UNC Health Registry Data Source Documentation

Table of Contents

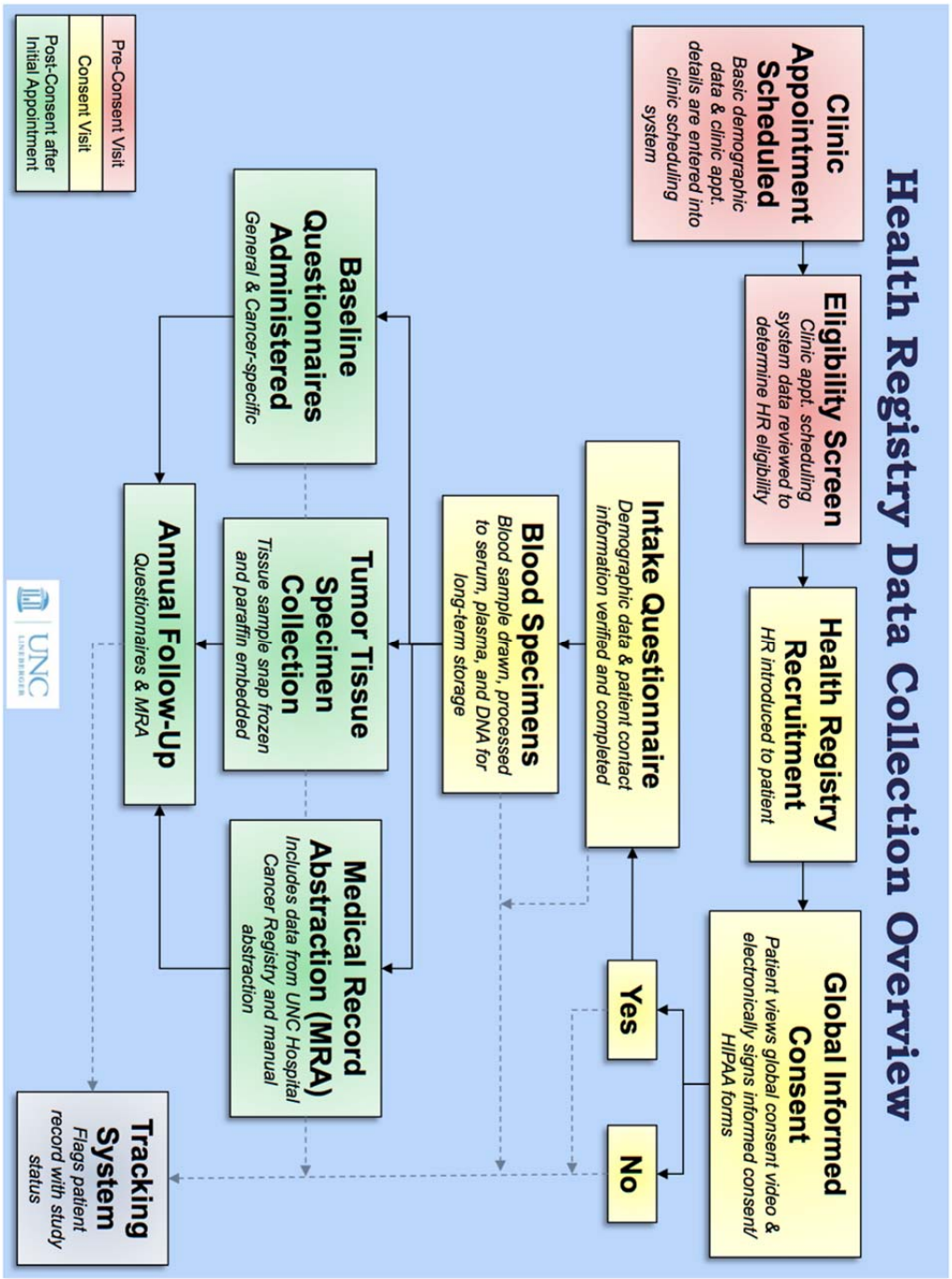

Figure 1 – Stages of Data Collection

Section 1: Tracking Study Events

With a large and complex study, ensuring a proscribed sequence of events requires operational systems that accurately and regularly record subject status so that the next event can be ascertained and future workloads be determined.

Subjects enrolled in the UNC Health Registry can take a variety of trajectories through study events depending on date of enrollment, specific cancer type, and length of follow-up. Tracking progress through study participation events requires a system for logging the sequence of events for each individual.

Upon a subject's enrollment into the study in the hospital (Recruitment Schedule → Successful Consent → Enrollment) they become eligible to receive the telephone-based baseline questionnaire. As cancer-specific billing codes are generated during visits to UNC Hospitals, the tracking system automatically matches codes that trigger cancer-specific questionnaires.

Interviewer dashboard, patient lookup, scheduling of interviews, and cancer-specific questionnaire are key features of the system interface. Additionally, the system is used for generation of patient contact materials and is a key administrative interface, and is a method for loading subject-specific study consent and HIPAA documents. The interface also provides staff with a way to record and verify a subject's vital/death status.

Details on the specific use of the study tracking system are detailed in the Interviewer Protocol Document.

Screenshots

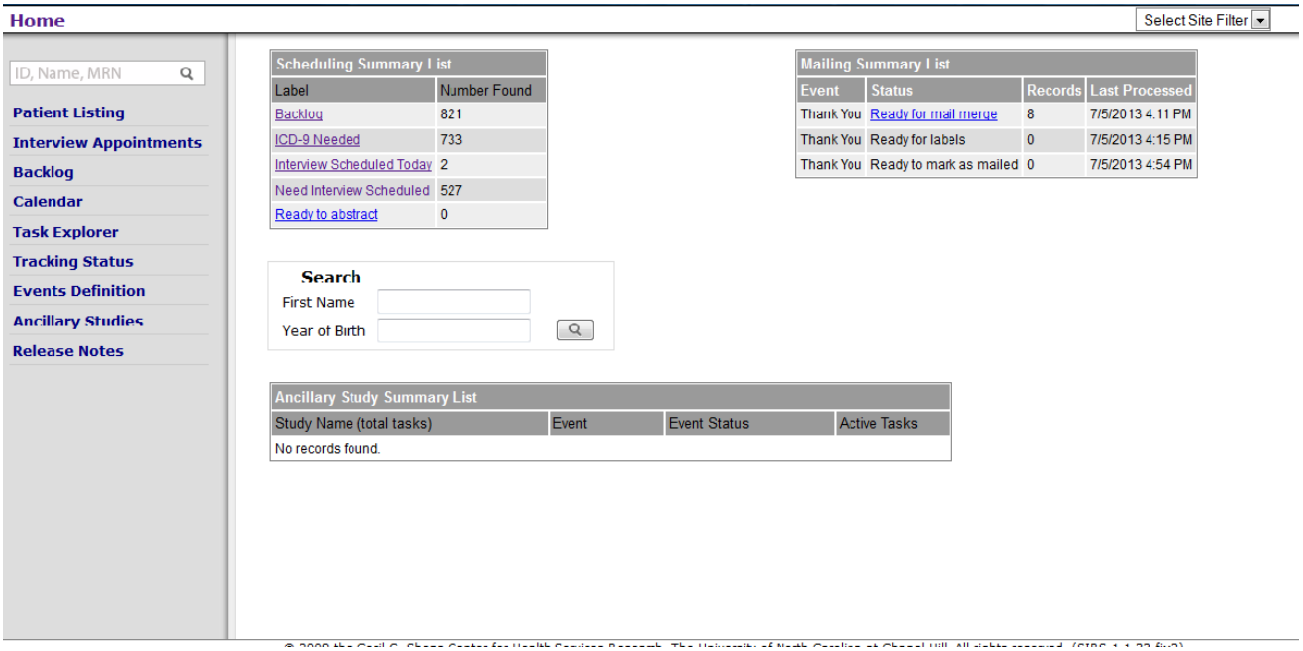

Figure 2. PaTSy Home Screen. This dashboard provides summary information on Interview Status, Thank You letters, and Ancillary Study Tasks

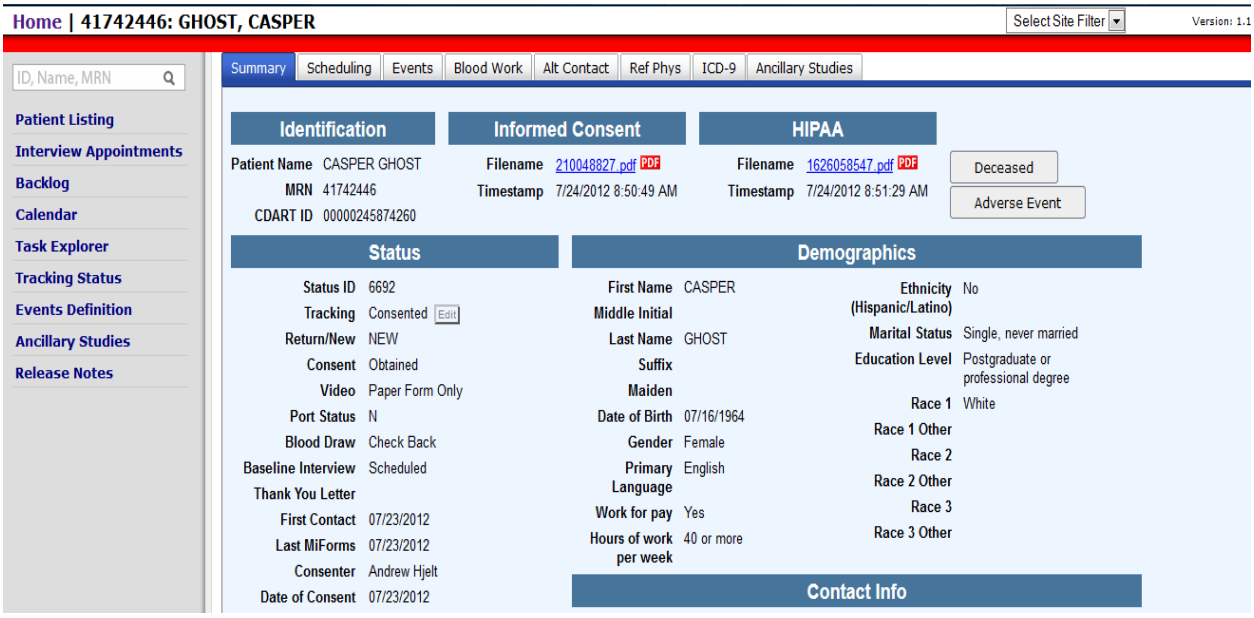

Figure 3. Patient Summary Page - Contains contact and demographic information and links to Informed Consent an HIPAA Documents

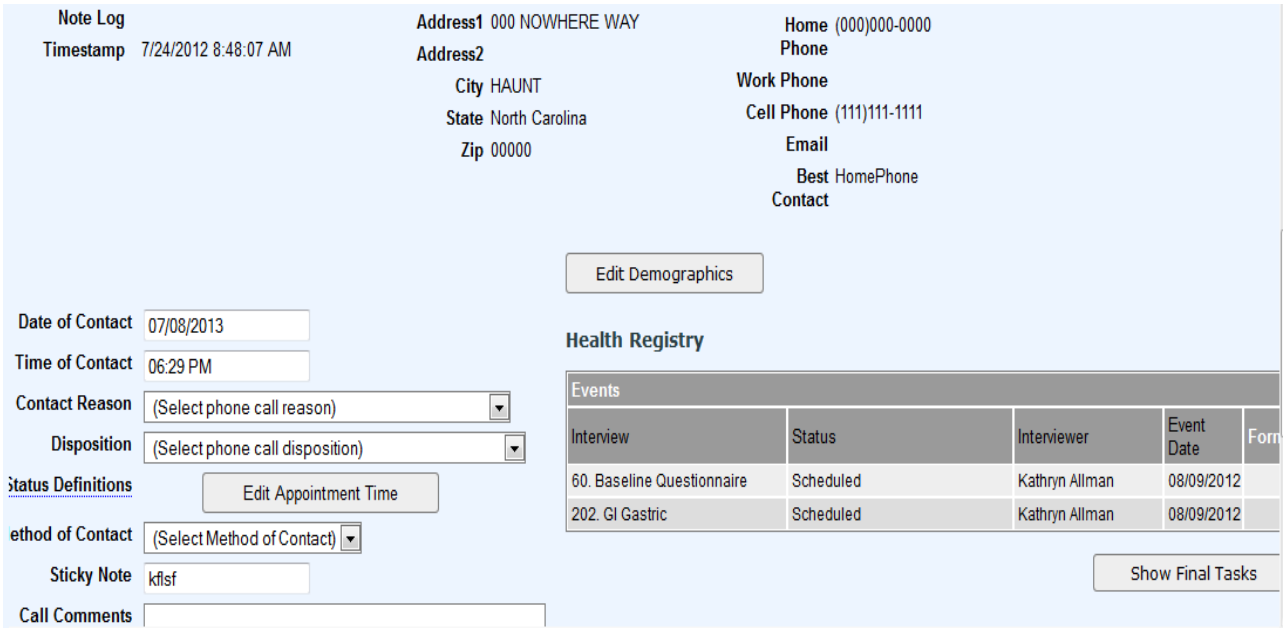

Figure 4. Patient Contact Information and Call Log

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

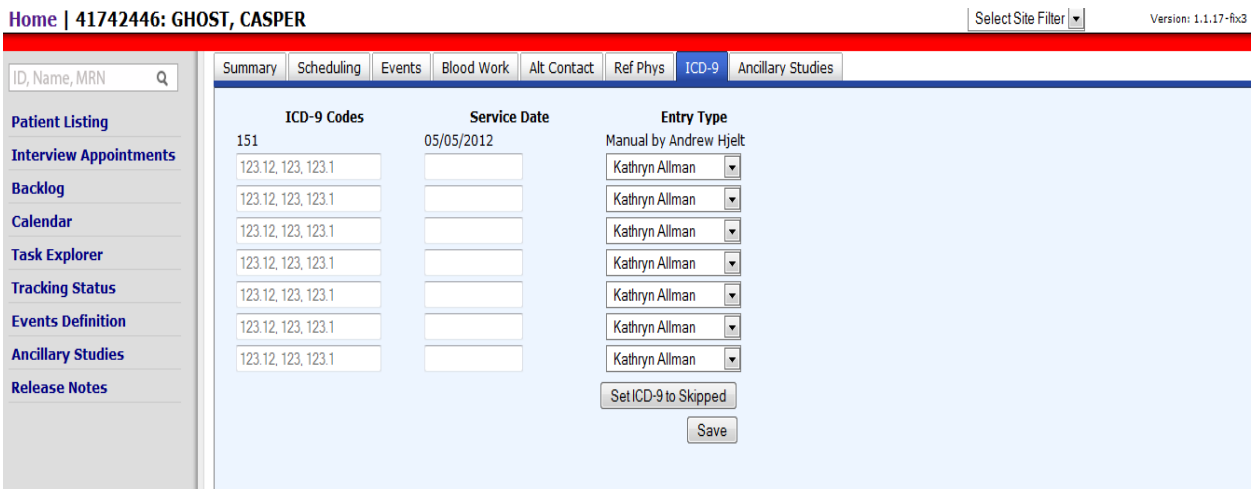

Figure 6. ICD-9 Entry Screen. Cancer-specific ICD-9 Codes trigger specific questionnaires.

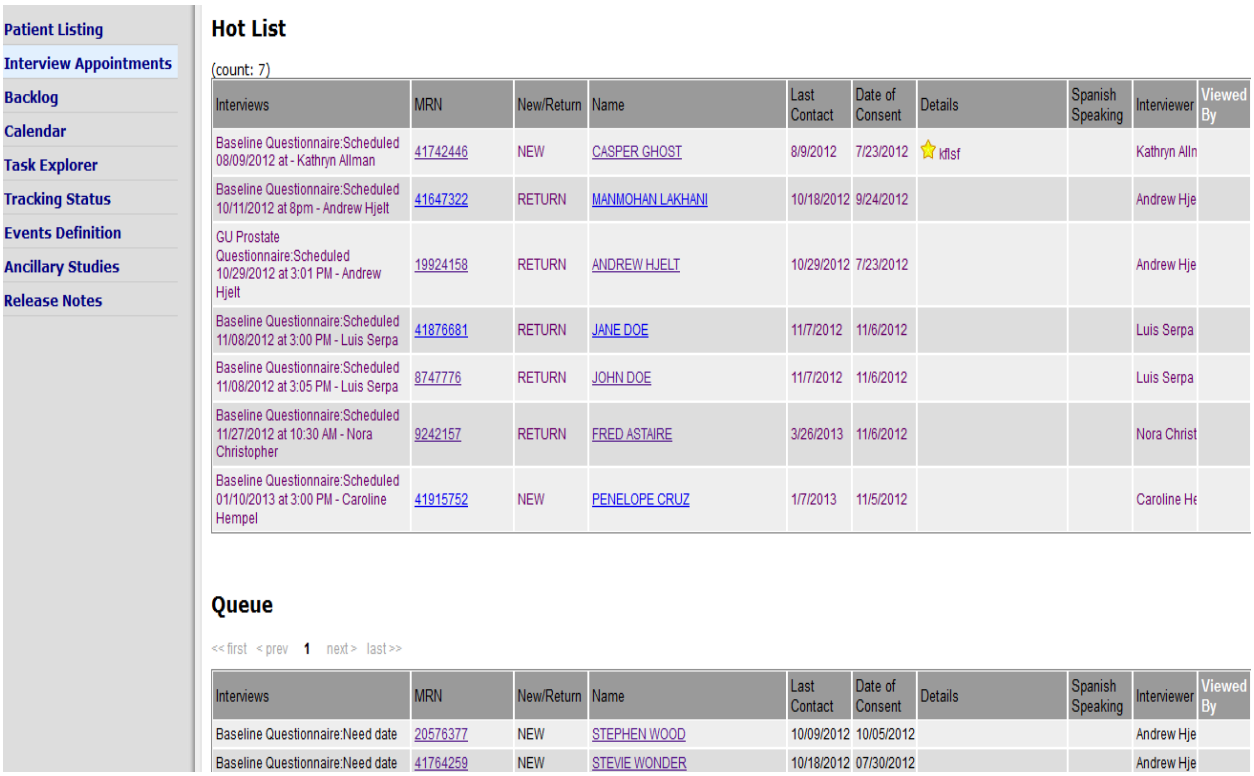

Figure 7. Interviewer Appointment Queue. The Hot List indicates interviews scheduled for the current day, while the Queue lists individuals with an outstanding interview status.

Section 2: Enrollment

Enrollment into the Health Registry occurs when an individual is deemed to be eligible based on clinic schedule data, is approached by a health registry recruiter, and informed consent is provided by the patient after learning about the risks and benefits of participation.

The enrollment process begins with a list of eligible individuals visiting cancer clinics at the NC Cancer Hospital each day. Clinic schedule data is obtained from UNC Healthcare appointment scheduling systems through the Carolina Data Warehouse for Health (CDW-H) and is prepared on a daily basis.

The clinic schedule report is populated with a list of individuals who screen to be eligible based on date of birth and residence, which are available from the hospital appointment scheduling system. Schedule reports are generated daily for Health Registry recruiters, who use this information to approach patients around the time of their appointment.

Once approached, an individual is informed about the purpose of the study, shown a brief video summarizing the purpose and scope of the Health Registry, and made aware of potential risks.

If consent is not given, the individual's final status will be set as "Refused." If the individual is interested but has not provided consent, the status is set to "Check Back" which can be set to

A brief intake questionnaire is completed during the consenting visit. Basic demographic and contact information is obtained, including name, address, date of birth, gender, and race. Additionally, marital status, employment status, and information about the referring physician are recorded.

To obtain the most up-to-date contact information for each subject, this information is confirmed during telephone interviews.

Health Registry recruiters utilize computer tablets running MiForms software (Mi-Co, Durham, NC) to record notes, conduct the intake questionnaire, and to process consent document signatures. The tablets are connected to a secure wireless network, and data is sent to a database and fileserver over an encrypted network connection. Paper copies of all consent documents are used, filed, and back-entered in the event of connectivity issues.

Section 3: Biospecimens

Upon enrollment into the Health Registry, a subject may provide biospecimens such as blood or tumor tissue.

On or after the participant's date of consent, blood samples are drawn by a phlebotomist, logged and collected by health registry staff, and transported to the UNC Biospecimen Process Facility. Blood samples are not collected for partipants who have a portacath installed. Subjects may choose delay or refuse the blood draw procedure. As a result, the final blood draw status and number of approaches are recorded in the study database.

All blood draw events are logged by Health Registry recruitment staff using computer tablets running MiForms software. Participants are assigned a unique barcode number used to track blood samples. A recruitment staff member brings the patient to a phlebotomy suite for the blood draw and gives the phlebotomist barcoded red and lavender tube vials. The time of blood collection is logged in MiForms and samples are temporarily refrigerated on-site. Within two hours of collection, batches of refrigerated participant blood samples are transported and logged into the BSP for further processing, aliquoting, and banking.

Biospecimen Process Facility – Blood Samples

Once participant blood samples are taken following consent into the study, they are sent to the UNC Biospecimen Process Facility (BSP). Each sample taken is collected in a tube with a serial number linked to the individual. Samples are batched, packed in a cooler, and walked to the BSP laboratory by a Health Registry staff member (a trip that lasting 5 to 10 minutes). Once the samples are delivered to the BSP, they are entered into the BSP laboratory information management system (LIMS) by the laboratory technician and Health Registry staff member.

Once blood samples are signed into the LIMS, BSP staff are assigned new work and apply a protocol to measure, aliquot, and store child samples – the sequence of which has been pre-programmed into the LIMS. All protocol steps are proscribed by the BSP LIMS system, which provides an interface to input measurements, record notes, and to assign child sample container identification numbers.

Child samples are banked in freezers in two separate locations to provide a fail-safe against equipment failure or physical breach, allowing samples to be securely stored. Requests to pull samples (to perform additional protocols, to mail, to destroy) are handled by communication between Health Registry staff and BSP technicians. Once a technician is requested to handle a sample, the detail of the request is logged. Sample(s) are identified in the LIMS system, pulled from storage, and prepared according to the request. A confirmation is sent back to Health Registry staff to notify that the request has been completed.

Tissue Procurement Facility – Tissue

The Tissue Procurement Facility is a group sponsored by UNC Lineberger Comprehensive Cancer Center and partner of the UNC Health Registry study. A staff member from the Tissue Procurement Facility identifies ”targetable” surgical procedures scheduled among consented patients that involve tumor tissue. In the event that a research specimen can be saved, the TPF banks the sample for future use.

TPF staff are present each morning and include documentation for surgical staff indicating patient consent in a study requesting tissue for research. After surgery is performed, TPF staff then take custody of tissue and send it on to pathology for tissue analysis.

After the pathology lab has sampled the tissue, any remainder is sent to the TPF for processing and banking. The process of transporting a sample from the surgical suite to the sample’s final banking normally takes an hour or less.

The TPF uses two interfaces to perform their work. Scheduled surgical procedures are cross-referenced with a list of consented subjects from the Health Registry study. Matching surgery events are examined by study staff and placed on a target list. Surgical appointment target lists are printed for TPF staff each morning.

The TPF LIMS system is used to record information on any samples collected. Information included in the LIMS system includes specimen ID, banking status, and notes from the pathology report.

Section 4: Questionnaires

Interview-based data is collected during the enrollment process and subsequent baseline and follow-up interviews.

Limited demographic data is collected at the time of enrollment onto tablet computers. The tablets include data pre-populated from the CDW-H schedule. The enrollment questionnaire provides an opportunity to collect and confirm basic data including name, date of birth, referring physician, and contact information.

Once an individual is consented and enrolled into the UNC Health Registry, subsequent questionnaire data is stored in CDART – Carolina Data Acquisition and Reporting Tool – a web-based application that is well suited for computer-assisted telephone interviewing (CATI).

The baseline questionnaire consists of a common set of interview elements collected of all subjects shortly after consent. Questionnaire elements cover a wide selection of topics including historic height and weight, exposure to alcohol, quality of life, and access to healthcare.

Annual follow-up consists of questionnaire forms administered on an annual basis.

Enrollment Questionnaire

At the time of consent, a brief enrollment questionnaire is administered to record basic demographic details including name, date of birth, gender, race, ethnicity, language, education status, marital status, referring physician information and contact details. This questionnaire is administered by consenting staff in the clinic. Questionnaire data is recorded on computer tablets with a secure wireless connection and is stored in a study-maintained database.

Baseline Questionnaire (CDART)

After enrollment into the study, subjects complete a baseline questionnaire consisting of more than 20 constituent questionnaires measuring health, symptoms, quality of life, and related variables. These questionnaires have been validated in the literature and underwent testing and IRB approval. The baseline interview lasts between 45 minutes to an hour and can be completed over multiple calls.

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 (dependent on answering a lifetime alcohol consumption) in the alcohol information form as well as age-based skips 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 (CDART)

A series of cancer-specific questionnaires were developed to capture data on specific disease-related points of interest and are included in baseline and follow-up questionnaires

administered to study participants. Example areas include Breast, Prostate, Colorectal, and Hepatobiliary cancers.
Assignment of cancer-specific questionnaires occurs through an ICD-9 billing-code based classification. The results of this classification appear to interviewers in the Patient Tracking System. A manual method of assignment is also specified in the event that the billing code feed fails to correctly classify an individual's disease.

Annual Follow-up (CDART)

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 any subsequent cancer-specific information as the result of new diagnoses.

Section 5: Medical Records Abstraction

Medical records information comes from multiple source systems. Billing code data from the Carolina Data Warehouse for Health (CDW-H) is used at the beginning of enrollment to assign cancer-specific questionnaires based on subject-specific ICD-9 billing codes.

At approximately the same time that an individual visits the clinic, the hospital tumor registrar begins to record information from a subject's medical records in order to fulfill state cancer-reporting requirements. This data is electronically stored using a software product called ERS (Electronic Registry Systems, Cincinnati OH). As diagnostic, tumor staging, and course of treatment information become available, this information is abstracted into software that can format state-compliant reports.

The UNC Health Registry Study purchased an instance of the ERS medical record abstraction software in order to enable staff to view abstracted information on the subset of UNC Hospital patients who have consented to the Health Registry study and enter additional information on treatments received. (See Hand-Abstraction for more information).

Finally, the SAS-UNC partnership is working to developing a health outcomes analysis tool based using a secure medical workspace concept. This tool will link to UNC Healthcare data and enable researchers to discover analytic cohorts and run analyses in a secured environment.

CDW-H Billing Codes – Cancer Specific Questionnaire Assignment

Cancer-specific questionnaire assignment is driven by ICD-9 billing codes from UNC Hospitals billing systems and serves as an "initial diagnosis" until more specific information is acquired through pathologic data. A rule-based algorithm has been implemented to assign questionnaires based on ICD-9 billing codes.

Source data originates from CDW-H billing code reports. These reports contain medical record number (MRN) corresponding to each subject, the patient's date of visit, and ICD-9 code.

CDW-H Data Sources Used by the Billing Code/Questionnaire Assignment Algorithm

Hospital Past Diagnosis Report	All hospital-side billing diagnoses since 4/2009 for patients who were seen at a UNC cancer clinic in the last 14 days
P&A Past Diagnosis Report	All Physician/Associates billing diagnoses since 4/2009 for patients who were seen at a UNC Cancer clinic in the last 14 days

This data is used in conjunction with individual date of 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 220 specific ICD-9 codes.

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

Abstracted Medical Record Data Elements

Medical Record information on consented patients consists of demographic data, tumor staging information, treatment information, and follow-up information. The following table contains the patient and cancer-specific information for research purposes.

UNC Cancer Registry abstracted medical record variables	
Demographics	
	Patient UNCHR ID
	First name
	Middle name
	Maiden name
	Last name
	Name prefix
	Name suffix
	Date of birth
	Current county
	Ethnicity
	Sex
	Race
	Tobacco history
	Family cancer history
	Alcohol history
Diagnosis and Staging	
	Sequence of the primary
	Site code
	Topography code
	Grade
	Histology
	Date of diagnosis

	Diagnosis zip code Comorbidities Site description Tumor size Tumor extension/involvement Lymph node involvement Number of positive nodes Number of nodes examined Prostate tumor extension/involvement Pathology report number Primary payer /insurance Survival months since diagnosis Tumor record number General stage AJCC stage basis Pathologic T Pathologic N Pathologic M Pathology AJCC stage group Clinical T stage Clinical N stage Clinical AJCC stage group Clinical staged by TNM edition number Justification for clinical staging elements Justification for pathologic staging elements General stage text Type of source documents Diagnosis comments Operative findings Comments
Surgery	
	Surgery date Surgery type Scope of lymph node surgery Lymph node removed Surgery of regional sites First or subsequent surgery Treating hospital code

Treatment	
	Treatment date Regional treatment modality Boost treatment modality Chemotherapy Hormone/steroid therapy Immunotherapy Hematologic transplant/endocrine procedure Other treatment Palliative procedure First or subsequent treatment Treating hospital code
Recurrence	
	Recurrence date Recurrence type Recurrence metastasis site Cancer status Patient status Date of last contact Follow-up method
Other	
	Cancer status Patient status Date of last contact Follow-up method

Medical Record Hand-Abstraction

A set of fields is created in the ERS system to detail information on first course of treatment for a study subject. These fields are entered by UNC HR/CSC staff and appended to the ERS record.

Hand-abstracted fields include a variable to signify whether first course of treatment was performed outside of the UNC cancer hospital. In a portion of those cases, documentation from outside clinics are scanned and received by the hospital tumor registrar and are available for enumeration of treatments.

Because first course of treatment will vary by cancer site and other individual factors, the detail of each record will vary. The availability of information for an individual's treatment will influence the completeness of these hand-abstracted fields. In many cases, this additional abstraction will serve as a basis to validate treatment information found in other sections of the ERS record.

Hormone, immunotherapy, and chemotherapy agents involved in the first course of therapy are noted in appropriate sections of the free fields. A best estimate of treatment start date is recorded based on records. This estimate is based on record of dispensing/administration 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 it's own NSC number. Each drug in the regimen is recorded.

Treatment completion status is recorded and corresponds to reasons for interruption or suspension of treatment. Successful completion and unknown completion status are also recorded in this field.

UNC HR/CRC abstracted medical records variables*

*Chemo Agent 1 - 6***	
	Chemo Agent NSC Start Date End date Number of Cycles Cycle duration Completion status
Hormone Agent 1	
	Start date Hormone Agent NSC
Immunotherapy Agent 1	
	Start date Immunotherapy Agent NSC
Other	
	Treatments Outside of UNC (Binary) Date of HR abstraction

** ERS free data fields modified to include those listed in this table*

*** Data will be abstracted for multiple treatments when necessary (1,2,etc)*

Appendix 1. System-Specific Information

Patient 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.

Overview of BSP 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 (ie 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 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 atleast 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 which 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 bythe Autopure LS (DNA extracton robot) and the Multiprobe 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 the Health Registry 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 TPF LIMS System

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.

Electronic Registry System (ERS)

Electronic Registry System (ERS) is a Cancer Registry software tool that includes the 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 via an ERS developed software patch. 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 reabstracted to check for accuracy. A quality control protocol involving reabstraction 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)

Appendix 2. System-Specific Data Mappings

CPR-OR Tissue Collection Tracking

The UNC Health Registry TPF Report contains tissue banking status information that is reported in coarser granularity based on several specific banking status categories. This table explains the scenarios where tissues are not collected.

Status	Explanation	Tissue Status	On TPF Report
Banked	Tissue was received and processed	TISSUE COLLECTED	
After Hours	Cases that run past 5:30pm Monday through Friday. Cases are then targeted the next morning before they are put on formalin by pathology (with the exception Friday cases, after hours cases will be updated the next day)	Temporary status due to pathology hours. Final status if collected on a Friday.	
Surgery Stopped (updated from Aborted)	Surgery was stopped (this generally occurs when frozen sections from complex surgery cases show metastatic disease, and the surgeon decides that the patient is better off without continuation of surgery)	NO TISSUE COLLECTED	
Completely Submitted to Path	Cases in which tissue procurement is not allowed to take any tissue because all of the tissue is needed by pathology (examples include: lymphoma workups, spleens submitted to hemepath, medical kidneys, uterus cases with complex atypical hyperplasia, BRCA ovaries)	NO TISSUE COLLECTED	"Not Enough Tissue" (Sample sent to Surg Path)
Canceled	Cases that were on the schedule but were either cancelled or moved to another date	NO TISSUE COLLECTED	
Improper Consent	Cases in which HIPAA was not signed, the consent form was not signed, dates were invalid, or surgeon did not consent patient prior to surgery	TISSUE NOT COLLECTED	

Infectious	Case that are known to be infected with HIV, HBV, HCV, or TB are not banked by TPF	TISSUE NOT COLLECTED	
Margins	Wide local excision (WLE) melanomas or breast re-operative cases in which margins were previously positive are not banked by TPF	TISSUE NOT COLLECTED	"Not Enough Tissue" (Sample sent to Surg Path)
Biopsies	Scope cases which produce little or no tissue are not banked by TPF	TISSUE NOT COLLECTED	"Not Enough Tissue" (Sample sent to Surg Path)
Too Small	Cases in which the PA can see tumor, but there is not enough tissue to be given to TPF	TISSUE NOT COLLECTED	"Not Enough Tissue" (Sample sent to Surg Path)
No Tumor Seen	- Cases in which there is either no cancer diagnosis, or the tumor cannot be seen or palpated (examples include: abnormal mammograms, benign disease cases, and cases with only normal tissue) are not banked by TPF	TISSUE NOT COLLECTED	"No Tissue" (No Sample Sent to Surg Path)
Not Targetable Procedure	Cases that will not produce tissue	TISSUE NOT COLLECTED	
No Consent	Cases that blank consent form is found in the chart (recently updated)	TISSUE NOT COLLECTED	
Withdraw	Cases that subject revoked the consent signed prior the surgery (recently updated)	TISSUE NOT COLLECTED	