# DISCOVERY in Hamakua:

## Epidemiological Information Collection in an Austere Environment

**Eric Rasmussen, MD, FACP**
Principal Investigator, TIDES, DARPA

**Objective:**

> To test a model for the sustained and iterative collection of epidemiological information from austere environments using local inhabitants, PDA-based software, and remote analysis into the TIDES data stream. The project was termed "Discovery", and funding was provided through the TIDES program within the Defense Advanced Research Projects Agency.

**Design:**

> Based on experience in Africa, the Middle East, and the Balkans, a domestic trial was designed to test the use of Personal Digital Assistants (PDAs) for sustained epidemiological collection in a remote and difficult region of the United States.  This trial was specifically to test the possible pathways for effective collection using moderately technical options. It was <u>not</u>, in this phase, designed to validate either the collection elements or the epidemiological status of a particular location. The focus was rather the optimizing of collection methods, data movement, and analytical possibilities. The requirement for such a capability, and so the trial, was confirmed by the UNICEF Country Coordinator in Uganda during personal conversations on-site in Kampala during the 2000-2001 Ebola outbreak.

> With assistance from United Nations relief agency staff members in Kenya, Uganda, Zambia, Switzerland, Italy, and the United States, desirable criteria were developed for collection in African village environments. After development of those criteria, former Global Epidemiologic Surveillance staff members from the World Health Organization

(WHO), now with the Centers for Disease Control in Atlanta (CDC), assisted in the design of a collection template for ongoing, baseline surveillance, striving to identify human and environmental trends before serious problems developed. The resulting template was discussed at length with collaborating centers and a working deployment draft was developed.

In order to reflect a reasonable approximation of the working environments in Africa, we chose a domestic, remote, tropical, agrarian site for the trial.  With assistance from local institutional representatives on site, we elected to pursue weekly epidemiological surveillance of remote Hawaiian sugarcane plantation camps along the northeastern Hamakua Coast of the Big Island of Hawaii.

Once both the site for the trial and a collection template were agreed upon, Palm OS devices (Handspring Visors) were loaded with custom software designed on the CDC template, a teaching curriculum was devised, and representatives were chosen from Hamakua camp volunteers for the weekly accrual of information from eight locations.

After weekly collections within the camps, the PDAs were synchronized at a central site well-known to all participants. The consolidated files were then submitted electronically to a server located at Syracuse University in New York for analysis. The consequent information was to be visualized into culturally appropriate images and returned to the camp representatives for discussion and decisions as they chose.

**Setting:**

The northeast Hamakua Coast of the Big Island of Hawaii is a spectacular natural setting, but has a population plagued by poverty, illiteracy, abuse, malnutrition, and violence. For fifty years before 1992, many of the slopes of the mountains leading to the sea were under sugar cane cultivation, and most of the cane fields were owned by the Hamakua Sugar Company. Until 1992 Hamakua Sugar was, by far, the dominant employer in the area, but the price competition from the Dominican Republic, Brazil, India, and elsewhere caused the cessation of all sugarcane cultivation on the Big Island by 1994. By late in that year all cane plantation employees had been dismissed.

As a condition of the corporate dissolution, structures used as homes by the camp employees were sold to the inhabitants at very low cost. Since the camps had been forcibly organized by ethnicity through the concerted actions of plantation management, the result, after the transfer of home ownership, was a collection of remote and remain quite ethnically distinct villages linked by a single road, police station, small post offices, school, two rural banks, and a credit union. The eight camps that still exist, therefore, have a small, common infrastructure and a similar environmental exposure history, but they are (in great part) culturally, ethnically and genetically distinct.

After the closure of Hamakua Sugar, each camp suffered high unemployment as the working-age population shifted from the lost plantation jobs into service-sector positions, predominantly within the hotels of the Kona Coast roughly 35 miles away over very narrow mountain roads. That service-based employment proved unstable as the economies of the Pacific Rim, particularly Japan and the United States, fluctuated, and unemployment has again become significant with the reduction of tourism and airline flights after 9/11. The result has been an increase in domestic abuse, especially in children and the elderly, illegal drug manufacturing, school dropout rates, and alcoholism.

It is important to note that the camps are essentially independent rural villages, but that collectively they have a fairly coherent epidemiological profile. From a medical perspective, physicians from the Medical Clinic in Honoka'a report the following from the set of camps:

- 7-fold greater incidence of pancreatic cancer over similar population baselines
- 5-fold greater incidence of Lou Gehrig's Disease (amyotrophic lateral sclerosis)
- 5-fold greater incidence of Lupus (Systemic Lupus Erythematosis, SLE)
- An apparent increase in congenital neurological disorders, some poorly characterized but predominantly behavioral
- Toxic exposures, both acute and chronic, predominantly organophosphate and chlorinated hydrocarbon pesticides. Alcoholism is also endemic.
- Poor access to medical care on the Western, biomedical model

These medical issues have never been studied and are only barely recognized as significant within the local population. As a result these camps seemed appropriate for a close and ongoing epidemiological assessment, beginning with this trial.

**Methods:**

**Surveillance Questions:**
The questions for collection were selected by local participants from a template of recommended questions composed by the CDC. Those questions comprised a set for both initial, "stable" data (location, water sources, baseline population, crops under cultivation, dominant employment, physical and social infrastructure, and the like) and the ongoing collection of "variable" data. The variable data included household illnesses, household employment that week, weather, births (normal and not), deaths (expected and not), arrests, children in school this week, available toilets, available water, areas unexpectedly not under cultivation, food shipment arrivals and more, with a number of medical variables for disease surveillance that were pursued only within the more sophisticated sites.

**Software Development:**
The final question template was sent to the developers of the Carebridge telemedicine system used within the British Royal Navy in the United Kingdom. Those programmers coded the template into the C programming language, optimized for the Palm OS on Handspring Visor PDAs. Each question had an associated "help" statement that explained the intent of the question by simply tapping on the letters of the question on the screen.  Most questions could then be answered very briefly with one or two digits.

**Submission sites and volunteer staff:**
A single volunteer was chosen as a project director on-site. That person (GM) was a lifelong member of the camps, was an ethnic native with a known and respected genealogy, spoke each of the associated languages, served on several local councils, had successfully raised children within the area, was the local Red Cross Disaster Services director, and was the president and CEO of the credit union that provided virtually all financial services for the camps, including automobile and educational loans, home refinancing, and funeral expenses. GM is also the chairman of the Hawaii Community Loan Fund, a micro-credit and small-business enterprise loan group.

That coordinator arranged for meetings within each of the prospective camps. Volunteers from the camps were then chosen to serve as collectors and separate

meetings were held to train those camp volunteers on the goals, the process, the PDAs, and the software. Virtually all of the camps between Hilo and Waipio elected to participate and volunteers were initially plentiful.

There were eight camp submission sites, the largest of which was Pa'auilo with 155 participating households. The other camps varied from roughly 40 up to about 95 households. Surveillance questions were written for households, not individuals, and no personally-identified information was ever collected.

For the collection process, each camp had a representative who, in turn, chose camp inhabitants for door-to-door surveillance each week. Participation at all levels was voluntary, but small incentives were offered for consistent participation throughout the six-month trial.

**Consolidation and analysis:**

The camp volunteers collected the data on the PDA, then brought the PDA to the consolidation point each week, usually on Monday morning after a weekend collection. The consolidation point was a dedicated laptop computer within the credit union. Each PDA was synchronized to the laptop through a unique ID, and then that set of files was consolidated into a single file and sent to a server at Syracuse University for analysis.

**Materials:**

1. Handspring Visor Palm OS PDAs with unique IDs
2. Discovery software suite on each PDA (representative screens in appendix 1)
   a. Stable data elements. **Parentheses** indicate the number of items to be completed within each category:
      i. Categories: 12, to be completed only once
         1. Document Information (3)
         2. Installation (10)
         3. Population (16)
         4. Shelters (7)
         5. Crops (14)
         6. Water (9)
         7. Power (5)

8. Industries (10)
9. Communication (15)
10. Government (4)
11. Religion (8)
12. Ethnic (9)
   b. Variable data elements
      i. Categories: 13, collected weekly
         1. Document information (3)
         2. Installation (1)
         3. Population (8)
         4. Households (5)
         5. Sanitation (2)
         6. Obstetrics (6)
         7. Health (52 – but optional)
         8. Food (9)
         9. Water (2)
         10. Power (5)
         11. Land use (9)
         12. Special events (15)
         13. Reporter note (1)

3. Laptop computer running Windows 2000 and both a PDA and ISP link
4. Dial-up ISP account
5. Paper-based collection sheets for variable data (appendix 2)
6. Analytical software on a Syracuse University web server
7. Perceptualization software on servers in San Diego
8. Educational presentations developed for the collection volunteers

Between the Stable and Variable programs there are a total of 138 data elements, including items contained in drop-down lists of religions, crops, educational levels, and the like. The number of actual questions within the Variable set each week was considerably smaller, roughly 60 elements, with some co-dependent variables.

**Main Outcome Measures:**

The desirable outcome was a consistent and comprehensive submission from each camp weekly that would be analyzed, visualized in an intelligible manner, and returned to the camp leadership in ways that would be useful for their understanding of the camp environment.

So:

1. Weekly interviewing of each participating household
2. Completion of the surveillance template for those households
3. Submission to the consolidation site
4. Submission of the consolidated file to the analysis site
5. Analysis of the submitted files
6. Visualization of the analysis in ways meaningful to the camp leadership
7. Return of the visualization to that local leadership
8. Iteration of the questions (based on the analysis) to accommodate findings, then restart the sequence of collection, submission, and analysis.

**Results:**

Between 01 August 2001 and 15 December 2001, the duration of the first phase of the trial, there were 54 consolidated camp submissions. The approximated time-sequence plot is included below.

Regrettably, the analysis and feedback loop – a critically important part of the sustainability of this project - was almost immediately re-directed into an analysis and visualization of Central Asian information on the evening on 9/11, roughly a month after the inception of the trial. That information, collected within Afghanistan, Pakistan, India, and the Middle East, was arriving through a separate project, had immediate National Security implications, and clearly (though regrettably) took precedence over the Hamakua Discovery project already underway. Unfortunately, the analysis and visualization programming team was difficult to split (it is three people) and the collection process therefore progressed in Hamakua from 9/12/01 forward with no incentive and no feedback to the collectors.

Despite that significant impediment, the table below reflects a persistent attempt by the Hamakua participants to collect information reliably despite having virtually no reciprocal encouragement. The first column is the file submissions (52 of the 54 sent are included; 2 were damaged in transmission and were unreadable) and the second column is the number of data elements submitted as changes from baseline.

Several of the variations can be traced to recognizable events within the camps. Some of the larger data point counts follow immediately after a meeting with the project director. Some of the smaller counts are associated with extreme weather and exceptional local occurrences (e.g. important sports events, funerals, religious festivals, and so forth).

## Data File submissions: 01 Aug – 31 Dec 2001

| File sequence | File submitted | Data points |
|---|---|---|
| 1 | B7EF85C1.xml | 1 |
| 2 | B7DDD988.xml | 5 |
| 3 | B7AA68B0.xml | 12 |
| 4 | B871BF05.xml | 0 |
| 5 | B7DDDB94.xml | 13 |
| 6 | B7B014E7.xml | 9 |
| 7 | B789BCB8.xml | 26 |
| 8 | B789BDAB.xml | 27 |
| 9 | B789BBE9.xml | 0 |
| 10 | B7C28BFD.xml | 11 |
| 11 | B789BBFB.xml | 26 |
| 12 | B802FCC9.xml | 5 |
| 13 | B7B014EC.xml | 24 |
| 14 | B802FD5A.xml | 9 |
| 15 | B80BF6E6.xml | 4 |
| 16 | B7998B7B.xml | 23 |
| 17 | B802FB39.xml | 0 |
| 18 | B7EF8604.xml | 1 |
| 19 | B7A674A3.xml | 19 |
| 20 | B789BE5A.xml | 0 |
| 21 | B789BB77.xml | 26 |
| 22 | B789BE4A.xml | 26 |
| 23 | B7A67E05.xml | 8 |
| 24 | B78B41CF.xml | 0 |
| 25 | B80BF785.xml | 9 |
| 26 | B789B9FA.xml | 0 |
| 27 | B78ACD36.xml | 0 |
| 28 | B875844D.xml | 7 |
| 29 | B789BD97.xml | 0 |

| 30 | B8758491.xml | 3 |
|---|---|---|
| 31 | B789BCC9.xml | 0 |
| 32 | B7A57D33.xml | 9 |
| 33 | B7A57DFE.xml | 8 |
| 34 | B7EF8510.xml | 1 |
| 35 | B7DDDAA9.xml | 7 |
| 36 | B80BF97B.xml | 1 |
| 37 | B78B2249.xml | 23 |
| 38 | B871BEBE.xml | 2 |
| 39 | B81BDB11.xml | 0 |
| 40 | B80BF93C.xml | 3 |
| 41 | B7B0157E.xml | 8 |
| 42 | B802FB83.xml | 2 |
| 43 | B79324C8.xml | 47 |
| 44 | B789BD83.xml | 26 |
| 45 | B7A67DA9.xml | 24 |
| 46 | B79326E6.xml | 12 |
| 47 | B7C28D3B.xml | 9 |
| 48 | B789BDBB.xml | 0 |
| 49 | B7B07426.xml | 0 |
| 50 | B7DDDA2D.xml | 6 |
| 51 | B78B41B2.xml | 0 |
| 52 | B7A98677.xml | 47 |

**Discussion:**

In our opinion there are several lessons that can be learned from this modest trial.

**Surveillance questions:**

The questions that were developed with the CDC were useful for initial (Stable) environmental demographics, but since the Variable data did not appreciably change from week to week, it became difficult to sustain the interest of the collectors. One critical lesson from this for us was the value to be had in the presence of a local and trusted champion. The communities, despite their hospitality, interest, and apparent good will, are very insular societies where small variations from the normal pattern of the days can have a significant impact on the comfort of the population. Strangers, even bearing gifts and a chance for the camps to improve their understanding of their environment, are welcome only to a point, apparently discussed endlessly throughout the community, and trusted not at all. Ready conclusions are that this sort of program has very little chance of accurate collection unless there is either:

1. a supportive member of the community who can both drive the internal host-population collection process and manage the cultural and political tribulations caused by the outsiders, or

2. a collection process that uses objective and disinterested outsiders for the repetitive collection process and leverages fear or significant rewards (e.g. disease containment or food provisioning) in order to accrue accurate information from the local population. We note that data validation will be challenging and that priorities and agendas between questioned and questioner may be quite different.

The collectors also noted that the alteration of the questions to accommodate new requirements was disappointingly cumbersome. The creation of the software for the Palm OS led to a degree of inflexibility that restricted our ability to respond with alacrity to environmental alterations. An example was the upsurge in local use and production of smokable D-methamphetamine (Ice) after a cascade of 9/11-related misfortune led to increased unemployment and a general rise in illegal activity. Because of the time constraints in re-coding the Palm software,

we had no ability to follow that specific topic even though it had a significant impact on the communities under study.

I suspect we would have been far better served had we dispensed with the Palm technology and used a simple paper form wedded to a flexible relational database, then scanned the form in order to accrue data.  The PDAs, by report of those working with the communities, were intimidating and problematic for those with little prior experience with electronics. Paradoxically, the PDAs were popular, and we reportedly had more volunteers because of the appeal of the gadget, but the collectors were apparently neither comfortable nor competent when faced with an unexpected problem.

Related to the inflexibility of the questions was the nature of the collection process. The collection volunteers were usually well-known community leaders who then had small numbers of community assistants actually walking door-to-door, analogous to the original concept of a small and remote village. However, those door-to-door collectors were often chosen because they knew everyone and so the households might well have been a bit reluctant in revealing embarrassing or otherwise potentially compromising information. We saw no satisfactory solution to the paradox of

1. requiring the collectors be familiar enough to the community that they were trusted by the households at all, yet
2. at the same time having the collectors learning so much about their neighbors that neither is comfortable and the collection process is hindered.

**Software development:**

The software was developed using CodeWarrior Palm OS development tools, with the reported generation of 230,000 lines of code to create a 13kb file. That is not a misprint. Of surprising importance, the coding of the questions was done with initial attention to the simplicity of the screen formatting, and less to the mechanism for the choice selection. Those ergonomics seem related, but they are not the same and it made a difference.

In retrospect there were two major flaws in the technical design:

1. The Palm OS screen design for collection, while quite careful in several ways, was limited by the speed at which data could be entered. The Palm OS recognition system (Graffiti) is not intuitive and errors are common. Though we had data validation structures in place, the act of writing on a one-inch square screen, having the choice rejected, then correcting and writing again, was frustrating. Drop-down menus offering choices that the collectors designed would probably have improved the data yield significantly by reducing the time requirement (and so increasing the throughput), while improving the reliability of the data element selection. However, it would have restricted the range of possible answers and further reduced the flexibility of the collection process – a problem mentioned above.

2. Noted above, as we now re-assess the core task to be performed, we would have been better served with much less technology. The software coding process was so laborious that, as noted above, the questions could not be readily adapted to changing circumstances.  We fully recognized that the initial template was likely to be inadequate and that changes would be required, but during the initial training sessions we briefed to the collectors a level of responsiveness that proved unsustainable because of the complexity of the software development. Unfortunately none of us involved in the project know any more flexible Palm OS coding structures, and the right answer is probably a far simpler method for collection using a relational database program, automatic form generation, paper-based collection, and then scanning the completed forms into electronic form for analysis. Such a method would have been far simpler, with a more robust and mature suite of tools, offering greater flexibility, more familiar acquisition methods for an unsophisticated audience, and very probably a greater yield of more reliable information.

However, it is important to recognize that we were specifically testing a technological approach for feasibility. The system as designed and implemented was expected to be unfamiliar and one of our goals was an assessment of the integration of such tools into a relatively naïve population using a supportive educational process and local champions.

That aspect of the study was informative, but did <u>not</u> sustain the use of PDAs for data collection <u>by local inhabitants</u> within in naïve populations.

Future efforts will use simpler tools for collection and allow on-site data element alterations by the local project director. Those alterations will respond to feedback from the community on issues of interest. The complex inter-relationships between the communities and their environment mandates an ability to follow "threads" as one might in a conversation. Conventional data collection methods are not designed along those lines, but we saw many alterations in community dynamics (as a result of slowly unfolding circumstances) that were worth tracking in such a fashion. For this study we lacked the adaptability we needed, but that lesson has been learned.

**Volunteer staff:**

Volunteers within a native population, while an available source of manpower in many areas, probably need ongoing motivation through some form of incentive when performing routine collection and we saw that during this study.  In many cases that ongoing incentive can be the providing of information useful to the community.  We suggest that <u>iterative feedback to the collecting communities is critical</u>, incorporated at each level so that some reason for a sustained effort is apparent to the community and that the community clearly gains from their effort.

**Consolidation:**

The process of data consolidation was relatively good and quite effective, given the constraints of the communications infrastructure in this remote area. A dedicated laptop was available for synching with each of the Visor PDAs at any point during working hours and the geographically central location within a key business was helpful in ensuring the collections took place. From that site a file was automatically generated, then emailed as an attachment to the analysis center. It is difficult to see where any improvements could have been made at this level of the study.

**Analysis and Visualization:**

The analysis of the information was not done in the dedicated fashion originally envisaged. The initial collections were under early evaluation for content material,

content representation, and then visualization of that representation over time in ways that made sense at several levels of interest. Then September 11th altered everything.

1. The Principle Investigator for this study is a Navy physician on active duty and was assigned to an operational billet on very short notice.
2. The analysis and visualization team (three people) is contracted to several Government agencies and was quickly pulled for full-time work that enlarged our understanding of the developing effort in Central Asia.
3. The software development team is a group within the British Royal Navy and they, too, were immediately and completely redirected into a focus on Central Asia

On the more positive side, the result of our concerted effort toward 9/11 response was the development of a more general visualization tool that has a great deal more capacity than originally designed for this small study. The analytical-visualization software we've collectively developed will be released to the public domain in the summer of 2002, and will be used to further assess the data gathered within this Hamakua study.

Further epidemiologic study within these communities  is under development as a second phase of this work.  It will be coordinated through The Kohala Center in Hamakua, an academic institute whose affiliations include the University of California, Cornell University, and the University of Minnesota. That second phase will build upon lessons learned in this study, evaluating the long-term health effects of low-level industrial chemical exposure on underserved Pacific Rim populations.


++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

END