

UNC HEALTH REGISTRY/ CANCER SURVIVORSHIP COHORT

Data Source Documentation

8/28/2013



UNC

**LINEBERGER COMPREHENSIVE
CANCER CENTER**

N.C. CANCER HOSPITAL

**University Cancer
Research Fund**

UNC Health Registry Data Source Documentation

Table of Contents

- i. Overview of The UNC Health Registry/Cancer Survivorship Cohort (HR/CSC) 2**
- ii. HR/CSC Data Collection Overview 3**
- Section 1: Cohort Eligibility and Enrollment 5**
- Section 2: Biospecimens 6**
 - Overview 6
 - Collection and Handling of Blood Samples 6
 - Collection and Handling of Tissue 6
- Section 3: Questionnaires 7**
 - Overview 7
 - Baseline Interview 7
 - General Questionnaire 7
 - Cancer-Specific Questionnaires 8
 - Annual Follow-up Interview 8
- Section 4: Medical Records Abstraction 9**
 - Overview 9
 - Hospital Tumor Registrar Abstraction of Medical Records 9
 - Augmentation of Medical Record Abstraction by HR/CSC Staff 9
 - Billing Code “Initial Diagnosis” for Cancer-Specific Questionnaire Assignment 10
- Section 5: Cohort/Patient Tracking System 11**
- Appendix: System-Specific Information 13**
 - Enrollment Tablet Software (Mi-Forms) 13
 - MI-Forms Implementation with HR/CSC: Frequently Asked Questions 13
 - MI-Forms Product Description 13
 - Overview of Biospecimen Process Facility Laboratory Information Management System (LIMS)
 - Quality Assurance / Quality Control 17
 - Biospecimen Process Facility Information Technologies (LIMS) 19
 - Future Tissue Procurement Facility Laboratory Information Management System (TPF LIMS) ... 20
 - Medical Record Abstraction: Electronic Registry System Software (ERS) 20
 - Participant Tracking System 22

Preparation Date: August 28, 2013

i. Overview of The UNC Health Registry/Cancer Survivorship Cohort (HR/CSC)

In 2007, the NC legislature funded the University Cancer Research Fund (UCRF - ucrf.unc.edu), to be used for cancer research at the University of North Carolina Hospitals (www.unchealthcare.org) and the UNC Lineberger Comprehensive Cancer Center (LCCC - unclineberger.org). The mission of the UCRF is to save lives and reduce suffering from cancer, and one of its first priorities was to establish The UNC Health Registry/Cancer Survivorship Cohort (HR/CSC) to aid research into cancer survivorship.

The HR/CSC is a hospital-based registry of cancer survivors who have agreed to the use of their data and specimen for research. This study integrates clinical, epidemiological and interview data with repositories of biologic specimens including DNA and tumor tissue. A global informed consent is implemented which covers all aspects of participation in the HR/CSC. HR/CSC participants agree to allow medical record abstraction, procurement and banking of biospecimens (e.g. tumor tissue and blood), research use of all specimens collected as a part of their medical care, and completion of questionnaire interviews at baseline and at annual follow-up, as well as re-contact for future studies. All study procedures, data collection instruments, and patient materials have been UNC Institutional Review Board approved.

ii. HR/CSC Data Collection Overview

The HR/CSC collects different pieces of information on participants at different phases of their enrollment. Figure 1 provides a graphical representation of data collection events. Key highlights of the process include:

- Prior to providing consent, select demographic data elements from clinical scheduling systems are screened to determine eligibility. This information is used to determine which patients should be approached for enrollment into the HR/CSC.
- Once approached in the clinic and informed of the cohort, patients choose to provide informed consent. Those who consent are then asked to confirm their contact information and answer a few brief questions.
- Ideally, a blood sample is drawn by a UNC Hospitals phlebotomist at the time of the enrollment visit. Blood specimens are delivered by HR/CSC staff to the UNC Biospecimen Process Facility (BSP) for processing and banking. Blood plasma, serum and DNA samples are prepared and banked.
- If a HR/CCS participant undergoes surgery at UNC Hospitals, leftover tumor tissue is banked by the UNC Lineberger Tumor Procurement Facility (TPF). The TPF processes surgical tissue specimens to snap frozen and formalin fixed and paraffin embedded (FFPE) tissue samples for research.
- Following the enrollment visit, a baseline questionnaire is administered that collects both general and cancer-specific information on health history, lifestyle, patient reported outcomes, and health care system interaction.
- Abstracted medical record information collected by the UNC Hospital Tumor Registry to comply with state cancer reporting requirements is shared with the HR/CSC. Cancer diagnosis, tumor characteristics, and general treatment information are available through this resource. For HR/CSC participants, HR/CSC medical records abstractors augment the Hospital Tumor Registry data with specific details not collected by the Hospital Tumor Registry (e.g., chemotherapy drug and dose).

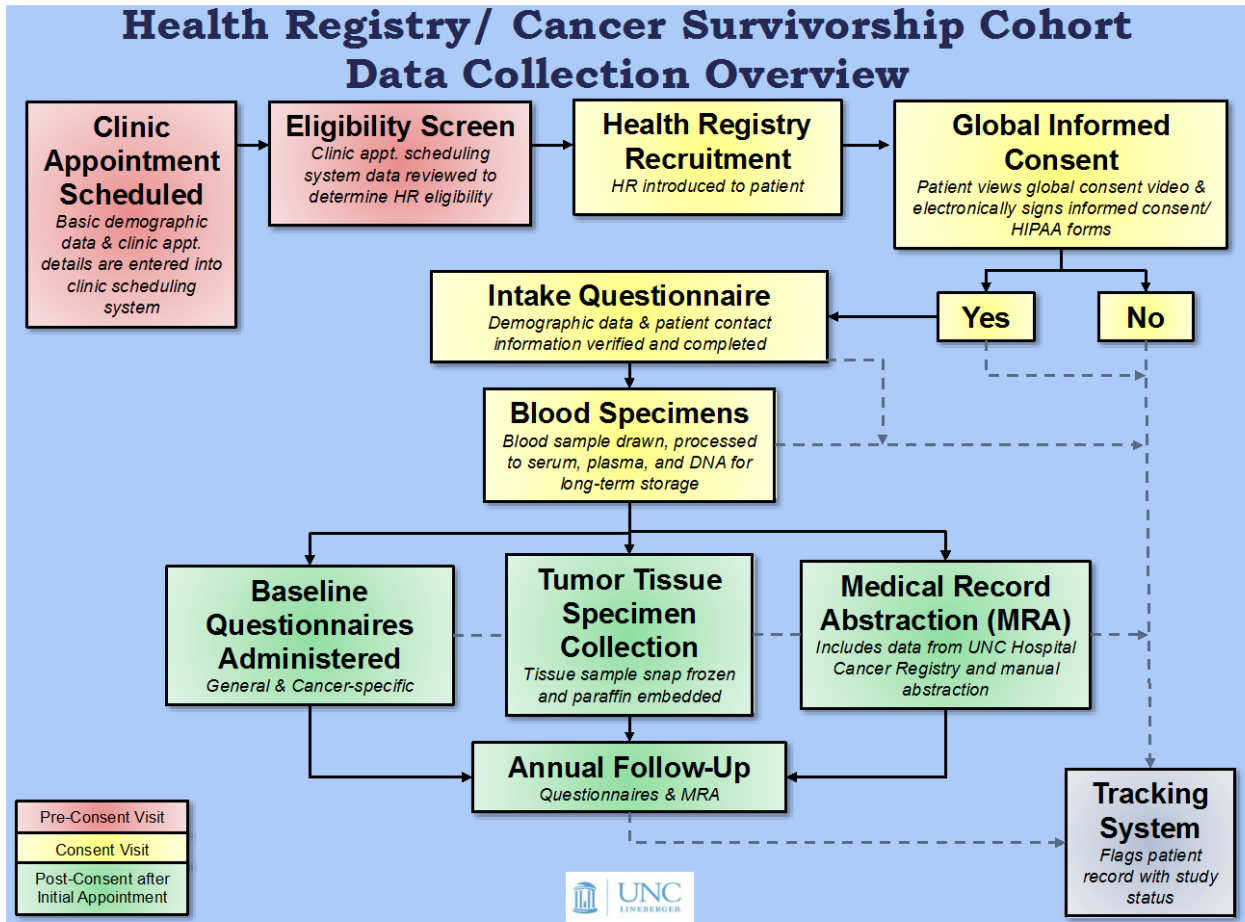


Figure 1 – UNC HR/CSC Stages of Data Collection

Section 1: Cohort Eligibility and Enrollment

The enrollment process begins with a list of eligible patients scheduled for appointments in cancer clinics at UNC Hospitals. The cohort is open to individuals 18 years or older who provide a North Carolina address. Clinic schedule data is obtained daily from the UNC Hospitals appointment scheduling systems.

The HR/CSC specific daily clinic schedule report is populated with a list of individuals who determined to be eligible based on date of birth and residence. These daily schedule reports are used by the HR/CSC recruiters to identify patients in the clinic who may be approached for enrollment in the HR/CSC.

Approached patients are shown a brief consent video. This video summarizes the purpose of the HR/CSC, and explains what is involved in participation. HR/CSC recruiters are available for questions that may arise.

After viewing the video, patients who desire to participate in the HR/CSC provide signatures electronically on both the HIPAA and the informed consent. Following consent, participants are asked to confirm their contact information and that of their referring physician, as well as, provide basic demographics information.

All data and events of HR/CSC participation are recorded electronically on password-protected tablets. Portable tablets used in the clinic for enrollment are connected to a secure wireless network, with data sent to a database and fileserver over an encrypted network connection.

Section 2: Biospecimens

Overview

As part of consent, participants agree to blood collection and banking of leftover tumor tissue should they have surgery at UNC Hospitals. The Biospecimen Processing (BSP) and Tissue Procurement (TPF) Facilities are partners of the UNC Health Registry/Cancer Survivorship Cohort and are tasked with responsibilities including collection, processing, and banking of HR/CSC participant biospecimens. A summary of biospecimen collection and handling is provided below.

For detailed information on the HR/CSC protocol for collection, processing, and storage of biospecimens (blood and tissue), please consult the UNC Health Registry/ Cancer Survivorship Cohort Biospecimens Collection, Processing, and Storage Protocol.

Collection and Handling of Blood Samples

At or after enrollment, participants who do not have portacaths ('ports') in place are provided with a bag containing unique barcode labeled blood tubes. This barcode label is linked to the participant and is used to track and annotate the blood specimen products. Participants are instructed to provide the tubes to the UNC Hospital phlebotomist at the time of blood draw. Once drawn, the blood tubes are handled by HR/CSC staff who ensure proper logging (including time of collection), handling, temporary storage and subsequent transport (within 2 hours of collection) to the **Biospecimen Processing Facility (BSP)** for further processing, aliquoting, and banking. These events, like those in enrollment, are recorded by HR/CSC staff in the electronic Patient Tracking System available on portable tablets in the clinic.

Once participant blood samples are transported to the UNC Biospecimen Process Facility (BSP), they are logged into the BSP Laboratory Information Management System (LIMS). BSP technicians then begin processing, aliquoting, and banking samples. The laboratory begins sample processing within 30 minutes of receipt. All protocol steps are predetermined and proscribed by the BSP LIMS system, which provides an interface to input measurements, record notes, and to assign sample container identification numbers. Samples are banked in freezers in two separate locations to provide a fail-safe against equipment failure.

Collection and Handling of Tissue

Nightly HR/CSC participants on the next day's UNC operating room schedule are identified and flagged by the **Tissue Procurement Facility (TPF)** and targeted for tissue collection. Following surgery, the surgical pathologist determines if leftover tissue is available for banking. If leftover tissue is available, the TPF technician is notified and retrieves the specimen from the surgical pathologist. Leftover tissue is taken to the Tissue Procurement Facility where it is processed and banked

Section 3: Questionnaires

Overview

Questionnaires may be administered at several different times during participation. The first time is at the enrollment visit when a brief intake interview is administered. The second is after enrollment at the time of baseline interviews (both general and cancer specific) and the third is at annual follow-up (annual follow-up has not been implemented in the HR/CSC at this time) (see Figure 1: Stages of data collection).

At enrollment, the intake questionnaire allows confirmation and collection of demographics data onto tablet computers. Demographics data include name, date of birth, contact information, marital, employment and education status, and referring physician name and address.

Baseline interviews are typically administered by trained HR/CSC interviewers within 2 weeks of the participant's enrollment visit. The baseline interview contains both general and cancer-specific questionnaires. Interviewers use a computer assisted telephone interview (CATI) system specifically developed for direct electronic data entry with built in quality control. While interviewers usually telephone participants to administer the interview, the CATI system can also be used to administered in-person interviews at the time of a participant's clinic appointment. The baseline interview (general and cancer-specific) lasts between 45 minutes to an hour and can be completed over more than one call if needed. Questionnaire data is recorded via computer using the CATI system known as CDART – Carolina Data Acquisition and Reporting Tool and stored in the study-maintained database.

Standardized, reliable, broadly used questionnaires were selected for the baseline interview. These questionnaires have been validated in the literature and are available in English and Spanish.

Baseline Interview

The baseline interview consists of both general and cancer-specific questionnaires. During this interview, each participant is asked a standard set of questions from the general questionnaire. Cancer-specific interview forms are selected for each participant based on indications. The following sections contain more information about each portion of the baseline interview.

General Questionnaire

The baseline general questionnaire consists of more than 20 constituent questionnaires measuring health, symptoms, quality of life, and related variables. A few of the questionnaires contain skip patterns that enable or disable further questions in a sequence depending on outcome. Examples include information on rate of alcohol consumption that is dependent on answering a lifetime alcohol consumption question on the alcohol information form as well as age-based skip patterns that are utilized in the historic height and weight form (questions refer to height and weight at ages 40, 50, and 60).

Cancer-Specific Questionnaires

With input from clinicians, a series of cancer-specific questionnaires were developed to capture data on specific disease-related points of interest and are included in the baseline questionnaire. Example areas include Breast, Prostate, Colon, and Hepatobiliary cancers. Assignment of cancer-specific questionnaires occurs through initial diagnosis assignment based on ICD-9 billing-code classification. The results of this classification (initial diagnosis) appear to interviewers in the HR/CSC Patient Tracking System to allow interviewers to select the appropriate cancer-specific questionnaire to administer to the patient.

Annual Follow-up Interview

After completion of the baseline questionnaire and applicable cancer-specific questionnaire forms, an annual follow-up interview serves to capture changes in this information, as well as numerous cancer outcomes including any subsequent treatment as the result of cancer recurrence. Annual follow-up of participants in the HR/CSC has not been implemented at this time.

Section 4: Medical Records Abstraction

Overview

Medical records information is gathered from electronic medical records at UNC Hospitals. Two mechanisms are used for this: 1) abstraction of medical records by the Hospital Tumor Registry are exported to the HR/CSC and 2) medical records data from the Hospital Tumor Registry is augmented for approximately 10-40 data elements primarily related to treatment details (e.g., chemotherapy drug and dose) that are abstracted from the UNC medical records by trained HR/CSC medical records abstractors. Additionally, as noted in the cancer-specific baseline questionnaire section, to expedite the assignment of cancer-specific questionnaires shortly after participant enrollment, ICD-9 billing code data from UNC Hospitals' Carolina Data Warehouse for Health (CDW-H) assigns the participant an initial diagnosis.

Hospital Tumor Registrar Abstraction of Medical Records

In order to fulfill state cancer-reporting requirements, the UNC Hospital's Tumor Registry abstracts information on diagnosis and first course of treatment from the medical records for all patients diagnosed and/or treated at UNC Hospitals. Medical records are examined and data abstracted from the medical record and manually entered into an electronic registry system by certified tumor registrars (CTRs). Categories of abstracted medical record data include: tumor diagnosis and staging information, biomarkers/tumor markers and treatment information. Please consult the HR/CSC data dictionary for more specific detail on abstracted fields. Completion of abstraction generally occurs between six to twelve months after a patients' initial diagnosis in order to capture full information on the first course of treatment.

The UNC Hospital Tumor Registrar utilizes a software product called ERS (Electronic Registry Systems, Cincinnati, OH) to collect medical records data. Following the entry of a patient's initial diagnosis, ERS software tailors the abstraction fields in a cancer-specific manner thus improving data quality. Furthermore as diagnostic, tumor staging, and course of treatment information coding are updated nationally, the ERS is updated automatically.

Augmentation of Medical Record Abstraction by HR/CSC Staff

The UNC HR/CSC purchased ERS, which affords the electronic export of data from the UNC Hospital Tumor Registry to the HR/CSC ERS. The ERS also has an additional feature, known as 'free fields' that can be customized for study-specific needs. In the case of the HR/CSC, 40 additional free fields have been customized for collection of additional treatment specific data (e.g., chemotherapy drug and dose). HR/CSC medical records abstractors abstract free field data into ERS after HR/CSC electronic receipt of the participant's data from the UNC Hospital Tumor Registry. Free field data include: hormone, immunotherapy, and chemotherapy agents involved in the first course of therapy, and best estimate of treatment start date based on dispensing/administration records within visit notes. Each agent's NSC (NCI National Service Center) number is recorded. Often, courses of treatment consist of regimens. For a particular drug regimen, multiple drugs may be administered, each with its own NSC number. Each drug in the regimen is recorded. Treatment start and stop dates are also recorded in ERS free fields along with treatment

completion status (e.g., treatment not completed as planned due to patient health/complications).

Full information on the scope of medical records is found in the UNC HR/CSC Medical Records Abstraction Protocol.

Billing Code “Initial Diagnosis” for Cancer-Specific Questionnaire Assignment

Billing codes provide an “initial” diagnosis for the purposes of the assignment of cancer-specific questionnaires; this step is necessary until pathologic data can be quickly extracted electronically (thereby providing a final diagnosis). In order to implement billing code-based initial diagnosis, a rule-based algorithm was developed to assign questionnaires based on ICD-9 billing codes around the time of enrollment.

Source data originates from UNC Hospitals billing system and is exported to the Lineberger Data Warehouse and Biospecimen Repository (LDBR). Reports contain medical record number (MRN) corresponding to each subject, the patient’s date of visit, and ICD-9 code.

This data is used in conjunction with the participant’s date of HR/CSC consent to select the billing records corresponding to the visit closest to the date of consent (this is often, but not always a visit on the subject’s date of consent). Billing codes from the selected date for each patient are matched against a questionnaire-mapping table. The mapping table identifies specific ICD-9 codes and their corresponding questionnaire.

Questionnaire assignments are displayed to interviewers in the Patient Tracking System so that the appropriate forms are selected during the time of interview.

Section 5: Cohort/Patient Tracking System

Managing a large cohort is a significant logistical undertaking. Ensuring a proscribed sequence of participation events requires operational systems that routinely record subject status accurately so that the next event can be ascertained and future workloads be determined.

Subjects enrolled in the HR/CSC can take a variety of trajectories through study events depending on date of enrollment, specific cancer type, and length of follow-up (see Figure 1 for general sequence of events). Tracking progress through study participation events requires a system for logging the sequence and completion of participation events for each individual.

During clinic recruitment, HR/CSC staff utilizes a schedule report based on information from the clinic scheduling system. This data is pre-populated onto a tablet computer running the enrollment tracking software (Mi-Co Mi-Forms, Mi-Corporation, Durham, NC) to expedite the enrollment process. When a patient is approached for enrollment, the staff member is able to quickly look up patient information and complete the electronic consent and demographic questions. Full information on enrollment procedures can be found in the UNC HR/CSC Clinic Enrollment Protocol.

After consent is obtained, blood is collected from the participant in the clinic. Several data elements are recorded for each participant during this process. The participant's blood draw eligibility is determined shortly after consent is given and is recorded within the enrollment software. Before a blood draw event, the recruitment staff member takes a preprinted unique barcode number and affixes one copy to each of two tubes (one red, one lavender) for tracking purposes. This number is then associated with the participant's unique HR/CSC ID within the enrollment software. After the phlebotomist collects the blood specimens, the blood collection is recorded as completed, and the time of blood draw and transport to the blood processing laboratory is recorded in the enrollment tracking system. More information on blood sample collection and transport are available in the Biospecimens Collection, Processing, and Storage Protocol.

Following a patient's consent they become eligible to receive the computer-assisted telephone-administered baseline general and cancer-specific questionnaires. At the time of enrollment, the telephone interview is scheduled using the HR/CSC tracking software. As cancer-specific billing codes are generated during visits to UNC Hospitals, the tracking system automatically matches codes that trigger cancer-specific questionnaires.

Study staff use the participant tracking system to load the interview appointment schedule. Each entry in the schedule links to the corresponding participant profile, which displays contact information and a contact record form to the interviewer for the purposes of annotation. Interviewers then open the computer assisted telephone interview system visualized on a second computer screen to perform the interview. Details on the specific use of the study tracking system are detailed in the Interviewer Protocol Document.

The Patient Tracking System has a number of key features. For a summary of each participant, the system provides a profile page that allows a user to see the status of key participation events at a glance for a particular participant. Additionally the interviewer dashboard provides a summary of outstanding scheduled interviews across the cohort to facilitate HR/CSC workflow. Other features include participant lookup, generation of patient materials (e.g. thank you letters, newsletters), and is an administrative interface for report generation. The interface also provides staff with a way to record and verify a subject's vital status. Screenshots of the interface are available in the [Patient Tracking System appendix](#).

The participant tracking system is also used by HR/CSC staff to track and record events for ancillary studies. Each ancillary study has its own dashboard displaying enrollment summaries and status of key study events, which may include interviews, mailings, and specimen processing.

In summary, this system provides a unified interface for collecting information important for monitoring global-level events, participant-specific status, and ancillary studies.

Appendix: System-Specific Information

Enrollment Tablet Software (Mi-Forms)

The HR/CSC study utilizes Mi-Forms software (Mi-Corporation, Durham, NC). This software runs on laptops with tablet/stylus capability running Windows XP.

MI-Forms Implementation with HR/CSC: Frequently Asked Questions

How is MICO Used?

MI-CO products are used during the consent process. Study staff approach patients in the clinic and seek a response regarding willingness to participate in the study. An electronic form is filled out for each participant.

How are records entered?

A form is loaded for each individual approached. Pen-based entry on laptops is used to complete forms. The forms auto-save to a network disk at the Lineberger Comprehensive Cancer Center.

How are new records created?

HR/CSC ID's are generated for eligible individuals before the study visit, based on clinic schedule records. These ID's are used to instantiate a unique intake questionnaire form identifier. Once an individual is approached for consent into the study, the staff member loads the form corresponding to the subject. The staff member then shows the consent video to the patient, and records consent status. If the patient consents, a demographics form is filled out, as well as information on blood collection status. If a patient refuses consent, only that consent status is logged.

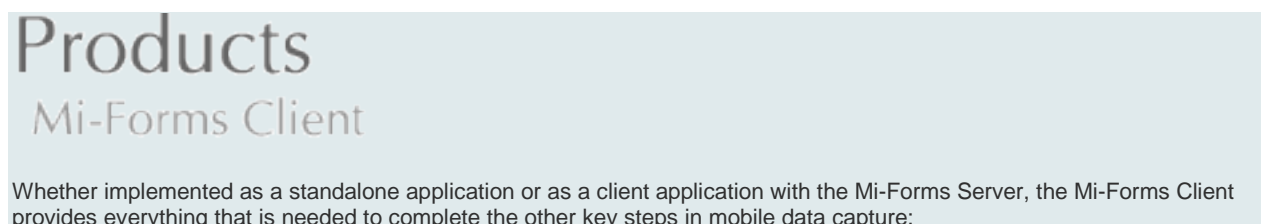
What is entered onto the tablet?

For individuals who do not provide consent, the appropriate consent status (e.g. "refused" or "check back") is noted in the consent status form. For individuals providing consent, the consent status is set to "obtained" and a subsequent form containing demographics questions is completely filled out.

Where does the information ultimately reside?

Data stored on tablet is exported to the HR/CSC database through the Mi-Forms Administrative Server application.

MI-Forms Product Description



The image shows a light blue rectangular box with the word "Products" in a large, dark grey serif font at the top left. Below it, "Mi-Forms Client" is written in a smaller, lighter grey sans-serif font. At the bottom of the box, there is a short paragraph of text in a small, dark grey sans-serif font.

Whether implemented as a standalone application or as a client application with the Mi-Forms Server, the Mi-Forms Client provides everything that is needed to complete the other key steps in mobile data capture:

- Prefill
- Capture
- Verify and Validate
- Communicate

Prefill: Prefill your forms with data you already have.

Data from another source may be 'pulled' and used as prefill information into a form. This reduces the amount of information needed for capture, reduces possible errors introduced without prefill data, and allows captured data to be linked back to the original record source, if needed.

Capture: Capture new data using a Tablet PC, Windows PC, or Digital Pen.

The Mi-Forms Client on a Tablet PC will convert handwriting into machine readable text as you write. Drawings, diagrams, and a number of other data sources may also be captured if needed (images, audio, barcodes, gps, attachments of any files, etc).

On a Windows XP PC or above, the Mi-Forms Client provides a keyboard and mouse entry interface for capturing data.

With Mi-Forms support for Anoto-based digital pens from Logitech (and other vendors as well), fill Anoto-based paper forms with the digital pen, dock the digital pen with your PC and Mi-Forms will capture and convert ink into machine readable form.

Verify and Validate: Verify the handwriting recognition results and validate data relationships (date / time evaluation, business rules, etc.) on your form.

On the Tablet PC, verification and validation checks are performed as the user is collecting data in real-time, at the 'scene' of data collection.

For digital pens, verification and validation checks can be performed on the PC when the pen is docked,.

Communicate: Communicate and share your data with existing enterprise systems.

Data may be exported from the Mi-Forms Client to the centralized Mi-Forms Server and then to other existing data repositories or exported directly from the Mi-Forms Client. Recognized data (text), ink, and form page background data may be exported via a variety of formats (csv, xml, ODBC, MSJet, Outlook, variety of image formats, etc.) Data exports may also be customized for greater flexibility in exporting your data to your existing data system, (ie. HL7, SVG, etc.).

In fact, in each of the steps Mi-Forms offers a host of options to provide you with the flexibility to implement mobile data capture applications in the way that best fits your organization's needs.

Mi-Forms Administrative Server

The Mi-Forms Server Version 7.0 allows the enterprise to control the flow of data, manage accounts, implement predefined workflow, and deliver solution updates to the users. Some of the key functions of the Mi-Forms Server include:

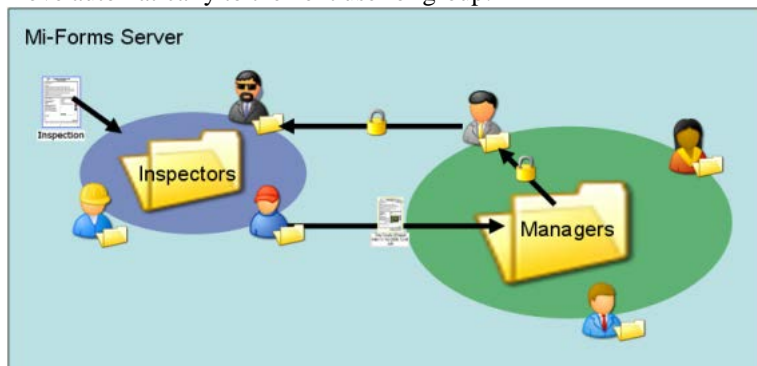
- Manage users, passwords, groups and privileges
- Manage form and solution revisions
- Forms based workflow
- Centralized data export including custom data exports from script
- Server-side script code using same APIs and event model as client / component
- Open web-services based communication over HTTP / HTTPS that can be integrated with Mi-Forms SDK built applications



Centralized Management with Mi-Forms Server

- Data validations can be performed server side
- Supports workflow via concept of queuing
- Enables data and form sharing between individual users or groups
- Forms can be routed based on data content
- Decisions on routing can be made either automatically via script code or via user interaction at form-finish time

Using the Mi-Forms Administrative Server you can create a predefined workflow pattern. The data collected can move automatically to the next user or group:



Supported Technologies

- Windows Server 2003, 2008 R2
- SQL for data storage (SQL Express or SQL 2005-2012)
- ASP.Net under .Net 3.5

Overview of Biospecimen Process Facility Laboratory Information Management System (LIMS) Quality Assurance / Quality Control

The Biospecimen Process Facility (BSP) Laboratory Information Management System (LIMS) uses Oracle as a data repository. There are three databases hosted on two servers supported by Instructional & Information Systems, UNC School of Public Health. Backups occur on both servers on a daily basis. As the need arises, snapshots of the production database are copied to the two additional databases (staging and development). In the unlikely event of data corruption on the production server, the School of Public Health will be able to restore the system from the latest backup.

The purpose of the staging database is to provide a test environment for application code prior to its release into the laboratory environment. With the use of custom SQL scripts, it is possible to rapidly setup a mirror image of the production database. This is useful for rapid response to LIMS maintenance issues. Users can report bugs by sending an email to informaticshelp@unc.edu. This will create a help request in the Lineberger Bioinformatics group ticket system. Application code running against the staging database is always found in the highest numbered branch within the code versioning system.

The development database is also periodically mirrored from production but with possible schema modifications. This database provides a data environment for major code changes, which sometimes continue over a period of months. Application code running against the development database will always be found in the HEAD branch of the code versioning system.

LIMS code development is done using the Eclipse software development tool. With few exceptions, aspects of the LIMS application (Java source code, SQL scripts, database packages, help documentation, etc.) are maintained within Eclipse and all of these files are backed up in a code versioning system on *bioinfostore.unc.edu*. There are two branches of code development: HEAD and Branch_X, where X represents the base-version number of the latest, or highest release. At the time of writing this document the latest release is 11, e.g. Branch_11 is the current branch.

Before releasing a new version of the LIMS, we conduct unit and integration testing. During unit tests, developers verify that the newly added code functions properly and solves reported bugs. Integration testing is used to make sure the changes don't break existing functionality. We have a series of general test cases that we run before releasing a new version. The steps for deploying a new version are documented and stored in our group's internal directory. We are implementing code reviews as a part of the deployment process.

Each release is tagged with a version number. The code versioning system (CVS) maintains a history of the code versions and a change history for each file in the repository. From Eclipse, you can view the change history for a given file and compare it to another version. We document the changes in each release on our group wiki.

When new studies are added to LIMS, the BSP lab managers meet with the principal investigator to discuss the sample processing workflow. Then, the BSP lab managers create protocols within the LIMS to standardize the procedures in each step of the workflow. Additionally, each lab technician receives periodic training to learn the lab's standard operating procedures as well as to maintain certifications pertinent to the work in the lab (i.e. blood borne pathogen, HIPAA, and CITI training). The LIMS has the functionality to maintain a listing of the certifications and validity dates for each user. The LIMS also has the functionality to track project-specific IRB certifications.

On a monthly basis, the BSP lab managers generate check-in reports from LIMS to verify the validity of sample collection and drop-off dates. In the LIMS project setup, there is an option to send check-in reports to study staff and lab managers on a weekly or daily basis. There is a scheduled task that runs daily and emails the report to registered users. This helps the study staff and lab validate what samples were sent and received each week. The LIMS maintains an audit trail for each sample processed in the system. Each audit record contains the technicians name and a date and timestamp. The audit trail tracks everything that is done to the sample from the time it was checked into the system to when it was discarded or shipped to another facility. There are audits for sample check-in, storage assignment, extractions, open and closing of protocols, discards, plan changes, etc.

The LIMS uses several different types of input fields on the interface depending on the type of data being entered. For numeric values such as volume and concentration, the LIMS displays an input field that allows the user to select a number from an ordered list. There are up and down arrows next to the data field for stepping through the values in the sequence. The LIMS uses checkboxes for yes and no fields. For example, to mark a sample as discarded, the user can check the discard box. Dropdown boxes are used to select values like units, container types, and project names. Input fields for comments, participant ids, drop off courier, etc. are free text.

The LIMS provides data quality assurance and quality control by validating data entered into the system. For example, Optical Density measurements are taken on the Nanodrop and the DropSense. The equipment saves the optical density measurements and concentrations to a file. When that file is loaded into the LIMS, the system will verify that the data is numeric. If it's not, the LIMS will display an error message to the user. If a donor code format is defined in a project check-in protocol, the LIMS will check to see that the entered text fits the desired format. The user will not be able to check-in the sample until he/she enters a participant id in the correct format. Each protocol can define certain fields as required. When processing samples through a protocol, technicians aren't able to complete a work set until they have entered data into all required fields. These are just a couple examples of cases where the LIMS validates user input.

Biospecimen Process Facility Information Technologies (LIMS)

The BSP utilizes a state-of-the-art specimen tracking system designed specifically to meet the immediate, inventory-style tracking needs of multiple health research projects. It is referred to as the LIMS (Laboratory Information Management System). The LIMS has been designed to be flexible and extendible for future needs. All samples entering the lab receive 2D bar code labels that allow us to follow any sample's progress through the specific procedures indicated by the appropriate study protocol. Samples that enter the lab are given a BSP ID number, further protecting a study subject's rights. These labels also contain a human readable portion and additional study encoded ID information.

The Biospecimen Processing Facility LIMS is Oracle-based and maintained by the School of Public Health with security and backup. The user interface (developed and maintained by the School of Medicine) is implemented as a Java thick-client with limited (read only) WEB access provided to principle investigators and their staff. The thick-client tracks creation, storage, movement, and discarding of containers (vials and plates) and the specifics of container contents (material type, volume, concentration, technician comments, etc.) and is accessed only by BSP Facility staff. All containers have a study context. BSP staff has access to all studies but only specific PI staff (as setup within the system) can access information associated with their project via the WEB client.

LIMS users are required to login using their user id and password. The passwords must be at least 8 characters and must contain a number and a special character (!,@,#,etc.). The LIMS will not allow users to re-use their last two passwords. Each user is assigned a role, which defines what features are available to them. In addition to roles, PI staff members are associated with projects. Staff members can log into the WEB client to view reports for samples in their projects.

Each container created by the system is assigned a BSP identifier that is unique - in effect, this is the facility asset identifier for that container/sample. Each container is labeled with the BSP identifier, project/donor identifiers, type of contained material, volume, concentration (if applicable), date of creation, and current storage location. Containers may be pre-assigned to a protocol path that eventually leads to the departure of the containers from the facility, but may also result in long-term storage within the facility. Through a query interface, sets of containers may be identified, pulled, and be further processed within the facility or be shipped to other facilities.

An audit trail is maintained on containers and content. This trail records such things as when and who drops off containers to the BSP, when the container was created as it moves through processing in the BSP as directed by the studies protocols, which technician created it, when and where it was stored, when and where it was pulled from, etc. The system will track materials that are transferred to other facilities or PI collaborators as well.

The LIMS has the capability of accepting processing logs generated by the Auto pure LS (DNA extraction robot) and the Multirole II (aliquoting robot). These logs can be

transferred to the LIMS at the completion of each run. These logs are linked to the individual samples histories and can be referred to in the future for a variety of reasons. The results from analysis done within the facility are recorded in the LIMS database. This includes agarose gel analysis (with image), optical density (with wave graph), real-time PCR and bioanalyzer results. Analysis reports are available to both the facility staff and project staff (through the Web interface). The current inventory of a given project is also available as a report. This shows the entire project history in terms of material that was checked into the facility, created by the facility, and shipped from the facility.

No Personal Health Information (PHI) is maintained anywhere in the BSP systems. When donor samples are checked into the facility the system allows for limiting the use of the samples based upon donor preferences (i.e. allow genetic research, destroy at project completion, or restrict research to original project).

On a daily basis, the Lineberger Data warehouse and Biospecimens Repository (LDBR) loads LIMS production data for reporting. The data warehouse uses a read-only account to connect to the LIMS database. Storing LIMS data in the data warehouse allows HR/CSC staff to create reports for their samples. These reports include information about the samples that have been checked in, the sample inventory, and nucleic acid analysis data.

Future Tissue Procurement Facility Laboratory Information Management System (TPF LIMS)

Development of the future TPF LIMS System will be based on the BSP LIMS system and will be similar in scope, development, and implementation. The Future TPF LIMS system will store information on collected tissue, pathology status, sample processing protocols and methods, and banking and inventory information.

Medical Record Abstraction: Electronic Registry System Software (ERS)

Electronic Registry System (ERS) is a proprietary Cancer Registry software tool for collecting and storing medical record data. The software standard data set required by the Commission on Cancer (CoC) and meets the requirements of the North American Association of Central Cancer Registries (NAACCR).

The UNC Cancer Registry's Certified Tumor Registrars use ERS as a data collection tool for manual electronic medical records abstraction – required by state law. Approximately 80% of the hospitals in North Carolina use ERS.

The ERS software tool for the UNC Cancer Registry is maintained and hosted by ERS and delivered as a virtualized desktop. It is accessed via website and requires entry of credentials. The UNC Health Registry/Cancer Survivorship Cohort (HR/CRC) purchased the ERS software tool to facilitate transfer of UNC Cancer Registry data directly to HR/CRC. The UNC HR/CRC ERS will be managed by the Lineberger Data Warehouse and HR/CSC IT

team and ERS programming support group in conjunction with ISD to assure versioning comparability. In addition to the required CoC data sets (see Table 1), ERS makes available to the UNC HR/CSC other data fields, including those that can be customized, to fit our specific needs. The additional fields include expanded research and diagnostic fields, clinical trial fields as well as user defined fields that can be created and maintained by the user. The HR/CRC staff will manually abstract these additional data fields (see Table 2) from the electronic medical record to complement the data that is provided directly in ERS from the UNC Cancer Registry. Coding manuals are online and maintained by the ERS software company. ERS provides users with a very extensive reporting module for data analysis to compare treatment analysis, survival, incidence by site, stage, race, etc. Any data item in the software can be selected and used for reporting purposes. The software contains a follow up module that allows the user to automatically schedule follow up lists and letters with patients.

ERS provides three days of initial on-site training for new users as well as unlimited on line training. A complete user manual is provided as well as a Video Library of training tutorials (see online Resource list below). The software is updated quarterly (ERS notifies ISD or user via email that the software update and document detailing update changes is available) and quarterly training webinars are provided that are awarded continuing education credits from the National Cancer Registrars Association (NCRA). ERS provides unlimited support to users for all registry and technical issues.

To be a CTR (Certified Tumor Registrar), one is required to have a minimum of a 2 year degree in a health related field and pass the national CTR test. As of 2012, each registrar must be certified to perform abstraction. There are internal edit checks within the software that come up if illogical information is entered, such as female for prostate cancer, zip code and county code matching errors as well as procedure and staging conflicts, such as entering pathologic staging when the patient has had neoadjuvant therapy. Required fields cannot be left blank. Additionally, about 10% of the cases are identified by the UNC Cancer Registrar and are re-abstracted to check for accuracy. A quality control protocol involving re-abstraction of a proportion of records to monitor data quality of additional HR/CSC abstracted data elements will be implemented. Additionally, both UNC Cancer Registry and UNC HR/CSC manually abstracted medical records data elements will be compared with planned EMR extractions for additionally quality assessment.

Resources:

1. ERS Website
<http://www.ers-can.com/>
2. FORDS manual for 2013
<http://www.facs.org/cancer/coc/fords/fords-manual-2013.pdf>
3. 7th edition AJCC staging manual: <http://www.cancerstaging.org/products/ajccproducts.html>
ISBN: 978-0-387-88440-0 (must purchase for \$65 – see website)

Participant Tracking System

The **Sheps Integrated Research System (“SIRS”)** is a secure, enterprise database and programming framework specifically designed to meet the needs of health research projects at the Cecil G. Sheps Center at UNC Chapel Hill (“Sheps”). SIRS provides reusable modules for data collection, management, and tracking for a diverse set of studies, registries, and research coordinating units. SIRS is specifically designed to provide a framework for programming that simplifies common programming tasks and allows customization while meeting applicable security requirements. The SIRS framework is completely web based and open source utilizing MySQL, Apache, JavaScript and PHP.

The Sheps Integrated Research System offers many advantages to investigators:

- Security
- Data management expertise
- Input and feedback from experienced SIRS programmers
- Ability to incorporate designs from previous studies
- Continuity within projects but across projects over time
- Faster “into the field” capability
- Lower cost
- Availability of common components with tested and verified code
- Thorough data auditing

All system login procedures and data submissions (e.g., from study staff or survey respondents through the Internet) are encrypted via the Secure Sockets Layer (SSL) protocol to a secure central database at the Sheps Center. User-level permissions are based on user roles and are defined within the project system to limit each user’s access to only those records they are authorized to see.

The central database for projects runs on a mirrored server system with automatic fail-over features, daily backups and transaction logs. This system is physically located in a Tier II Data Center providing backup power sources, climate control, fire protection and 24x7 surveillance.

Audit logs are reviewed routinely to verify security measures are operational. The servers are scanned weekly for vulnerabilities and are currently maintained at the highest level of vendor and CERT security recommendations. Data are never shared outside the project unless authorized by the project leader. User authentication is based on user passwords. Password creation requirements are in place to guarantee “strong passwords” as defined by the CERT security recommendations. The lead systems administrator is GIAC Security Essentials certified through May, 2014.

The UNC Sheps Center endeavors to preserve the privacy, confidentiality, and security of protected health information that may be part of health records or research datasets. Protected Health Information (PHI) is handled according to appropriate Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy and Security Regulations. Sheps Center programmers and research staff who work with sensitive data are required to complete appropriate HIPAA training with periodic updates, complete Sheps Center internal training, comply with the UNC IT Security Policies, and agree to the provisions of

the Sheps Center’s Rules of Behavior and Sanction Policy. The Sheps Center strives to implement reasonable security controls guided by FISMA, HIPAA, and OMB Circular A-130, Appendix III.

Participant Tracking System Screenshots

The participant tracking system is a web-based application that provides real-time information on cohort status and participant information.

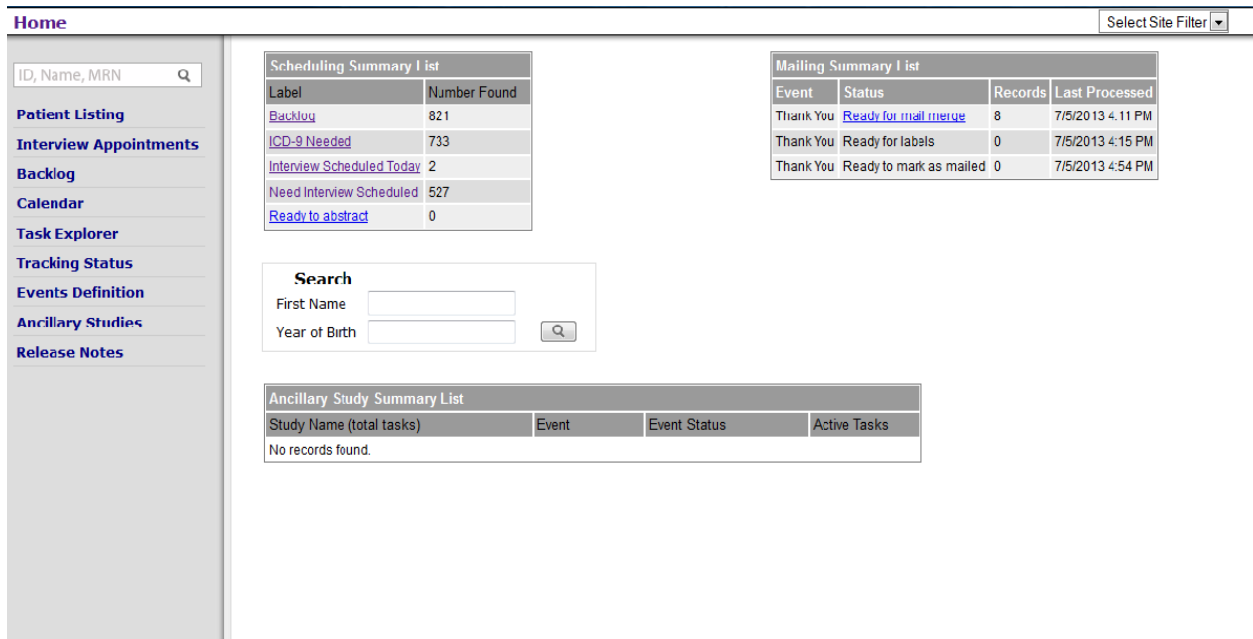


Figure 2. PaTSy Home Screen. This dashboard provides summary information on Interview Status, Thank You letters, and Ancillary Study Tasks

The participant tracking system provides a global cohort dashboard (Figure 2) as well as a participant dashboard (Figure 3). This enables staff the capability to see both “tree” and “forest” views.

Home | 41742446: GHOST, CASPER Select Site Filter Version: 1.1

Summary | Scheduling | Events | Blood Work | Alt Contact | Ref Phys | ICD-9 | Ancillary Studies

ID, Name, MRN

Identification
 Patient Name CASPER GHOST
 MRN 41742446
 CDART ID 00000245874260

Informed Consent
 Filename [210048827.pdf PDF](#)
 Timestamp 7/24/2012 8:50:49 AM

HIPAA
 Filename [1626058547.pdf PDF](#)
 Timestamp 7/24/2012 8:51:29 AM

Deceased
 Adverse Event

Status

Status ID	6692
Tracking	Consented <input type="button" value="Edit"/>
Return/New	NEW
Consent	Obtained
Video	Paper Form Only
Port Status	N
Blood Draw	Check Back
Baseline Interview	Scheduled
Thank You Letter	
First Contact	07/23/2012
Last MIForms	07/23/2012
Consenter	Andrew Hjelt
Date of Consent	07/23/2012

Demographics

First Name	CASPER	Ethnicity	No (Hispanic/Latino)
Middle Initial		Marital Status	Single, never married
Last Name	GHOST	Education Level	Postgraduate or professional degree
Suffix		Race 1	White
Maiden		Race 1 Other	
Date of Birth	07/16/1964	Race 2	
Gender	Female	Race 2 Other	
Primary Language	English	Race 3	
Work for pay	Yes	Race 3 Other	
Hours of work per week	40 or more		

Contact Info

Figure 3. Patient Summary Page - Contains contact and demographic information and links to Informed Consent and HIPAA Documents

The patient contact information and call log section (Figure 4) provides a convenient space to record contact events and completion status.

Note Log

Timestamp 7/24/2012 8:48:07 AM

Address1 000 NOWHERE WAY
 Address2
 City HAUNT
 State North Carolina
 Zip 00000

Home (000)000-0000
 Phone
 Work Phone
 Cell Phone (111)111-1111
 Email
 Best HomePhone
 Contact

Date of Contact
 Time of Contact
 Contact Reason
 Disposition

Status Definitions

Method of Contact
 Sticky Note
 Call Comments

Health Registry

Events				
Interview	Status	Interviewer	Event Date	Form
60. Baseline Questionnaire	Scheduled	Kathryn Allman	08/09/2012	
202. GI Gastric	Scheduled	Kathryn Allman	08/09/2012	

Figure 4. Patient Contact Information and Call Log

The participant event status page (Figure 5) provides a brief summary of key study events.

Statuses		Cancer Specific	
Status ID	6692	202. GI Gastric	Scheduled
Tracking	Consented		
Consent	Obtained		
Informed Consent	Saved		
HIPAA	Saved		
Demographics	Saved		
Video	Paper Form Only		
Parking Voucher	Saved		
Consent Checkback			
Blood Draw Checkback	Needed		
ICD-9	Completed		
Port Status	N		
Blood Draw	Check Back		
Barcode Link			
Blood Draw: Red Tube			
Blood Draw: Lavender Tube			
Baseline Interview	Scheduled		
Thank You Letter			
First Contact Date	2012-07-23 00:00:00		
Last Contact	2012-07-23 00:00:00		
Date of Consent	2012-07-23 00:00:00		
Note Log			
Timestamp	7/24/2012 8:48:07 AM		

Figure 5. Participant Event Status Page. Specific Information on Events and Cancer-Specific Interview.

The ICD-9 code entry screen (Figure 6) provides an interface to set cancer-specific questionnaires for each participant. These questionnaires can be automatically assigned by the billing code algorithm, manually entered, or temporarily skipped.

ICD-9 Codes	Service Date	Entry Type
151	05/05/2012	Manual by Andrew Hjelt
123.12, 123, 123.1		Kathryn Allman
123.12, 123, 123.1		Kathryn Allman
123.12, 123, 123.1		Kathryn Allman
123.12, 123, 123.1		Kathryn Allman
123.12, 123, 123.1		Kathryn Allman
123.12, 123, 123.1		Kathryn Allman
123.12, 123, 123.1		Kathryn Allman
123.12, 123, 123.1		Kathryn Allman

Figure 6. ICD-9 Entry Screen. Cancer-specific ICD-9 Codes trigger specific questionnaires.

The Interviewer Appointment Queue (Figure 7) screen provides both an interview schedule and ordered list of telephone interviews outstanding. Interviewers can utilize this screen to select the next interview to perform.

Patient Listing

Interview Appointments

Backlog

Calendar

Task Explorer

Tracking Status

Events Definition

Ancillary Studies

Release Notes

Hot List

(count: 7)

Interviews	MRN	New/Return	Name	Last Contact	Date of Consent	Details	Spanish Speaking	Interviewer	Viewed By
Baseline Questionnaire:Scheduled 08/09/2012 at - Kathryn Allman	41742446	NEW	CASPER GHOST	8/9/2012	7/23/2012	★ ktlstf		Kathryn All	
Baseline Questionnaire:Scheduled 10/11/2012 at 8pm - Andrew Hjelt	41647322	RETURN	MANMOHAN LAKHANI	10/18/2012	9/24/2012			Andrew Hje	
GU Prostate Questionnaire:Scheduled 10/29/2012 at 3:01 PM - Andrew Hjelt	19924158	RETURN	ANDREW HJELT	10/29/2012	7/23/2012			Andrew Hje	
Baseline Questionnaire:Scheduled 11/08/2012 at 3:00 PM - Luis Serpa	41876681	RETURN	JANE DOE	11/7/2012	11/6/2012			Luis Serpa	
Baseline Questionnaire:Scheduled 11/08/2012 at 3:05 PM - Luis Serpa	8747776	RETURN	JOHN DOE	11/7/2012	11/6/2012			Luis Serpa	
Baseline Questionnaire:Scheduled 11/27/2012 at 10:30 AM - Nora Christopher	9242157	RETURN	FRED ASTAIRE	3/26/2013	11/6/2012			Nora Christ	
Baseline Questionnaire:Scheduled 01/10/2013 at 3:00 PM - Caroline Hempel	41915752	NEW	PENELOPE CRUZ	1/7/2013	11/5/2012			Caroline H	

Queue

<< first < prev 1 next > last >>

Interviews	MRN	New/Return	Name	Last Contact	Date of Consent	Details	Spanish Speaking	Interviewer	Viewed By
Baseline Questionnaire:Need date	20576377	NEW	STEPHEN WOOD	10/09/2012	10/05/2012			Andrew Hje	
Baseline Questionnaire:Need date	41764259	NEW	STEVIE WONDER	10/18/2012	07/30/2012			Andrew Hje	

Figure 7. Interviewer Appointment Queue. The Hot List indicates interviews scheduled for the current day, while the Queue lists individuals with an outstanding interview status.