

# Using ArcGIS/ArcPad to Locate Households Potentially Exposed to Agricultural Pesticides

Justine Allpress, RTI International  
Ross Curry, RTI International

## Abstract

This paper explains the methods used during a government-funded health grant to identify households potentially exposed to agricultural pesticides. The work constraint of this study required the use of aerial photography and demographic data to identify cropfields and surrounding houses that met study criteria. The functionality in ArcGIS to overlay geocoded demographic data on top of aerial photographs was invaluable to study investigators, because the majority of site selection was conducted with limited knowledge of the study area. The success of the study relied upon concise measurements from field edge to surrounding houses, which made it imperative to verify the locational findings. Data were validated using ArcPad with a global positioning system (GPS) receiver to ensure the cropfields and selected households were precisely located. These technologies significantly reduced time spent in the field and eliminated the risk of data transfer inaccuracies between paper and digital maps.

## Introduction

The National Cancer Institute Division of Cancer Control and Population Sciences awarded RTI International (RTI) a grant to study indirect human exposure to agricultural pesticides. This study involved collecting biological and environmental samples to determine the likelihood of pesticide exposures for households close to cropfields. Although various studies have implied that living in close proximity to fields increases the chances of pesticide exposure, the studies have not examined in depth a possible exposure gradient related to distance from these fields. RTI's study attempted to identify such a pattern in the hope of increasing disease prediction rates in populations close to agricultural fields, as well as to determine whether, in fact, an exposure gradient is a function of distance from fields. The study focused on the herbicide atrazine because it is commonly used on at least 75 percent of corn cropfields.

Because of the geographic nature of this project and the need for precise measurement of distances between the selected households and fields, the selection process required a significant geographic information systems (GIS) component. The GIS team sought to develop methodologies and databases that could be used to select corn fields with a substantial number of potential candidate households in close proximity. Households included in the study could not be closer to any fields other than those being investigated. In addition, the study criteria required that an adult and a child aged 4 or younger be living in the home during the study period. To allow for the study of distance the households that met the above criteria were also required to be spread across four buffer zones at different distances from the candidate fields. Various methods were used to meet the specifications of this task, including aerial photograph interpretation, analysis of U.S. Census Bureau data, and the use of ArcPad in the field.

## Identifying Potential Candidate Fields

Prior to conducting any analysis, RTI investigators identified potential field locations within the state of Illinois. During these initial stages, investigators had not chosen specific towns or counties for study inclusion; therefore, they attempted to define a methodology to locate candidate fields across the state. Upon review of aerial photographs, it was ascertained that the most likely locations to find conforming fields would be on the edge of medium to large towns, thus allowing for an appropriate housing density moving away from fields edge. During this process, investigators downloaded and examined 284 Digital Orthophoto Quadrangles (DOQs) within 29 counties.

As noted above, for a field to satisfy the study criteria, no other fields could be located in a proximity close enough to contribute to residential exposures through pesticide drift or windblown soil. To achieve this goal, investigators digitized the field boundary closest to target households of each chosen field in ArcMap. They then created buffers away from the edge of the field at distances of 200, 500, 800, and 1,000 meters, which designated the study area for each of the fields. Once the study area had been defined, investigators identified all potential conflict fields (i.e., other nearby fields that might contribute to the household atrazine exposures) and created 1,000 m buffers away from their closest boundary line.



**Figure 1. Candidate Field Buffers and Conflict Field Resolution**

Using the above method, investigators had the ability to predict how much of the study area would be lost as a result of conflict fields, upon confirmation of the crops being grown in the

fields. At this point in the process, there was no way to determine which crops were being grown in each field; therefore, candidate fields could potentially contain crops other than corn. Moreover, field selection was reliant on aerial photography taken in 1998 and 1999. Although these photographs were relatively current, many of the areas identified for possible study were on the edge of towns where new development was likely. Because such construction could completely change the study area as it had been defined, or remove the selection of a field as a candidate altogether, it was impossible for investigators to predict which fields would actually be eligible for the study without traveling to Illinois to confirm crop rotation patterns and current land use.

## **Cropfield Verification**

Through the process of locating potential fields using aerial photographs, investigators identified 95 cropfields located across Illinois. This number was reduced to 80 to facilitate visits to all of the sites to verify crop growth patterns and current housing development. These 80 fields were then split into 2 groups of 40 to enable field trips covering different areas of the state.

On the initial field trips, investigators used rudimentary techniques to keep track of their findings and to navigate between locations. While visiting the fields, investigators marked the paper maps to indicate any changes in a field's boundary, as well as any important changes to the local landscape. They also photographed field boundaries in the direction of houses and noted potential barriers to pesticide drift.

Although accurate representation of field boundaries was an important part of the initial visits to the study area, the most important task for the investigators was the verification of the type of crop being grown in the field. The field trips took place shortly after the 2002 harvest, so crop residue in the fields was observed to determine which crops had been grown the previous year. Because recruitment of household members to participate in the study was scheduled for the following growing season, RTI's project design required that investigators predict which crop was going to be grown in a field at that time. For a field to be eligible for the study, corn had to be the crop grown during the season when RTI was to collect the biological and environmental samples, because atrazine would be likely be sprayed on such fields. Due to this study requirement, the desired crop residue found on potential fields during the first field trip was soybeans, which would imply that corn would be grown during the following growing season because of crop rotation. Although this method was not foolproof, it was the best methodology aside from contacting farmers directly.

The field selection process was also repeated for all of the conflict fields. This made it possible to ascertain how many houses would be located in the areas of study. If corn residue was found on the conflict field, it was likely that soybeans would be grown the following year, and the field would no longer pose as a conflict exposure for the candidate field.

At the end of the two initial field trips, investigators were left with 34 potential fields. Those removed from consideration either had corn residue on the fields at the time of observation, suggesting that soybeans would be the crop for the following year, or the location was no longer a cropfield.



**Figure 2. Potential Candidate Field with Soybean Residue Present**

## **Acquiring Demographic Data**

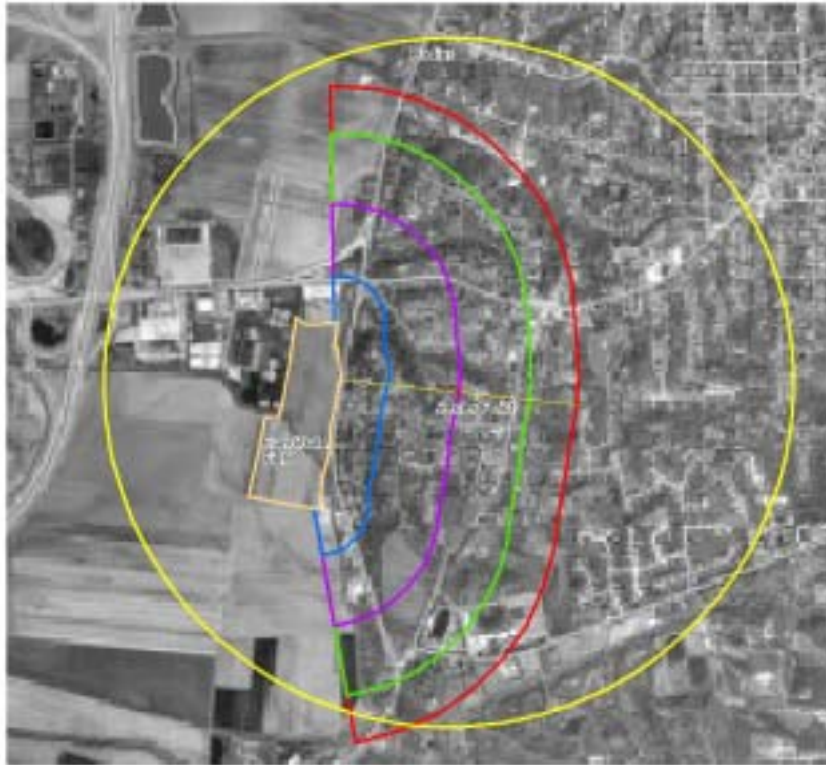
After the field investigators verified that corn was being grown on the candidate fields, new boundaries and buffers were created. These spatial data were used to locate the households meeting participation criteria for the study with regard to their distance from the field boundaries. Specific information was also needed concerning household composition, including occupants' ages, phone numbers, and type of home dwelling (e.g., single family, apartment). Investigators also needed the latitude/longitude coordinates of each house in decimal degrees. To acquire demographic and contact information, investigators contacted marketing list vendors and, after researching several service providers, decided to use a vendor who could provide all addresses fitting the study requirements in a radius from any known address. This method was determined to be the most cost-effective, because investigators could locate an address at the center of the study area and then create a radius to incorporate all the homes within the study buffers. A large part of the radius would include the candidate field, thus reducing the number of unnecessary homes selected when the radius was created.

A consistent method was needed to obtain the addresses from which study radii would be created. To achieve this, investigators located the centroid of the 1,000 m buffer, as well as the half-way point along the field boundary closest to the study area. A straight line was then drawn between the two points, running from the field boundary to the edge of the 1,000 m buffer. The center point of this line was then designated as the center point of the study area.

Once the center point had been located, it was necessary to find an actual address that corresponded as closely as possible to this point. Investigators overlaid a streets layer on top of the map layout they were working on. From here, they chose the closest road and examined the address range on that street. Using the address range as a guide, investigators then selected a fictitious address on the street and attempted to find a corresponding ZIP+4 for the address on the

U.S. Postal Service Web site. When a ZIP+4 was found, the information was sent to the data-provision company, which then returned geocoded addresses with latitude/longitude information for all homes in that specific ZIP+4. Once these addresses were brought in to the layout, investigators identified the closest address to the center point of the study area.

The final step in this process was to determine the size of the radius needed to incorporate all of the potential candidate households in the study area. To create the radius, investigators drew a circle originating at the central address and moved out in increments of one-tenth of a mile, until all of the homes in the study area were contained within the radius. On completion, the central address and radius length were provided to the data vendor.



**Figure 3. Locating the Study Area's Central Address and Radius.**

### **House Location Evaluation**

Using the provided radius and central address location, the data vendor supplied investigators with a list of all homes (with latitude/longitude coordinates) within the circle, which included all homes in the study area. Using these data, investigators identified all the single-family homes that had at least one child aged 4 or younger. For each field, investigators compiled a layer of only those homes that met the study criteria and overlaid this with the aerial photography, buffer zones, and conflict fields. This procedure allowed investigators to determine the approximate number of potential candidate households in the area, as well as how these households were dispersed across the four buffer zones. An RTI statistician determined that the optimum number of participating households in a study area was six, with two in each of the 0 to 200 m and 800 to

1000 m buffers, and one household in each of the other two buffers. From this initial overlay, investigators were able to determine whether any of the fields were unacceptable because of a lack of potential household participants or because of inconsistently spaced households throughout the study area. Some homes were also excluded because of their proximity to conflict fields. At the end of this process, 24 candidate fields remained, located between Rockford and East St. Louis, Illinois.

## **Crop and House Location Verification**

Before recruitment of houses could take place, investigators had to verify that corn was being grown in the candidate fields in early spring and that the geocoded houses were located accurately. The location of the households was important to ensure that they were spaced far enough among the buffers to indicate an exposure gradient.

Verification was achieved through two more field trips to Illinois in the late spring. This timeframe made it possible to identify with certainty the crops that were growing in the potential study fields. The primary task for investigators was to determine that corn was being cultivated in the candidate fields. Once this was achieved, investigators confirmed the borders that had been drawn from the trip the previous fall. If the field had been delineated properly, the next task was to confirm actual house locations.

Prior to the second field trip, investigators loaded aerial photography and layers containing geocoded house locations, field boundaries, buffers, and conflict fields into ArcPad on a handheld computer, with a GPS unit attached. Using a spreadsheet listing all potential house addresses, as well as the aerial photography, investigators navigated to each of the selected households and compared the geocoded point to the actual house location. It was determined during this process that the geocoded points were inaccurate; therefore, each point representing a house had to be linked to a corresponding rooftop on the aerial photograph. Investigators did this by editing the houses shapefile in ArcPad, which enabled real time changes and reduced the chance of inaccuracies during data transfer from paper to digital formats. In several cases, investigators found that this editing process led to houses being moved between buffer zones or even removed from the study area completely. The inaccuracy of the original geocoded houses confirmed that the location verification process was integral to the accuracy of the study's final results.

During this process, 304 houses were visited and their locations verified, and 21 fields remained viable for the study.



**Figure 4: Corn growing in candidate field.**

## **Candidate House Selection**

Once all of the data was collected from the field, it became RTI's task to select the households that would be contacted for possible participation in the study. Investigators achieved this goal by importing the house layers from ArcPad back into ArcMap and displaying these layers over the buffers that had been updated with any changes observed from the field. Upon review of the households that had been verified, investigators could finally see how the houses were spread through the buffer zones for each of the candidate fields. This final location review made it possible to identify any fields that would not be feasible to study due to the dispersion of houses in the study area. It was decided that the optimum situation was one where all of the candidate households participating in the study would be located in a relatively straight line moving away from the field border through the buffers. This was not possible for all fields, but was a policy that RTI investigators attempted to follow. The houses that closest matched this pattern were identified and listed for future contact by recruiters.

Once investigators had identified all of the households selected for contact for recruitment, the investigators' role in the process was almost fulfilled. The final task was to plot all of the homes that accepted the invitation to participate in the study in order to track how the houses were spread through the study areas. This made it possible for investigators to advise the recruiters if further efforts were needed to achieve a complete sample for each field.

## **Summary**

A number of procedures were implemented to meet the requirements of locating candidate fields and homes. Through aerial photograph interpretation, RTI investigators were able to identify potential fields and areas of study, however ground truthing was necessary to confirm these selections. Without visiting potential locations, this portion of the investigation would have been

impossible. Once in the field, the utilization of digital technologies with ArcPad aided in data collection and increased accuracy of field delineation and household location.

Using geocoded points provided by marketing list vendors was integral to the success of the investigators' task. However, the fallibility of geocoding as a means for precise address location was highlighted. Without ground truthing the geocoded results, investigators would have been unable to accurately locate homes that possessed the desired demographic characteristics in relation to candidate fields, which would have significantly compromised the project.

Despite the limitations presented, RTI investigators were able to accurately and efficiently place homes within the study area and locate those which met the criteria imposed by the project. This allowed them to deliver the necessary data to the researchers on the RTI team, leading to successful recruitment of study participants.

### **Author Information**

Justine Allpress  
RTI International  
3040 Cornwallis Rd.  
RTP, NC 27709  
Tel: (919) 485-5688  
E-mail: [jla@rti.org](mailto:jla@rti.org)

Ross Curry  
RTI International  
3040 Cornwallis Rd.  
RTP, NC 27709  
Tel: (919) 541-6239  
E-mail: [rjc@rti.org](mailto:rjc@rti.org)