME	START POSITION	LAST POSITION	LENGTH	COMMENTS OR PARAGRAPH REFERENCE
SEQUENCE-NUMBER	1	10	10	2.3.1
RECORD-NAME	11	15	5	VALUE IS "BASIN"
NOT USED	16	20	5	
RIVER-BASIN-NAME	21	44	24	2.3.19
RIVER-BASIN-CODE	45	48	4	2.3.20
HYDROLOGIC-UNIT-CODE	49	56	8	2.3.21
NOT USED	57	120	64	

2.2.7 RECORD DESCRIPTION FOR HOL (MINERAL HOLDING)

DATA ELEMENT NAME	START POSITION	LAST POSITION	LENGTH	COMMENTS OR PARAGRAPH REFERENCE
SEQUENCE-NUMBER	1	10	10	2.3.1
RECORD-NAME	11	15	5	VALUE IS "HOL "
NOT USED	16	20	5	
HOLDING-1	21	33	13	2.3.22
HOLDING-2	34	46	13	2.3.22
HOLDING-3	47	59	13	2.3.22
NOT USED	60	120	61	

2.2.8 RECORD DESCRIPTION FOR REFER (REFERENCE)

DATA ELEMENT NAME	START POSITION	LAST POSITION	LENGTH	COMMENTS OR PARAGRAPH REFERENCE
SEQUENCE-NUMBER	1	10	10	2.3.1
RECORD-NAME	11	15	5	VALUE IS "REFER"
NOT USED	16	20	5	
EVALUATOR	21	30	10	2.3.23
MINERAL-PROPERTY-FILE	31	36	6	2.3.24
MINE-MAP-REPOSITORY	37	37	1	2.3.25
CORE-LIBRARY	38	38	1	2.3.26
MSHA-IDENTIFICATION-NUMBER	39	45	7	2.3.27
GEOLOGICAL-SURVEY-CRIB	46	52	7	2.3.28
TYPE-OF-EVALUATION	53	53	1	2.3.29
YEAR-OF-INFORMATION	54	57	4	2.3.30
MAINTENANCE-DATE	58	61	4	2.3.31
NOT USED	61	120	59	

2.2.9 RECORD DESCRIPTION FOR PLANT (TYPE AND IDENTIFIER)

DATA ELEMENT NAME	START POSITION	LAST POSITION	LENGTH	COMMENTS OR PARAGRAPH REFERENCE
SEQUENCE-NUMBER	1	10	10	2.3.1
RECORD-NAME	11	15	5	VALUE IS "PLANT"
NOT USED	16	20	5	
PLANT-TYPE	21	26	6	2.3.32
PLANT-IDENTIFIER	27	32	6	2.3.33
NOT USED	33	120	87	

2.2.10 RECORD DESCRIPTION FOR PLS (PUBLIC LAND SURVEY)

DATA ELEMENT NAME	START POSITION	LAST POSITION	LENGTH	COMMENTS OR PARAGRAPH REFERENCE
SEQUENCE-NUMBER	1	10	10	2.3.1
RECORD-NAME	11	15	5	VALUE IS "PLS "
NOT USED	16	20	5	
PRINCIPAL-MERIDIAN	21	34	14	2.3.34
TOWNSHIP-NUMBER-AND-DIRECTION	35	39	5	2.3.35
RANGE-NUMBER-AND-DIRECTION	40	44	5	2.3.36
SECTION-NUMBER	45	46	2	2.3.37
SECTION-SUBDIVISION	47	52	6	2.3.38
SURVEY-STATUS	53	58	6	2.3.39
NOT USED	59	120	62	

2.2.11 RECORD DESCRIPTION FOR COMMO (COMMODITY)

DATA ELEMENT NAME	START POSITION	LAST POSITION	LENGTH	COMMENTS OR PARAGRAPH REFERENCE
SEQUENCE-NUMBER	1	10	10	2.3.1
RECORD-NAME	11	15	5	VALUE IS "COMMO"
RECORD-IDENTIFIER	16	19	4	2.3.40
NOT USED	20	20	1	
COMMODITY	21	34	14	2.3.41
MODIFIER	35	56	22	2.3.42
MARKETABILITY	57	57	1	2.3.43
STANDARD-INDUSTRIAL-CLASSIFICATION-CODE	58	61	4	2.3.44
NOT USED	62	120	59	

2.2.12 RECORD DESCRIPTION FOR N-NAM (ADDITIONAL OR ALTERNATE NAME)

DATA ELEMENT NAME	START POSITION	LAST POSITION	LENGTH	COMMENTS OR PARAGRAPH REFERENCE
SEQUENCE-NUMBER	1	10	10	2.3.1
RECORD-NAME	11	15	5	VALUE IS "N-NAM"
RECORD-IDENTIFIER	16	19	4	2.3.40
NOT USED	20	20	1	
NAME	21	55	35	2.3.45
NOT USED	56	120	65	

2.2.13 RECORD DESCRIPTION FOR O-NAM (OWNER NAME)

DATA ELEMENT NAME	START POSITION	LAST POSITION	LENGTH	COMMENTS OR PARAGRAPH REFERENCE
SEQUENCE-NUMBER	1	10	10	2.3.1
RECORD-NAME	11	15	5	VALUE IS "O-NAM"
RECORD-IDENTIFIER	16	19	4	2.3.40
NOT USED	20	20	1	
OWNER-OR-OPERATOR-NAME	21	76	56	2.3.46
NOT USED	77	120	44	

2.2.14 RECORD DESCRIPTION FOR OWNER (OWNERSHIP)

DATA ELEMENT NAME	START POSITION	LAST POSITION	LENGTH	COMMENTS OR PARAGRAPH REFERENCE
SEQUENCE-NUMBER	1	10	10	2.3.1
RECORD-NAME	11	15	5	VALUE IS "OWNER"
RECORD-IDENTIFIER	16	19	4	2.3.40
NOT USED	20	20	1	
STATUS	21	28	8	2.3.47
OWNERSHIP-PERCENT	29	34	6	2.3.48
HOME-OFFICE	35	54	20	2.3.49
YEAR-OF-INFORMATION	55	58	4	2.3.50
NOT USED	59	120	62	

2.2.15 RECORD DESCRIPTION FOR B-BIB (BIBLIOGRAPHY)

DATA ELEMENT NAME	START POSITION	LAST POSITION	LENGTH	COMMENTS OR PARAGRAPH REFERENCE
SEQUENCE-NUMBER	1	10	10	2.3.1
RECORD-NAME	11	15	5	VALUE IS "B-BIB"
RECORD-IDENTIFIER	16	19	4	2.3.40
NOT USED	20	20	1	
BIBLIOGRAPHY	21	80	60	2.3.51
NOT USED	81	120	40	

Appendix D

Plotting Procedures Developed for the CCP by J. F. Uhlinger

MAPPING OF MOTHER'S RESIDENCE ADDRESSES
(From CCP Protocol Book)

I INTRODUCTION:

Identification of residence location for cases and controls is the first step in the determination of case control ratios within geographic areas for ecological analysis. Once located and plotted on topological maps, location coordinates and elevation information can be captured for use in ecologic analysis. Because the birth certificate records are confidential, addresses must be plotted without disclosing the information contained in the birth certificates to organizations or personnel that could assist in the plotting of the addresses. Thus, the following procedures for plotting the addresses was designed to capture the required information confidentially.

The location of the case and control addresses as well as some format of environmental information listed above will serve as a data base for utilizing computerized Geographical Informaion Systems (GIS) capabilities in the ecological study of craniosynostosis. The use of GIS capabilities essentially enables comparison of geographic attributes associated with the data. The number of cases and controls within specified areas is to be automatically generated for comparison to the distribution of potential environmental exposure measures. This comparison is expected to identify distributional relationships that can be utilized in generating more specific hypotheses regarding the the occurance of craniosynostosis in Colorado.

The specific purpose of plotting the study subject's addresses on to 1:50k and 1:100k scale USGS topographic maps is to capture universal transverse mercator easting and northing coordinates (UTM coordinates in meters), elevation information, and associated hydrological unit codes. This information is to be read from the plot map manually or by utilizing computerized mapping processes resident in the GIS. The resulting data is to be condensed and formatted into a data base that will facilitate the use of overlay maps of the various data on 1:500,000 scale maps to generate disease occurrence values that are intended to be used in epidemiological analyses.

II DATA SOURCES

RETROSPECTIVE STUDY DATA:

Birth certificate data for cases and controls utilized in the study are downloaded from computer tapes into computer files temporarily resident on the UCHSC VAX. These files were copied to a 5 1/4 floppy diskette. The files on this diskette were copied to the resident disk drive of the Craniosynostosis P.C. for data processing using programs contained within SAS. The 5 1/4 inch floppy was preserved as a back up copy of the birth certificate data incase of unforseen or irrepairable alteration of the resident files on the personal computer.

A PCSAS program was written to extract, format, sort, and list variables from the birth certificate data that were deemed pertinent to the plotting of the mother's residence at the time of birth of the child for cases and controls. This program, variable list, and variable records is documented in appendix XX. The resulting list of address variables was sorted by county codes for use in the plotting procedures.

PLOT MAP RESOURCES:

The Craniosynostosis Registry has three sets of USGS topographic maps (USGS topo maps) on which addresses for the retrospective studies were plotted. The first set contains 23 1:50,000 scale maps of Routt, Eagle, Clear Creek, Boulder, Jefferson, Douglas, Denver, Arapahoe and Adams counties. This set was chosen because the map scale allows better resolution to plot the larger number of cases and controls in these counties.

The second map set includes 54 1:100,000 scale USGS topo maps. This set covers all of Colorado at a scale appropriate to capture UTM coordinates and elevation information for the study.

The last map set includes a 1:500,000 scale USGS Colorado State topo map and a 1:500,000 scale USGS Hydrologic Unit Boundary map of Colorado. These maps were used to composite data from the 1:50,000 and 1:100,000 scale plot maps as a reference for quality control steps presented in the Quality Contol and Quality Analysis section.

The 1:50,000, 1:100,000, and 1:500,000 scale maps are stored alphabetically by scale in the map filing cabnet in the DCEED of the Department of Health.

REFERENCE MAP RESOURCES:

Because the topo maps generally do not have labelled street names, required in the plotting procedures, higher resolution street maps are used as reference for the location of addresses. These reference maps or street maps contain a street index, street address ranges are detailed and current enough to show all streets and most landmarks. The following is a list of reference map resources available for use in this project:

PIERSON GRAPHICS:

Pierson Graphics Corp. "Denver Regional (Close Up) Atlas - Street Guild", and maps of Loveland and Ft. Collins, 1986.

The Pierson Graphics Corp. "Denver Regional (Close Up) Atlas - Street Guild" is a catalogue of "high resolution" street maps. Each page is a map approximately 5 city blocks by 5 city blocks in area. The catalogue contains county indexes of the metro area. This index contains street names and address ranges along the streets. The maps have all streets named and address range numbers indicated generally by block.

Locating an address on these maps is accomplished by using the index and street range. Once the street is located the address range along the street is interpolated in a linear fashion between range numbers as indicated on the map. Most, if not all addresses should be identifiable on these maps.

To transfere an address location to the USGS topo, recognition techniques are usually effective. Triangulating by direction and/or scale distance also works well. The scale factor that relates the Pierson Graphics Atlas maps to the 1:50,000 scale USGS topos is 5/12 (5 inches on the topos = 12 inches on the atlas maps thus). Another technique that can be used is quartering using section, township and range information.

"REDI" MICO-FICHE:

Real Estate Data Inc. (REDI) Micro - fiche resources (for Adams, Arapahoe, Boulder, Denver, Douglas, Elbert, El Paso, and Larimer counties), 1986.
Contact: Gaylan Conrad - Sales Representative

The Real Estate Data Inc. (REDI) Micro - fiche resources (for Admas, Arapahoe, Boulder, Denver, Douglas, Elbert, El Paso, and Larimer counties), 1986 resource presents various formats of maps. These include county maps, tax maps, and platt maps. These maps are also indexed in various ways; by owner name, tax criteria, subdivision and section, township and range. To use this resource a high degree of familiarity with the various indexing systems must be acquired. Thus, use of REDI should be reserved for problem addresses.

Transferring of points to USGS topos must be done by using recognizable streets and landmarks or quartering using section,

township and range information.

STATE HIGHWAYS MAPS:

Rural Address Program Maps

Rural Address Program (RAP) maps from Colorado State Dept. of Highways (for Eastern Plains and 50% of Mountain counties), circa 1975.

The Rural Address Program (RAP) maps from Colorado State Dept. of Highways (for Eastern Plains and 50% of Mountain counties), circa 1975 are also in various formats. They are referenced by section, township and range as well as on county reference maps. These maps are generally 10 years old and therefore lacking more recent information. The RAP maps include addressing systems for inferring addresses especially on county roads and outside city limits. Here again, some familiarity with the format is advised, but is generally quickly learned.

Transfer of points to the USGS topo is accomplished using any of the above mentioned techniques.

City Platt Maps

City Platt maps Colorado State Dept. of Highways, circa 1960.

Contact: George Pugh, Rodger Gilpin

Another resource available at the Dept. of Highways is the catalogue of City Plats. These are city maps for most incorporated cities and towns in Colorado. These maps contain street names within incorporated city boundaries, but do not contain any address range information. To locate addresses with these maps, the plotter must first find the appropriate street on the map and have the map xeroxed. This map is then used as reference when telephoning the local city planning or assessor's offices. The telephoning procedure is described in depth in reference to use of assessor's maps and resources.

POST OFFICE "AIS" RESOURCES:

Carrier Route Maps at Denver, Grand Junction, and Colorado Springs Post Offices Address Information Service (AIS) maps (various update dates for various localities).

Contact: Denver - Adolf Ruiz
Grand Junction - Darlene Swenson
Colorado Springs - Mary Jane Carter

Carrier Route Maps at Denver, Grand Junction, and Colorado Springs Post Offices Address Information Service (AIS) offices can also be utlilized. These maps are designed for use in delivering mail. Thus, information pertaining to many rural communities using only Post Office Boxes for mail delivery is not

referenced in these maps.

The Postal Service assists in determining the ZIP +4 code for the addresses and appropriate Carrier Route maps are pulled as reference maps. Postal Personnel are available as a source of information and procedures. The three management System Offices are located in Denver, Colorado Springs and Grand Junction. Thus, a conserted effort to plot all points using the available reference maps must be made. Lists of addresses that can not be plotted using other reference map resources must be kept accurately and the organization of work flow must adhere rigidly to this protocol in order to facilitate the efficient plotting of addresses. Essentially, only one trip to the postal facilities should be required.

The AIS offices have offered their services for small lists of addresses (5-10 addresses). A list of address ranges is mailed to the appropriate AIS office, and they return xerox copies of maps containing the address ranges in question. These maps are then used as reference maps in plotting. For larger lists of addresses, a trip to the appropriate AIS is in order to reduce the labor burden on the Postal Service.

Address location transfer from the carrier route maps is best done using recognizable map features or quartering techniques.

ASSESSOR MAPS:

County Assessor Maps and Indexes, located at specific County Assessors' Offices or County Planners' Offices through out Colorado.

County Assessor Maps and Indexes, located at specific County Assessors' Offices or County Planners' Offices through out Colorado, can be used in two ways. The first is a trip to the appropriate office to utilize their maps (usually updated RAP maps or plat maps similar to that of the REDI resources). The second way of utilizing these resources is through the use of telephoning procedures described below.

REALTOR MAPS:

Realtor Maps, upon request from Realtors in very specific areas, can also be used. Realtor are usually happy to mail subdivision maps of localized areas. These maps usually do not contain address range information, but are a good resource for telephoning (descibed below).

III MAPPING PROCEDURES

WORK FLOW:

Prior to plotting of addresses, the address data for the study is sorted and listed by the count-city code (variable). Two separate lists were generated one for cases one for controls. Plotting is done on a county by county basis according to the WORK FLOW SCHEMATIC DIAGRAM (last page). This is done to most efficiently utilize the most current, complete and availible reference map resources first. As addresses are plotted according to plotting procedures and identified according to the POINT ID SCHEME, they are checked off on the list. Problem addresses and difficult to plot addresses are listed on separate lists for future plotting.

The plotting of addresses starts in the metropolitan areas for which the best reference material is availible. The 1:50,000 scale topo maps are used preferentially to the 1:100,000 scale topo maps. Where 1:50,000 scale maps do not exist according to the coverage index maps, 1:100,000 scale topo maps are used.

In general all addresses that can be plotted according to the technical procedures described in TECHNICAL PLOTTING PROCEDURES SECTION for a county are plotted using a specific reference map resource. Those that cannot be plotted will fall into one of two categories and noted on the data list. The first is "impossible to plot addresses". The second is "problem addresses".

"Impossible to plot addresses" are those that are indicated on the data address lists by:

Post Office Box numbers (P.O. Boxes)
Rural Route or Postal Carrier Route numbers
Incomplete or missing addresses

(NOTE: If Mother's street address is unavailable, mother's mailing address is plotted only if state, county, and zip codes are consistent with mother's street address; otherwise it is treated as impossible to plot).

Attempts are made to identify the locations of the imposiible to plot addresses using identified procedures and resources as well as phone calls. Because addressing systems vary from county to county, and systems often change, efforts to plot these adresses may not be fruitful. When all procedures and resources have been exhuasted, these "impossible to plot" addresses are plotted at the location of the Post Office designated by the ZIP code of the particular address. The procedure for plotting of the Post Office locations includes making a list of the Post Offices to be plotted, identifying an address for the Post Offices, and following the TECHNICAL PLOTTING PROCEDURES for location and transfer. The locations of

these Post Offices are designated by an "+ or x" of the appropriate color (as described in the POINT ID SCHEME section). The corresponding identification number has a "P" added as a suffix indicator to indicate that the point is an impossible to plot address.

"Problem addresses" are those for which locations cannot be located using a specific reference map resource. These addresses should eventually be plotted utilizing the next resource in the WORK FLOW SCHEMATIC DIAGRAM. If not contained in the subsequent reference map resources, the address can be checked by Vital Records from which the data originated.

POINT ID SCHEME:

The plot point identification scheme for the retrospective CDC study utilizes color and an identification number ranges to identify and differentiate between cases and controls on maps and in data lists. The color scheme utilizes red for cases and black for controls. The identification number scheme is described as follows:

Data Type	# Range	Color
Retrospective CDC Cases	= 4000 - 4999	RED
Retrospective CDC Controls	= 5000 - 5999	BLACK

The scheme for designating case/control ID numbers within each data type category varies. For cases, the three digit CDCID numbers are preceded by a four (4). Thus, numbers are assigned within the range of 4000 - 4999. For example, a retrospective study case with a CDCID of 114 is assigned a map number of 4114 and colored red. If an address is represented as a Post Office Box on the data list a "P"indicator suffix is added (see "impossible to plot" addresses).

Controls are assigned a number consistent with the data list of addresses as sorted by county, city codes and date of birth. They range between 5000 - 5999 inclusive. The identification (ID) number corresponds to the observation number in this list. The three digit representation of the observation number or ID number is preceded by a five (5). For example, a retrospective study control with an observation number of 325 is designated as 5325 and colored black. If an address is represented as a Post Office Box on the data list a "P"indicator suffix is added (see "impossible to plot" addresses).
(note: during digitizing the P indicator is transformed to a .9 numerical suffix in digitizing)

TECHNICAL PLOTTING PROCEDURES

LOCATION OF ADDRESSES ON REFERENCE MAPS:

An address is precisely located on the appropriate reference map utilizing the reference map road indexing system and address range information. The specific technique for locating addresses on the reference maps varies due to differences between the resources and their designed purposes. Many of these resources require familiarity and experience to use due to their complexity and format.

When using hard copy or paper reference maps (Pierson Graphics maps) the following general procedure is as follows. The location is noted with a pencil as a circle approximately 1/4" - 1/2" in diameter. The location of this circle is then transfered to the appropriate USGS topo map and accurately plotted in ink and identified as described in the POINT ID SCHEME section of this protocol.

TRANSFERE OF REFERENCE MAP ADDRESS LOCATIONS TO USGS TOPOS:

The transfer procedure varies for each reference map resource that is used. In general, map reading and location identification skills are necessary. There are three ways of transferring point location information from the reference resources to the USGS topos. For the purposes of this project the are called "recognition", "triangulation" and "quartering".

The recognition transfer technique requires identification of specific map features recognizable on both the reference maps and on the USGS topo. The plotter should locate recognizable street intersections or land marks that may be consistent with both maps and locate the point in relation to these recognizable landmarks. For instance, if an address is located on the reference map half way between two cross streets along an identifiable street, the address is located half way between those cross streets along the identifiable street on the USGS topo. Likewise, if an address is located on the reference map along a street with an identifiable curve or near an identifiable land mark, it is plotted accordingly on the USGS topo.

If there are no identifiable map features in close proximity to the address location on the reference map or the USGS topo (as is the case in some of the newer sub-developments where the USGS topo has not been updated), other techniques may be used.

Triangulation is used when there are at least two decernable features identifiable on both maps (the reference map and the USGS topo), but they are located at a distance far enough away from the address location that the location of the point cannot be transferred using map feature recognition techniques.

The plotter measures directions to the address location point from each of the decernable land marks using a reference

line and protractor on the reference map. These directions lines are then drawn on the USGS topo from the two decernable land marks. Where these directional lines intersect on the plot map is the transferred address location.

Triangulation can also be done using linear features such as perpendicular streets if a map scale is known. The plotter measures the map distance (in inches) from the address location on the reference map to a north-south street. This distance is converted to real distance by multiplying the map distance in inches by the scale of the map. Then this real distance is divided by the map scale on the USGS topo to give a plot map distance. This is measured and drawn on the USGS topo as a line parallel to the north-south street at the appropriate map distance. This is repeated for an east-west street. The intersection of the two drawn lines is the transferred location of the address.

Quartering is another technique. This requires knowledge of the section, township and range in which the reference map address location is plotted (this information is generally on every reference map that is used).

The plotter locates the appropriate section, township and range on the reference map and divides that section evenly into quarter sections. The plotter then further divides the quarter section containing the reference map address location into four quarters and so on. This process is similarly repeated for the appropriate section on the USGS topo until the point is transferred with certainty. Thus, the point is transferred by narrowing down square areas consistantly on both maps.

The location of an address is plotted in ink on the appropriate 1:50,000, or 1:100,000 scale topo map. A standard size circle of 5/64 inch diameter (from a template) is used to represent the location of an address that is plotted with a fairly high level of confidence on either scale USGS topo. This confidence is attributed to locations of addresses in areas where street names and address ranges are well defined on reference maps, and where the transfer of the points from the reference map to the topo map is facilitated by numerous or well defined landmarks such as street intersections or parks. In general, if the mapper feels confident that the location of the address is very close to its "real life" location on the reference map and one of the three transfere procedures works effectively, then the point will be represented as a 5/64 inch diameter circle according to the appropriate POINT ID SCHEME.

For addresses that cannot be located so confidently on the reference map and thus are problematically transferred, a larger plot map circle is used. Generally the smallest circle that encompasses the probable location(s) is used. A 5/16 inch diameter circle (found on the same template as the 5/64 inch diameter circle) is the largest circle that should be used due to the various reference map resources availible and the work flow

protocol (described later).

The causes of potential uncertainty may result from inconsistency between the reference map and the topo map. It may result from the lack of desernable landmarks, or more typically, lack of resolution in the street address range on the reference map. With respect to these cases, the location of the address is plotted to a best guess location and noted appropriately with a maximum size circle of 5/16 inch diameter on the plot map.

A special address location plotting convention is implimented if two addresses are coincident on the plot map. This occurs when two separately reported addresses are located in the same apartment building or one household reports more than one case, control or both. These coincident points are represented as a 5/64 inch diameter circle inside a 5/64 inch diameter cicle. Each circle is colored appropriately according to the POINT ID SCHEME.

TELEPHONING TO GET ADDRESS LOCATIONS:

For counties with address lists of 1 to 5 addresses and limited reference map resources or any problem addresses, telephoning Post Offices, County Assessors' Offices and realtors is the most efficient means of getting address locations. Caution must be excersized to protect the confidentiality of the addresses. The best means of doing this is to give only address range in formation instead of the actual address. Confidentiallity can also be protected by not identifying yourself, but this is only appropriate for inquiry about one or two addresses.

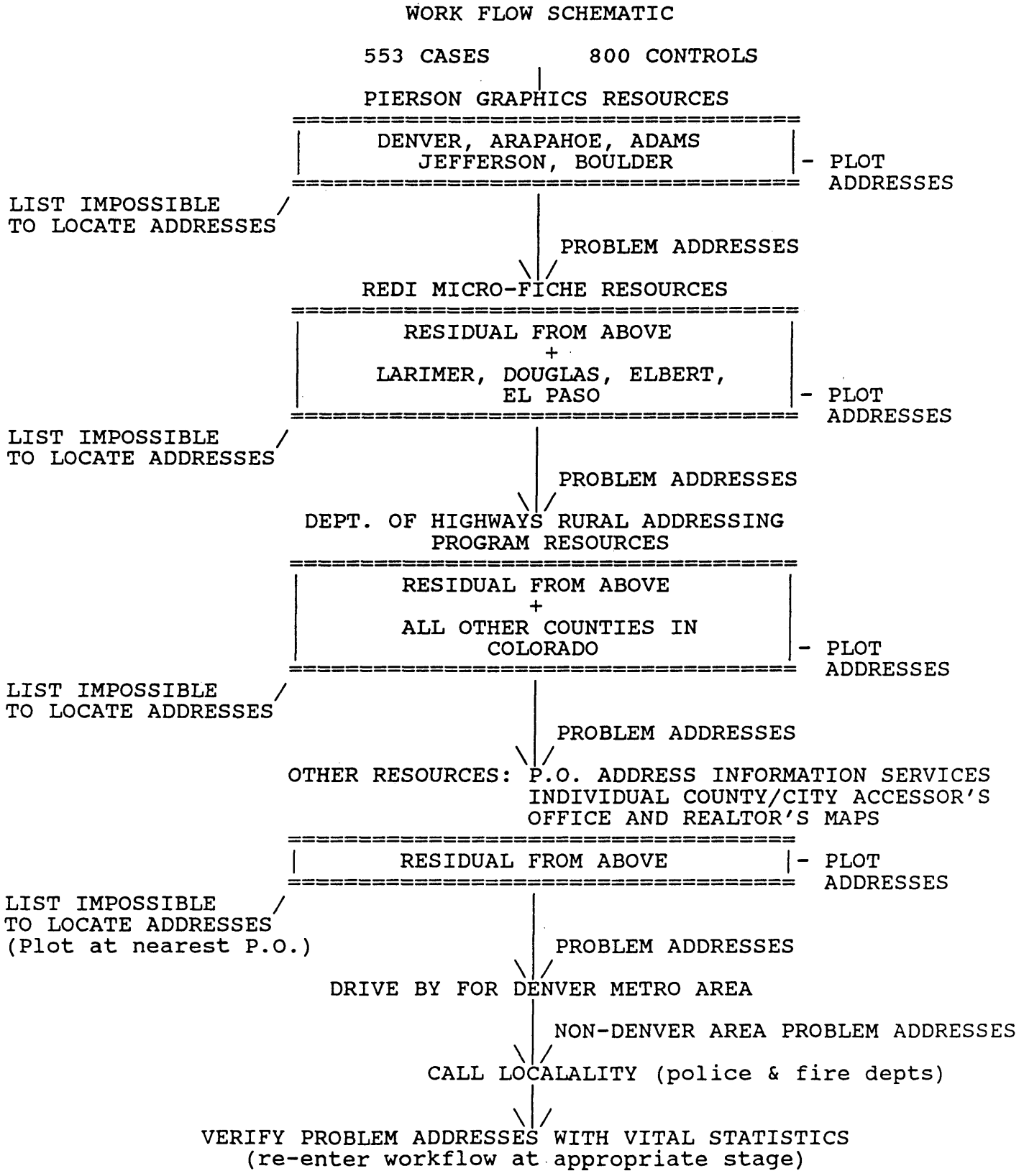
The process of address location is best facilitated using a reference map with street names such as CITY PLATS, REALTOR MAPS or other maps. The caller asks for the location of addresses ranges and narrows down the exact location of the address number by interrogation. The address is the located on the reference map used in the phone call. The following are hypothetical conversations that exemplify techniques to get information from realtors, city planning offices, county accessor's offices, or any other resource with out giving out the actual address number which would jeopardize the confidentiallity of the birth certificate data.

1) For address within city limits and reference map has street labeled with no address range information:

Caller: "Hello, I need to locate an address."
Assessor: "OK, How can I Help?"
Caller: "I have an address on Main Street in the city of Buena Vista. Can you tell me what cross streets designate the 600 block of main?"
Assessor: "Yes", ...gives response.....
Caller: Indicates 600 block on reference map, "thank you very much."

2) For rural areas (county roads) (address in question is 11640 County Road 321) County Road 321 located on reference map:

Caller: "Hello, can you help me locate an address on County Road 321."
Assessor: "OK, how can I help?"
Caller: "Can you tell me the section, township and range, and any distinguishing road characteristics for 11645 CR321?"
Assessor: "Let me check our maps... I can't find 11645, but I did find 11640 in" .... gives section, township and range and quarter section for address.
Caller: Locates section,township and range and quareter section on reference map. "Is it near any distinguishing map features like a river, bend in the road or intersection?"
Assessor: "Yes"..... gives details.
Caller: Plots location, "Thank you very much."


WORK FLOW SCHEMATIC
553 CASES
800 CONTROLS
PIERSON GRAPHICS RESOURCES
DENVER, ARAPAHOE, ADAMS
JEFFERSON, BOULDER
- PLOT ADDRESSES
LIST IMPOSSIBLE TO LOCATE ADDRESSES
PROBLEM ADDRESSES
REDI MICRO-FICHE RESOURCES
RESIDUAL FROM ABOVE
+
LARIMER, DOUGLAS, ELBERT,
EL PASO
- PLOT ADDRESSES
LIST IMPOSSIBLE TO LOCATE ADDRESSES
PROBLEM ADDRESSES
DEPT. OF HIGHWAYS RURAL ADDRESSING PROGRAM RESOURCES
RESIDUAL FROM ABOVE
+
ALL OTHER COUNTIES IN
COLORADO
- PLOT ADDRESSES
LIST IMPOSSIBLE TO LOCATE ADDRESSES
PROBLEM ADDRESSES
OTHER RESOURCES: P.O. ADDRESS INFORMATION SERVICES
INDIVIDUAL COUNTY/CITY ACCESSOR'S OFFICE AND REALTOR'S MAPS
RESIDUAL FROM ABOVE
- PLOT ADDRESSES
LIST IMPOSSIBLE TO LOCATE ADDRESSES
(Plot at nearest P.O.)
PROBLEM ADDRESSES
DRIVE BY FOR DENVER METRO AREA
NON-DENVER AREA PROBLEM ADDRESSES
CALL LOCALALITY (police & fire depts)
VERIFY PROBLEM ADDRESSES WITH VITAL STATISTICS
(re-enter workflow at appropriate stage)

Appendix E

FORTRAN Programs Written by L. Ayers at CIRES

```
      PROGRAM BREAKOUT
C
C     THIS PROGRAM READS A MINERALS AVAILABILITY FIELD OFFICE
C     TAPE FOR COLORADO AND WRITES SELECTED INFORMATION TO A
C     NEW FILE. THIS INFORMATION WILL LATER BE JOINED TO AN
C     ARC/INFO COVERAGE WHICH IS PRODUCED USING THE POINTS DEFINED
C     IN GENPOINTS.DAT (DURING COMBO2.FOR). FOLLOW THIS PROGRAM WITH
C     COMBO.FOR AND COMBO2.FOR.
C
C     WRITTEN OCTOBER 1988 BY LEE MARGARET AYERS
C     CSES/CIRES  U. OF COLORADO, BOULDER
C
      INTEGER CID,COMODK,MINEID,ZONE,NORTH,EAST,USERID
      REAL RNORTH, REAST
      CHARACTER*5 RECID
      CHARACTER*12 OTYPE
      CHARACTER*13 STATUS
      CHARACTER*14 COMMOD
      CHARACTER*22 MODIF, COM(8,2)
      CHARACTER*35 MNAME
      CHARACTER*105 THREST
      CHARACTER*120 REC
C
C
      OPEN(UNIT=11,FILE='MANSC2.TAP',STATUS='OLD',RECL=120,ERR=99)
      OPEN(UNIT=12,FILE='OUT.DAT',STATUS='NEW',RECL=450,ERR=99)
      OPEN(UNIT=13,FILE='POINTS.DAT',STATUS='NEW',ERR=99)
      OPEN(UNIT=14,FILE='POINTS12.DAT',STATUS='NEW',ERR=99)
      OPEN(UNIT=15,FILE='LOLA12.DAT',STATUS='NEW',ERR=99)

C
C
      READ(11,101,END=300)REC
101   FORMAT(A120)
      READ(REC,103)MINEID,RECID,THREST
103   FORMAT(3X,I7,A5,A105)
      CID = MINEID
      USERID = 1
C
C
      DO 50 I = 1, 200000
        CID = MINEID
        L = 1
        M = 1
           DO 25 J = 1, 8
             DO 30 K = 1, 2
               COM(J,K) = ' '
30           CONTINUE
25         CONTINUE
C
C
75      IF (MINEID .EQ. CID) THEN
          IF (RECID .EQ. 'IDENT') THEN
            READ(THREST,105)MNAME, OTYPE, STATUS
105         FORMAT(5X,A35,A12,A13)
          ELSEIF (RECID .EQ. 'UTM') THEN
            READ(THREST,107)ZONE, NORTH, EAST
107         FORMAT(5X,I2,X,I7,I6)
          ELSEIF (RECID .EQ. 'COMMO') THEN
            READ(THREST,109)COMMOD, MODIF
109         FORMAT(5X,A14,A22)
              IF (M .EQ. 1) COM(L,M) = COMMOD
90            CONTINUE
              IF (M .EQ. 2) COM(L,M) = MODIF
              M = M + 1
              IF (M .EQ. 2) GOTO 90
```

```
              L = L + 1
              M = 1
            ENDIF
              READ(11,101,END=300)REC
              READ(REC,103,ERR=300)MINEID,RECID,THREST
            GOTO 75
          ENDIF
              RNORTH = NORTH/1
              REAST = EAST/1
          WRITE(12,201)CID,MNAME,OTYPE,STATUS,ZONE,RNORTH,REAST,
     +    ((COM(L,M),M = 1, 2), L = 1,8),USERID
          WRITE(13,202)CID,EAST,NORTH
            IF (ZONE .EQ. 12) THEN
              WRITE(14,202)CID,EAST,NORTH
              WRITE(15,203)EAST, NORTH
            ENDIF
201       FORMAT(I7,A35,A12,A13,I2,F12.4,F12.4,16A22,I5)
202       FORMAT(I7,X,I6,X,I7)
203       FORMAT(I6,X,I7)
          USERID = USERID + 1
50      CONTINUE
C
C
300     CONTINUE
        CLOSE(UNIT=11,STATUS='KEEP',ERR=9999)
        CLOSE(UNIT=12,STATUS='KEEP',ERR=9999)
        CLOSE(UNIT=13,STATUS='KEEP',ERR=9999)
        CLOSE(UNIT=14,STATUS='KEEP',ERR=9999)
        CLOSE(UNIT=15,STATUS='KEEP',ERR=9999)
C
C
        STOP
99      STOP 'ERROR OPENING'
9999    STOP 'ERROR CLOSING'
        END
```

```
      PROGRAM COMBO2
C     READS TWO SEPARATE FILES AND WRITES THE PARTS OF BOTH
C     INTO A NEW FILE. COMBO.DAT CONTAINS MINEIDS AND ZONE13
C     COORDINATES. POINTS.DAT CONTAIN BOTH ZONE 12 AND 13 COORDS.
C     PLUS MINEIDS. BECAUSE ARC/INFO FILES WILL ONLY HANDLE
C     5 INTEGERS, A NEW ITEM, USER ID, HAS BEEN ADDED AS THE REFERENCE
C     POINT FOR MINEIDS. COMBO2.DAT WILL CONTAIN USER-IDS,
C     MINEIDS AND ZONE 13 COORDS. ANOTHER FILE, GENPOINTS.DAT WILL
C     CONTAIN USER-IDS, AND XY COORD. VALUES.
C
C     WRITTEN BY LEE MARGARET AYERS, OCTOBER, 1988
C     CSES/CIRES U. COLORADO, BOULDER
C
      INTEGER MINEID, MINEID2, USERID
      REAL NORTH, EAST, NORTH2, EAST2
C
C
      OPEN(UNIT=12,FILE='COMBO.DAT',STATUS='OLD',ERR=99)
      OPEN(UNIT=13,FILE='POINTS.DAT',STATUS='OLD',ERR=99)
      OPEN(UNIT=14,FILE='COMBO2.DAT',STATUS='NEW',ERR=99)
      OPEN(UNIT=15,FILE='GENPOINTS.DAT',STATUS='NEW',ERR=99)

C
C
      READ(12,102)MINEID2, EAST2, NORTH2
      USERID = 1
C
C
      DO 50 I = 1, 70000
        READ(13,101,END=300)MINEID, EAST, NORTH
101     FORMAT(I7,X,F6.0,X,F7.0)
102     FORMAT(I7,2F12.4)
C
C
          IF (MINEID .NE. MINEID2) THEN
            WRITE(14,201)USERID,MINEID
            WRITE(15,202)USERID,EAST,NORTH
          ELSEIF(MINEID .EQ. MINEID2) THEN
            WRITE(14,201)USERID,MINEID
            WRITE(15,202)USERID,EAST2,NORTH2
            READ(12,102,END=50)MINEID2, EAST2, NORTH2
          ENDIF
201         FORMAT(I5,I7)
202         FORMAT(I5,X,F12.4,X,F12.4)
            USERID = USERID + 1
50    CONTINUE
C
C
300   CONTINUE
      CLOSE(UNIT=12, STATUS='KEEP',ERR=9999)
      CLOSE(UNIT=13, STATUS='KEEP',ERR=9999)
      CLOSE(UNIT=14, STATUS='KEEP',ERR=9999)
      CLOSE(UNIT=15, STATUS='KEEP',ERR=9999)
C
C
      STOP
99    STOP 'ERROR OPENING'
9999  STOP 'ERROR CLOSING'
      END
```