



**Southwest
Oncology Group**
A National Clinical Research Group

STATISTICAL CENTER

July 2004



FRED
HUTCHINSON
CANCER
RESEARCH
CENTER



**Cancer
Research
And
Biostatistics**

Introduction

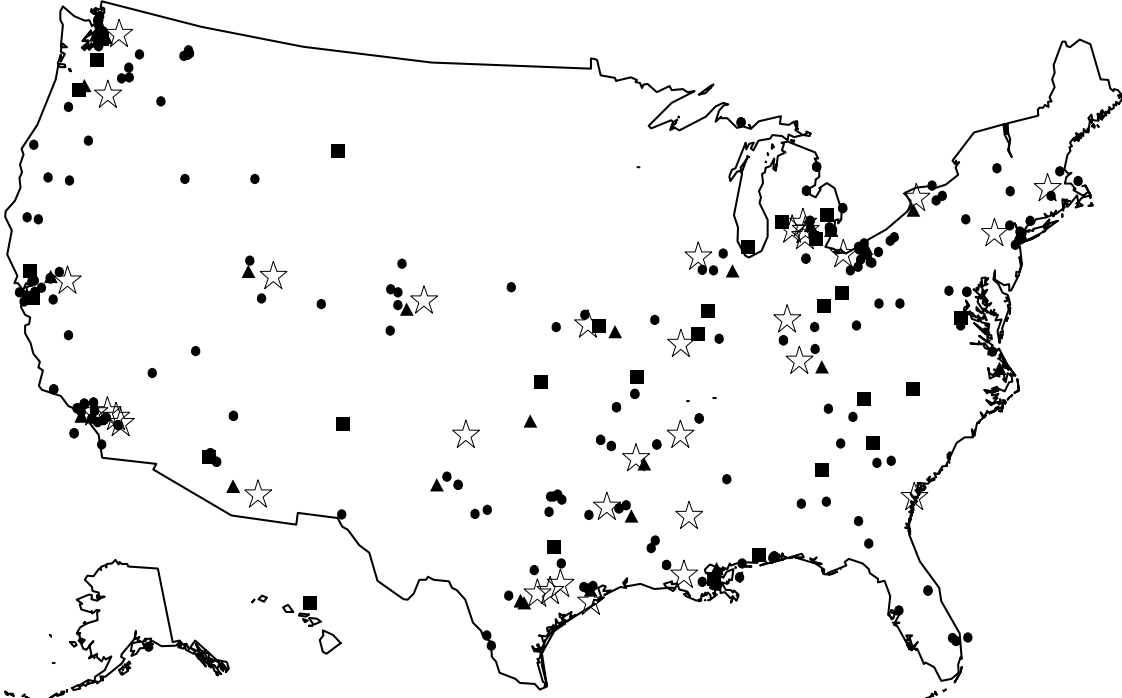
The Southwest Oncology Group is a national consortium of institutions and investigators organized for the purpose of improving the survival of cancer patients through clinical research. The Group began in 1956 as the Southwest Cancer Chemotherapy Study Group; it has expanded to include all modalities of cancer therapy, and institutions in all regions of the country. Most of the studies done by the Group are designed to assess whether a regimen merits further study (Phase II), or to compare two or more regimens (Phase III). Studies in cancer control research (prevention, symptom control or quality of life) are also carried out. The group is also coordinating two large studies aimed at preventing prostate cancer.

The Statistical Center of the Southwest Oncology Group is located at the Fred Hutchinson Cancer Research Center and at Cancer Research And Biostatistics in Seattle, Washington, having been moved from M. D. Anderson in Houston in 1984. The Director is John Crowley, Ph.D. The Operations Office of the Group is located at the Cancer Therapy and Research Center in San Antonio, Texas and is under the direction of the Group Chair, Charles A. Coltman, Jr., M.D.

Figure 1 is a map giving the location of the cooperating institutions. Member institutions are the academic centers, which formed the original nucleus of the Group. Affiliate institutions are groups of community physicians affiliated with a member institution. Community Clinical Oncology Program (CCOP) institutions are community hospitals or consortia with a mandate for both clinical and cancer control research. Urologic Cancer Outreach Program (UCOP) institutions are organizations specializing in the treatment of genitourinary cancers. Prostate Cancer Prevention Trial (PCPT) sites are involved in a large randomized chemoprevention trial testing the agent finasteride (see Figure 2). The Selenium and Vitamin E Cancer Prevention Trial (SELECT) sites are presented in Figure 3. Locations of patients registered to Group studies in 2003 are depicted in Figure 4, using zip codes of their residences. The organization of the Statistical Center is shown in Figure 5.

Figure 1

Southwest Oncology Group Institutions



☆ Member

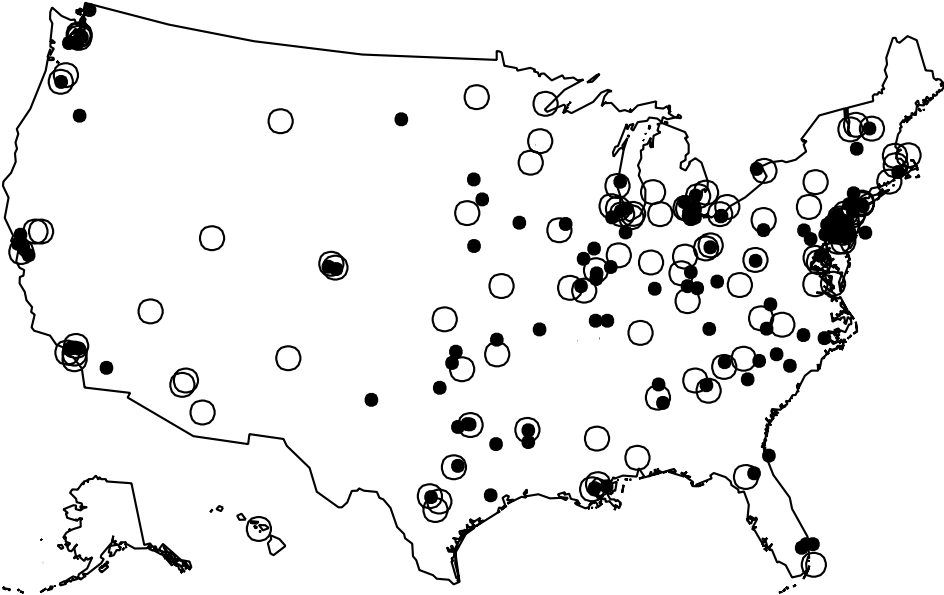
■ CCOP

● Affiliate

▲ UCOP

Figure 2

Prostate Cancer Prevention Trial Study Centers



○ Center ■ Site

Figure 3

SELECT Study Trial Centers



Figure 4

Southwest Oncology Group 2003 Registrations

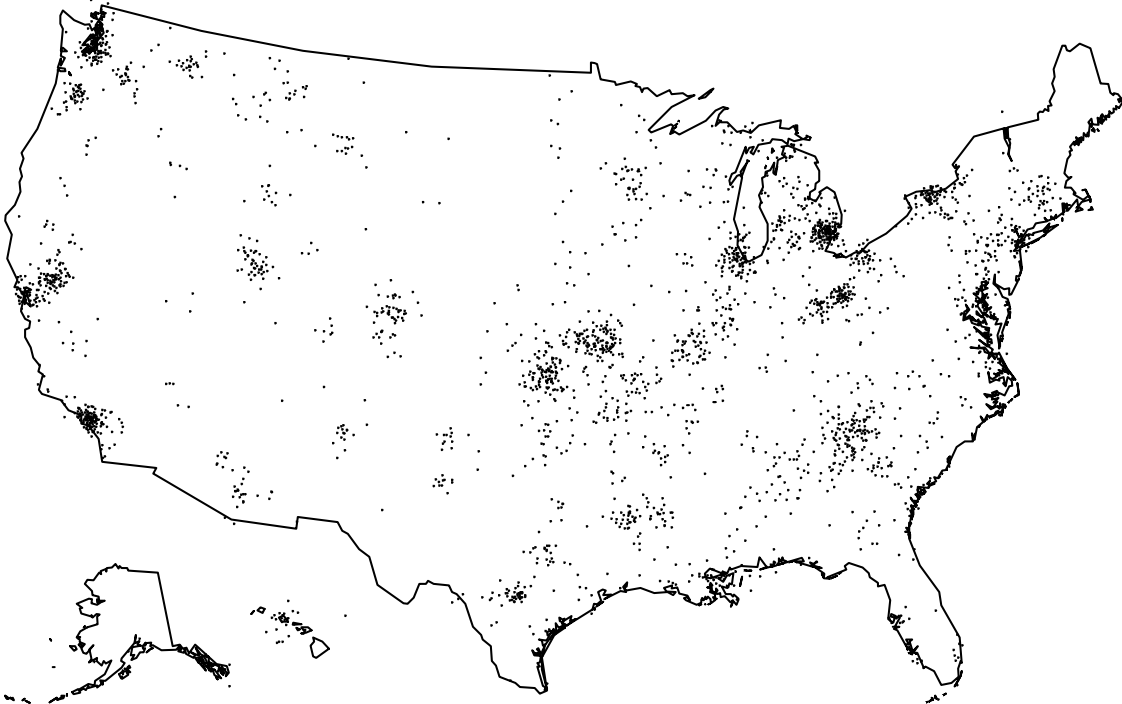
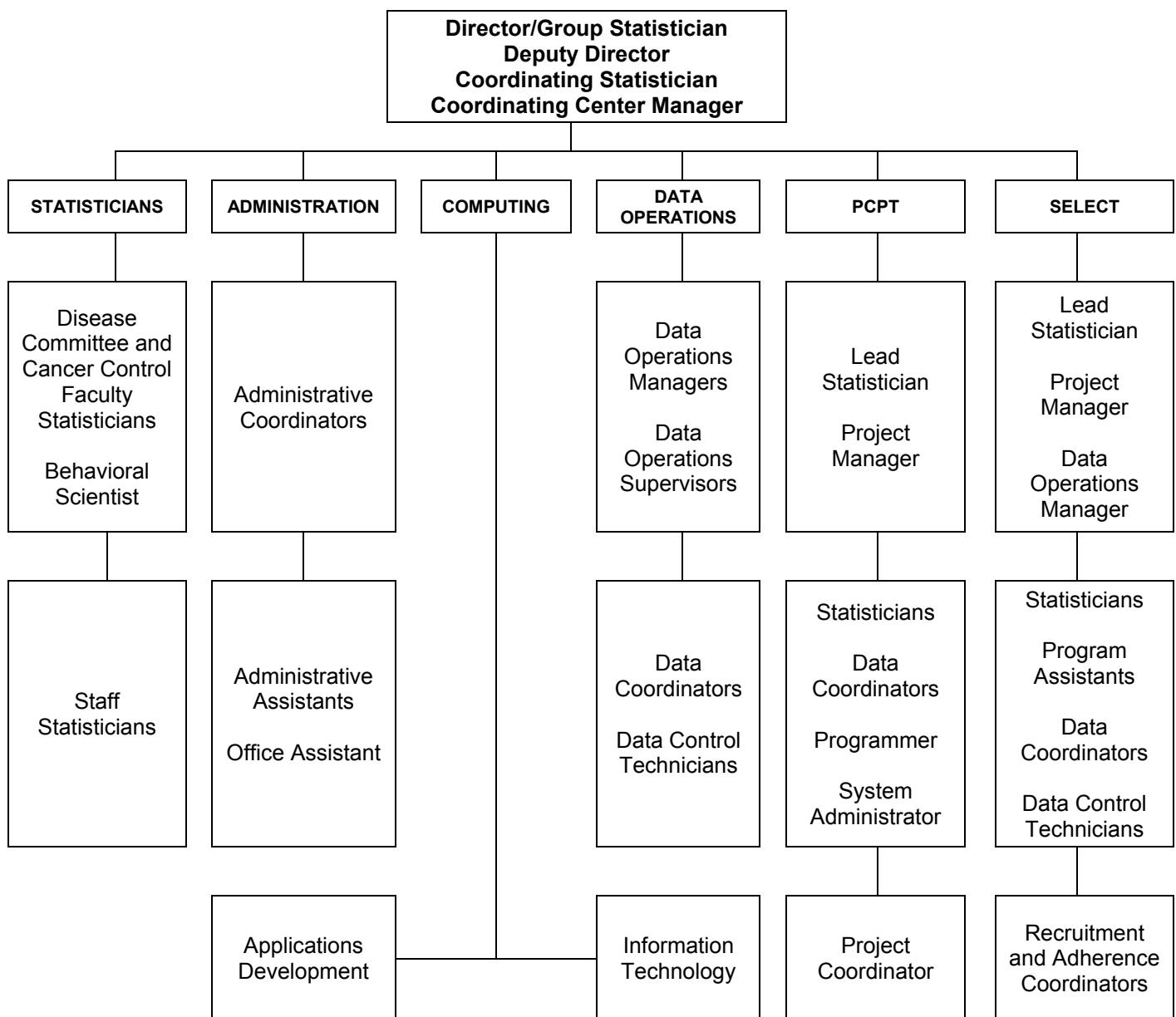


Figure 5



SOUTHWEST ONCOLOGY GROUP STATISTICAL CENTER

SWOG STATISTICAL CENTER ORGANIZATIONAL CHART



The Group consists of 34 member institutions, 28 CCOPs, 211 affiliates, and 26 UCOPs. In 2003, 3,579 patients were registered to clinical trials, cancer control studies and ancillary studies. There are currently 93 studies open for accrual.

Patient registrations over the past five years by study type, committee, and institution type are as follows:

**Table 1: SWOG Patient Registrations Over 5 Years
(Numbers Exclude Prostate Cancer Prevention Trial and SELECT)**

	Jan 98- Dec 98	Jan 99- Dec 99	Jan 00- Dec 00	Jan 01- Dec 01	Jan 02- Dec 02	Jan 03- Dec 03	<u>Total</u>
Total	4,392	4,468	4,039	5,493	4,242	3,579	26,213
Study Type							
Treatment:							
Phase I	11	0	5	25	0	0	41
Phase II	1,106	911	575	520	804	568	4,484
Phase III	2,334	3,018	3,047	4,378	2,677	2,364	17,818
Other	941	539	412	570	761	647	3,870
Committee							
Brain	96	69	21	35	61	70	352
Breast	923	1,581	1,178	1,621	784	798	6,885
GI	1,322	902	585	607	424	122	3,962
GU	319	457	1,011	1,029	1,018	863	4,697
GYN	118	118	158	185	119	129	827
Head/Neck	156	148	135	44	66	75	624
IMT	0	0	0	0	0	5	5
Leukemia	435	144	164	163	208	228	1,342
Lung	307	482	267	634	1,178	851	3,719
Lymphoma	292	198	295	280	163	231	1,459
Melanoma	178	169	90	111	140	103	791
Myeloma	201	154	94	30	25	75	579
ICAS (Sarcoma)	45	46	41	753	56	29	970
Institution							
Member	1,289	1,200	908	1,325	1,128	983	6,833
Affiliates	788	948	894	1,104	809	685	5,228
CCOP	1,138	1,412	1,148	1,648	1,239	923	7,508
UCOP	202	239	389	428	440	399	2,097

Registrations From Other Groups:	Jan 98- Dec 98	Jan 99- Dec 99	Jan 00- Dec 00	Jan 01- Dec 01	Jan 02- Dec 02	Jan 03- Dec 03	Total
ACSOG	0	0	0	0	2	0	2
CALGB	341	171	189	300	118	89	1,208
CTSU	0	0	0	15	134	121	270
ECOG	419	273	154	292	139	171	1,448
EORTC	0	17	107	87	73	67	351
EPP	0	5	20	33	28	3	89
GOG	3	1	62	90	0	0	156
NCCTG	86	65	63	69	68	37	388
NCIC CTG	49	58	43	75	40	97	362
NSABP	0	0	0	0	1	0	1
RTOG	77	79	62	27	23	4	272

Statistical Center Functions and Organization

The Statistical Center of the Southwest Oncology Group is directed by the Group Biostatistician, John J. Crowley, Ph.D., and is located at both the Fred Hutchinson Cancer Research Center (FHCRC) and Cancer Research And Biostatistics (CRAB) in Seattle, Washington. Dr. Crowley and most other doctoral level statisticians also have faculty appointments in the Department of Biostatistics at the University of Washington located approximately three miles from the Statistical Center.

CRAB is a non-profit organization founded primarily to support the work of the Southwest Oncology Group Statistical Center. Statistical review, data management, software development, and network administration for Southwest Oncology Group Institutions are jointly performed by the Statistical Center housed at both the FHCRC and CRAB. The entire Statistical Center—both at FHCRC and CRAB—were previously located in the same downtown Seattle building. However, in March 2004, the FHCRC-housed component moved to the new Public Health Sciences building on the FHCRC campus in the South Lake Union area, about a mile distant from CRAB. The two now physically separate staffs are connected electronically.

Statisticians and data coordinators are each assigned to at least one disease and/or discipline committee. Statisticians review all protocols for feasibility, experimental design, and the appropriate number of patients needed to answer the research objectives. Statisticians also perform analyses of study results for the semi-annual Report of Studies, as well as for the data and safety monitoring committee (where appropriate) and for publications. Data coordinators review protocols for clarity and consistency, register and randomize patients on protocols, review patient data forms for consistency and completeness, help in study monitoring, oversee mailings to Group participants, and coordinate training for Clinic Research Associates new to the Southwest Oncology Group.

Administrative support staff process incoming and outgoing data. Computer programmers maintain the Statistical Center's hardware and software and develop new software as needed to accomplish the Statistical Center's objectives. Administrative and clerical tasks are carried out by the coordinating center manager, administrative coordinator, administrative assistants, and office workers. Finance tasks are managed by the finance administrator and the finance personnel. Each staff member has access to the network to carry out a share of the activities.

Southwest Oncology Group Statistical Center (FHCRC) Staff

John Crowley, Ph.D., Director
Jacqueline Benedetti, Ph.D., Deputy Director
Catherine Tangen, Dr.P.H., Coordinating Statistician

Administration

Mark Blitzer, B.A., Administrative Coordinator
Tess Hurley, Finance Administrator
Jeffrey Zacko-Smith, M.P.A., Administrative Assistant

Behavioral Scientist

Carol Moinpour, Ph.D.

Biostatistics

Donna Ankerst, Ph.D.
Kenneth Kopecky, Ph.D.
Michael LeBlanc, Ph.D.
PY Liu, Ph.D.
James Faulkner, M.S.
Holly Gundacker, M.S.
Caroline Jiang, M.S.
Danika Lew, M.A.
Sheryl McCoy, M.S.
James Moon, M.S.
Cathryn Rankin, M.S.
Joseph Unger, M.S.

Prostate Cancer Prevention Trial

Phyllis Goodman, M.S., Lead Statistician
Susan Carlin, B.A., Project Manager
Chen Chi, M.S., Statistician
Amy Darke, M.S., Statistician
Katie Gower, M.S., Statistician
Shannon Hill, B.M., Adherence Project Coordinator

Data Coordinators:

Jessica Hussell, B.A.
Bernard Moore, B.B.A.

Programming:

Carl Benson, B.S.
Laura McDonald, B.A.

**Southwest Oncology Group Statistical Center
Cancer Research and Biostatistics (CRAB) Staff**

John Crowley, Ph.D., President and CEO
Evonne Lackey, C.C.R.P., Vice President and CTO
Catherine Crowley, Ph.D., Secretary-Treasurer and COO

Administration

Cheryl Angle, B.A., Administrative Assistant/Receptionist
Lori Schumacher, B.A., Administrative Assistant
Marcia Foster, B.A., Office Assistant

Applications Development

Ron Bredehoeft, Director of Applications Development

Angela Ribble, B.A., Project Manager

Andy Reynolds, Net Web Programmer
Ben Kleinman, B.A., Systems Analyst/Programmer
Carlos Marin, Systems Analyst/Programmer
David Law, B.S., Systems Analyst/Programmer
Darlene Davis, B.S., Systems Analyst/Programmer
Deborah Sopher, M.S., Systems Analyst/Programmer
Jennifer Farrell, B.S., Systems Analyst/Programmer
Kelly Balch, B.A., Case Report Forms Programmer
Kent Harris, Graphics/Web Designer
Marius Sorescu, B.S., Systems Analyst/Programmer
Mark Jones, B.S., Systems Analyst/Programmer
Teresa Chern, A.A.S., M.C.S.E., Systems Analyst/Programmer
Rick Mize, Case Report Forms Programmer
Vic Bredehoeft, Systems Analyst/Programmer
Yoko Toyama, B.A., Software Tester

Biostatistics

William Barlow, Ph.D., Statistician

Jason McCoy, M.S., Statistician
Kari Chansky, M.S., Statistician

Information Technology

Kelly Landreth, B.S., Director of Computer Services

Anthony McLaughlin, B.A., Network Administrator
Brian Hanners, B.A., Network Administrator
Brian Hartline, B.S., Network Administrator
Jane Xie, B.A., Database Administrator
Michael Fujinaka, B.S., Senior Network Administrator
Steve Dong, B.A., System Support Specialist

Data Operations Center

Rodney Sutter, Data Operations Manager

Diana Lowry, B.A., Data Operations Technical Manager

Camille White, B.S., C.C.R.P., Data Operations Supervisor
Laura Kingsbury, M.R.T., Quality Assurance Coordinator
Scott Kurruk, B.A., Data Operations Supervisor
Stephanie Edwards, Data Operations Supervisor

Data Coordinators:

Brian Zeller
Christine McLeod
Cynthia Hill
Janice Leaman
Jean Barce, B.A.
Jenni McNurlin
Jennie Barrett
Larry Kaye, B.A.
Lisa Gavigan
Tracy Maher, B.S.

Data Control Technicians:

Leigh Boelman, B.A., Discipline Coordinator and Data Control
Tech Lead
Amy Edwards, B.A., B.S.
Antoine Frederick, B.A.
Devin Kearns, B.A.
Dianna Mize
Julia Jackson
Karen Elliott
Richard W. Mize

**Selenium and Vitamin E Cancer Prevention Trial
(SELECT)**

Jo Ann Hartline, M.P.H., Project Manager

Phyllis Goodman, M.S., Lead Statistician

Amy Darke, M.S., Statistician
Chen Chi, M.S., Statistician
Jeff Probstfield, M.D., Medical Consultant
Karen Anderson, Recruitment and Adherence Coordinator
Katie Gower, M.S., Statistician
Monica Yee, B.A., Data Operations Manager
Russell Campbell, M.A., Minority Recruitment
and Adherence Coordinator
Sharon Moon, Recruitment and Adherence Assistant
Kathy Conner, B.A., Administrative/Technical Assistant

Data Coordinators:

Dona Marrah, Data Operations Supervisor

Matthew Scott, A.A.S., Data Coordinator Lead
Diane Liggett, B.S.
Jenny Peck, B.A.
Sarah Effert, B.A.

Data Control Technicians:

Claudia Vio, A.A., A.A.A.S
Ginnie Bauman
Karen Wesolowski
Michelle Marrah

Statistical Center Disease and Discipline Committee Teams

<u>Committee</u>	<u>Data Coordinators</u>	<u>Biostatisticians</u>
Brain	Lisa Gavigan	Cathryn Rankin
Breast	Christine McLeod Cynthia Hill Diana Lowry Lisa Gavigan	William Barlow Caroline Jiang Danika Lew
Gastrointestinal	Christine McLeod Jennie Barrett Rodney Sutter	Jacqueline Benedetti Sheryl McCoy
Genitourinary	Brian Zeller Janice Leaman Jean Barce Jenni McNurlin	Catherine Tangen James Faulkner
Gynecology	Jean Barce	PY Liu Caroline Jiang
Head & Neck	Cynthia Hill Jenni McNurlin	Michael LeBlanc James Moon
Leukemia	Tracy Maher	Kenneth Kopecky Holly Gundacker
Lung	Camille White Janice Leaman Larry Kaye	John Crowley Kari Chansky Jason McCoy
Lymphoma	Scott Kurruk Tracy Maher	Michael LeBlanc Joseph Unger
Melanoma	Brian Zeller Scott Kurruk	PY Liu James Moon
Myeloma	Diana Lowry	John Crowley Jason McCoy
Intergroup Coalition Against Sarcomas	Stephanie Edwards	Cathryn Rankin John Crowley
Cancer Control Research	Diana Lowry	Donna Ankerst Caroline Jiang
Correlative Sciences	-----	Donna Ankerst Caroline Jiang
Pathology	Leigh Boelman	Holly Gundacker
Radiation Therapy	Leigh Boelman	Danika Lew
Surgery	-----	James Moon
Bone Marrow or Stem Cell Transplant	-----	Kenneth Kopecky
Special Populations	-----	Joseph Unger
Cytogenetics	Jennie Barrett Tracy Maher	Holly Gundacker
Immunomolecular Therapeutics	Jean Barce Laura Kingsbury	PY Liu James Moon
Early Therapeutics	Laura Kingsbury	PY Liu Sheryl McCoy

Primary Mission of the Statistical Center

The primary mission of the Southwest Oncology Group Statistical Center is to make progress in the prevention and treatment of cancer through clinical research. The mission is accomplished through the conduct of important trials and through translation of biologic concepts to clinical care. Quality research, quality data, and publication of results are critical to the effort. The Statistical Center contributes through the following:

- **Study Design**
The Statistical Center has a fundamental role in clarifying study objectives and in designing statistically sound studies to meet those objectives.
- **Protocol Review**
The Statistical Center reviews all protocols for logical consistency and completeness, in order that study conduct not be compromised through use of an inaccurate protocol document.
- **Data Quality Control and Study Monitoring**
The Statistical Center continually enters, forwards to study coordinators, reviews, corrects, updates, and stores data from all active Southwest Oncology Group studies, in order that study results not be compromised by flawed data and that studies be monitored for patient safety.
- **Analysis and Publication**
The Statistical Center is responsible for statistical analysis and interpretation of all Southwest Oncology Group coordinated studies and all Southwest Oncology Group database studies.
- **Statistical Research**
The Statistical Center has an active research program addressing unresolved design and analysis issues important to the conduct of cancer clinical trials and to ancillary biologic studies.
- **Training**
The Statistical Center plays a key role in the training of new clinical research associates (CRAs) and of Young Investigators in the Group.

Main Objectives of the Statistical Center

1. To participate in the development of proposed protocols, particularly as regards experimental design, sample size and feasibility. Biostatisticians work with the study coordinator on the statistical aspects for each protocol, including a specification of the major objectives and the number of patients required to meet those objectives. Data coordinators and biostatisticians comment on concept sheets for proposed studies and work with protocol coordinators in the Operations Office to produce protocols that are concise and clear.
2. To provide for registration of all patients on all studies, and for randomization of patients where appropriate. Registrations for all studies are available by direct Internet Web-based connection to the WebReg program.
3. To develop Web and software technology for paperless submission of data to the Statistical Center.
4. To provide for review and quality control of data collected during studies. Data coordinators screen all incoming data and query institutions regarding any incompleteness or inconsistency. Further range and logical checks are made at data entry. Data are collected on timeliness and accuracy of data submitted by each institution, as part of an effort to improve quality throughout the Group.
5. To provide for data entry, and for computer processing, and storage and retrieval of data. Data entry is performed by trained data control technicians. There are three data entry submission routes: paper, fax or the Internet.
6. To work with the Group Chair and other investigators in the Group to improve the quality of clinical trials through the use of improved data forms, uniform and reproducible data definitions and economical data flow, and efficient use of Statistical Center resources.
7. To assist the Group Chair in the administration of the Group.
8. To analyze and publish the results of studies in conjunction with the study coordinators. Improvements are continually being made in the analytic and design tools available to the biostatisticians. The main analytic tool is SAS™, with locally developed software to extract SAS™ files from Oracle™ and to produce tables and survival curves for the semi-annual Report of Studies and for publication.
9. To use the data collected to try to find new leads regarding prognostic factors and late effects.
10. To perform statistical research on the efficient design, conduct and analysis of cancer clinical trials and cancer control research. In particular, research is being done on the analysis of survival data, on design and analysis strategies for clinical trials, on monitoring strategies for Phase III studies, on analysis of longitudinal data subject to non-ignorable missingness, and on methods for the analysis of high-dimensional cytogenetic, microarray and proteomic data.

11. To educate investigators, nurse oncologists and CRAs in statistical analysis, research design and the utilization of the most advanced scientific and data management strategies.

Major Accomplishments

In fulfilling these objectives, major accomplishments during the years 1998-2003 were:

1. Developed a new Web based system to allow direct registration by institutions. Expanded this system to allow use by members of other Cooperative Groups, and registration to studies conducted by other Groups.
2. Developed a Web page providing access to the Report of Studies, accrual reports, training manuals and other research information.
3. Offered a training program for experienced clinical research associates on the use of the new Web tools.
4. Improved training. Developed a formal training program and a Data Operations Procedure Manual for new data coordinators. Revamped the training program and updated the manual for institutional clinical research associates. Significantly improved Standard Operating Procedures (SOP) documentation.
5. Contributed to the development and conduct of a Young Investigators Workshop, an intensive training program designed to teach clinical trials principles to new researchers.
6. Revised the program for training new biostatisticians, and instituted a biostatisticians' meeting to discuss procedures, policies, and methodologic issues.
7. Developed an improved system for documenting and tracking serious adverse events (SAEs), formerly referred to as adverse drug reactions (ADRs).
8. Simplified all forms in response to the need for efficiencies in the clinical trials process.
9. Participated in a national cooperative effort to streamline data collection through the Common Data Element project initiated by the Cooperative Group Chairs and Group Statisticians, and supported by the NCI.
10. Served on external committees in the form of faculty participation on data monitoring committees, NIH review panels, American Joint Commission on Cancer task forces, development of CTC and RECIST definitions, editorial boards, etc.
11. Adopted a new data submission and entry system that allows data to be scanned, faxed or submitted via the Web for direct entry into the database.
12. Developed an electronic patient chart system.
13. For intergroup communication, initiated collaboration with other cooperative group statistical centers to allow direct data submission on intergroup trials. Worked with CTSU on regulatory and data submission procedures.
14. Revised guidelines for the design and assessment of quality of life on Southwest Oncology Group studies.

15. Continued development of the database management system, including more flexibility in handling multiple registrations, cycle-specific data collection, extensive additional logic checks at the time of entry of patient evaluations, and enhanced capability for managing double blind studies in anticipation of planned prevention studies.
16. Met responsibilities with respect to the conduct, design, and analysis of Southwest Oncology Group studies.
17. Made advances in statistical methods for therapeutic and prevention trials, for quality of life studies, and for the analysis of high-dimensional data, as evidenced by presentations at national meetings and publications. Developed and improved tools for the design of trials and made them available on the Web.
18. Continued the development of analytic software, including the Statistician's Report Worksheet (SRW) for producing the Report of Studies, and programs for exploratory and longitudinal data analysis.
19. Managed follow up of multiple large intergroup trials, including SWOG 8814, 8897, 9313, and 9304. Managed a number of successful Phase III trials, including an intergroup sarcoma trial, S0033, testing two doses of Geevac in GIST tumors.
20. Completed accrual to a large chemoprevention study, the Prostate Cancer Prevention Trial (PCPT). Study results were published in the New England Journal of Medicine in July 2004.
21. Opened a second cancer prevention study, the Selenium and Vitamin E Cancer Prevention Trial (SELECT), which accrued a total of 35,534 men between activation in August 2001 and the end of accrual in June 2004.
22. Participated with a group of Japanese investigators in a series of clinical trials workshops.
23. Published a book on clinical trials, Clinical Trials in Oncology, co-authored by three Statistical Center faculty. Offered short courses based on this text. A second edition of this text appeared in 2002. Edited a Handbook of Statistics in Clinical Oncology, with contributions from five Statistical Center faculty.
24. Joined a consortium of SWOG researchers in the development of an Early Therapeutic Program for the Group. Developed a new Web program to allow rapid data transmission for this initiative.
25. In conjunction with the Group Operations Office, expanded the Web security systems to require individual password access to the Group Web site, thus allowing more detailed Web-based communications between institutions and the statistical center.
26. Initiated development of a system for collection and transmission of data from off-site laboratories and institutions to enable better tracking of pathology specimens and more consistent reporting of results.
27. Moved the cytogenetics business office to the Statistical Center.

28. Developed an internal Quality Assurance program to audit the data coding quality of Data Coordinators, and to provide continuing education as needed.
29. Established a correlative sciences statistical team to support the increasingly complex analyses involving high dimensional data.
30. Managed 26,213 registrations and 31,613 patients in follow-up.

Computing Infrastructure

Information Technology Environment

The Statistical Center computing resources are based on the Microsoft Windows 2000/2003 network. The key services include the patient databases, E-mail, WEB, application and file sharing, electronic document and image management, Cardiff Teleform[®] data forms design and submission, batch application processing, disaster recovery, virus protection services, Citrix-based terminal services, remote access, desktop configuration and network monitoring/management.

Desktop Systems and Network

Each staff at FHCRC and CRAB has an Intel Pentium III or higher system. Each desktop PC runs MS Windows 2000 Professional or MS Windows XP Professional. Over 35 desktop applications are supported. All workstations access the Statistical Center servers and the Internet. The Statistical Center database, WEB, file and related services are housed at CRAB. Statistical Center staff at FHCRC access CRAB resources (over a dedicated network connection to improve performance and security) using Citrix terminal services and Internet Explorer. Sensitive information is transmitted using encryption. In addition to the dedicated network link between CRAB and FHCRC, each organization has independent connection to the Internet.

MS Windows Servers

The Southwest Oncology Group Statistical Center uses a modified, distributed architecture with the majority of its MS 2000/2003 Servers, i.e. different servers perform different functions. There are over 50 network servers including a small number of development and test servers. Production servers are Pentium III or higher servers with multiple processors, a minimum of 1 Gigabyte (GB) of memory, have fault tolerant disk subsystems and carry on-site maintenance. Several tape library systems are used for backup.

Key network server services offered at the Statistical Center include:

- **Oracle[®]** Database
- **Microsoft[®]** SQL Server
- **Microsoft[®]** Exchange Post Office
- **Microsoft[®]** IIS WEB Servers (Internet and Intranet)
- **Lyris[™]** Listserv
- **Ecora[©]** Auditor
- **TNT[©]** ELM Event Monitoring
- **Cardiff[™]** Teleform[®]
- **Citrix[®]** Terminal Services
- **Camellia[©]** C/S Batch Services
- **Veritas[™]** Backup Exec Enterprise
- **Trend Micro[™]** Virus Protection
- **Symantec[™]** Ghost Enterprise and WinInstall
- **Microsoft[®]** System Management Services
- **SAS[®]** and **Splus[®]** Statistical Packages

Data Security and Disaster Recovery

Data Security and Disaster Recovery is based on what are often referred to as “best practices” in electronic computing and networking. CRAB Network Administrators periodically review and compare current network and security best practices with existing CRAB policies and procedures. Outside professional reviews and audits also provide critical information. Updates to policies, procedures and training are incorporated as appropriate.

A summary of key Statistical Center network policies and procedures requirements are:

Contingency Plans

All servers are backed up to tape (full and incremental). Backup media are transported to and from a bonded, secure off-site storage facility (vault) on a daily and weekly basis. All tapes stored on-site are located in fireproof cabinets in restricted areas. The most recent, and critical backups are stored in fireproof safes with combination locks. Original software, other media and directions are stored in centralized fireproof cabinets.

Servers are configured using fault tolerant disk subsystems. Special disk imaging and file recovery applications are used to speed up restoration on key file servers including database systems. Not only has this process been tested, it is used as a part of normal server construction and upgrades.

Emergency mode plans cover varying levels of disaster recovery developed to address the severity and extent of disaster. This may include a combination of manual and electronic replacement systems until such time critical network services can be re-instated to a fully working level.

Information Access Control

Supervisors submit Employee Action Forms (EAF) to key account management staff for all employee hires, terminations, or job function changes for staff. User accounts are **user-based**. Individually identifiable patient data, collected only at the time of registration, are stored in a secure table where read and write access is highly restricted and requires prior approval of the Group Statistician, based on demonstration of the user's "need to know" to perform their job functions. All accounts have controlled access to resources based on individual user account definitions.

The highest-level account for network administration is renamed/changed periodically to increase security. For network and database accounts, all staff are required to use “strong” passwords (those with a combination of lower and upper case alphabetic characters, numbers and special characters) to reduce the risk of password cracking. Passwords are aged and staff is required to change them on a regular basis. Passwords are not transmitted via E-mail or over the Internet.

The CRAB Firewalls provide restricted access to and from resources and are monitored 7 X 24. Verisign® Secure Socket Layer (SSL) encryption is used on Web and E-mail servers. Secure File Transfer Protocol (SFTP) is used for file transfer with remote clients.

Information Technology Security and Monitoring

The CRAB Security and Disaster Recovery Handbook along with the Information Technology Standard Operating Procedures define overall electronic security policies and procedures for the Statistical Center Resources.

Senior level Information Technology staff at CRAB are responsible for monitoring and addressing network security and host/server resources, including the patient database. All Information Technology support staff have appropriate levels of supervision.

Network administrators review network and server event, security and application logs and other reports on a daily basis to monitor login, file access, security incidents and the status of hardware and services. Notification software is configured to provide immediate notification (paging and E-mail) to network administrators for unexpected network and server events.

Servers and workstations are proactively updated with security patches as well as OS and application updates. Network Administrators subscribe to notification lists to stay on top of emerging problems and corresponding updates or fixes.

All desktop computers, servers and the E-mail post office have active, real-time virus protection. Virus protection software is automatically pushed to the CRAB network and subsequently updated on each system.

Computer equipment (including equipment checked out to staff) is logged and tracked in an online inventory database including location and user (for individual desktops).

Identified real or suspected security incidents are logged, addressed and reported to and/or by senior network administrators and the Director of Information Technology. Problems are in turn reported and/or escalated to CRAB Executive Offices and the CRAB Professional Conduct Board as appropriate.

Senior management including senior network administrators perform risk assessment on new systems and events to define the cost and benefits of different solutions and the solution impact on: Confidentiality, Integrity and Accessibility.

Media Controls

All software media and licensing are filed in fireproof file cabinets and restricted areas. Only authorized personnel may access original software media and licensing. All software upgrades to workstations, servers or other systems are done by Information Technology staff or by explicit permission of the Information Technology Director (in very rare situations, some users may install specified software licensing).

All on-site backup media are stored in either a fireproof safe or fireproof cabinet located in restricted access areas where only authorized staff may enter. All off-site backup media are stored in a secure, bonded, protected vault at professional facilities.

All old server and PC disk drives, CDs, portable media and other storage material are destroyed by a bonded, professional media destruction company.

All server rooms, server configuration areas and media storage are in restricted access locations. Signs are clearly posted on restricted areas noting that they are restricted. Only a

limited number of authorized staff may enter restricted areas. All server rooms and server configuration areas have security cameras, which record 7 X 24 hours per week. All servers have screen/keyboard locks. All vendors or other third party visitors accessing restricted areas are logged into an electronic file and escorted by authorized staff.

Information Technology Resource Access Policies

All staff are required to read and sign software, network, computer and data protection policies. These policies clearly outline and define proper use of desktop computers, E-mail, access to other network services, software usage, protection of patient information and other related practices. Operating System policies may further restrict staff from inappropriate computer or software usage. Network software monitoring applications track software use by user and workstation. All users must have valid accounts and passwords. Logon sessions have enforced password protected screen savers that lock the systems after 30 minutes of inactivity. Staff workstations are located in specific work areas that are locked after normal business hours. Remote and local users must maintain integrity and confidentiality when accessing CRAB resources.

Data Authentication and Encryption

The use of encryption (VPN, SSL and SFTP) reduces the risk of alteration or easily viewing of Internet traffic (packets) containing sensitive information. Operating system and database controls restrict inappropriate access privileges to data, files and other objects that require protection from modification.

Software

The operation of the Southwest Oncology Group Statistical Center depends on six major classes of software: database management, statistical analysis, desktop applications, network services, report processing, and data management.

Database Management

The database management software used is Oracle, one of the major commercially marketed systems. Oracle is based on the relational model of database management and is built around the industry-standard SQL language. The Statistical Center's data management operation is built around Oracle's capability for multiple users to manage simultaneous database modifications. In addition to the core relational database management module, Oracle has components for ad hoc queries, report writing, generation of screen-based data maintenance applications, interfacing to high-level languages (C++ or Visual Basic, for example), and database administration and tuning.

Statistical Analysis

The main statistical package used is SAS™. Several in-house programs have been written using SAS™ and Splus™ to perform tasks such as Cox regression diagnostics, Kaplan-Meier survival curves, sample size computations, exact methods, recursive partitioning, and longitudinal data analysis. The Center has recently acquired licenses for GenePlus and Insightful ArrayAnalyzer software for the analysis of microarray data.

Desktop Applications

Several Microsoft desktop applications are used by staff including Office Professional (Word, Excel, PowerPoint and Access), Visio and Project. Another key application is Adobe Acrobat.

Network Services

Electronic mail is used extensively for communication within the Statistical Center as well as with the Operations Office and with other Group members. CRAB and FHCRC staff access respective Microsoft Exchange Post Offices with MS Outlook and Outlook Web Access (OWA). The Statistical Center uses MS Internet Information Server (IIS) for Web services. Cardiff Teleform® is used for forms design, data submission and data entry.

The Southwest Oncology Group Home page can be found at <http://swog.org>. This site is maintained at the Operations Office in San Antonio. The swog.org website has links to the Statistical Center's services.

Report Processing

The Statistical Center Report of Studies is created using an application developed at the Statistical Center (Statisticians' Report Worksheet, or SRW) incorporating a web-based interface, creation of a SAS data set from Oracle, and the word processing tools of Microsoft Word.

SRW is based on a "thin" Client/Server (C/S) model using Web publishing technology. Web pages are driven from two primary database sources: Oracle and SAS data sets. MS Internet Explorer provides a program interface for input of textual and study parameters needed to define and set up charts, tables, graphs, and descriptive information. SAS extracts the data from the patient database via Open Database Connectivity to create a SAS data set, i.e., a "snapshot" of the patient data. Study chapter generation is done on a Web server based on input from the SAS data sets, study information defined in the Oracle database and end user input (such as text, label definitions, and table format information).

Users are able to view and output the study/chapter results in three ways:

1. As Web pages for preliminary browsing/viewing of the document.
2. As Web pages for final publication.
3. As formal output to a printer, a postscript file, and other file format, e.g., MS Word or Portable Data Format (PDF), for professional printing of the Report of Studies. PDF copies of the Report of Studies are available on the SWOG Home page (<http://swog.org>).

Hyper Text Markup Language (HTML) templates represent the various tables/charts. Microsoft's IIS Web Server and some of its key components are used to provide Web and final publication. A program process fills in much of the templates based on the SAS data sets and other stored database information, in order to generate more complete HTML

documents. Additional programming filters further refine the Web pages, and the formal output of the Report of Studies (ROS) is an executed "object linking and embedded-enabled document production manager", as the final ROS requires more extensive formatting (headers, footers, page numbers) compared to the Web pages.

The study chapters are made available as static publications based on the "snapshot" data sets. Data sets, other interim priority documents and the final chapter output are archived for future retrieval and reference. Since this model is a mix of "thin" C/S and Web publication with some portions being batched off to back-end Windows servers, it works well for both Local Area Network and Remote Access clients.

Security is built into the systems: correct user account information and password protection are required for accessing these services and firewalls are used to audit and restrict access.

Other in-house reports are created using SQL queries and Microsoft Active Server Pages (ASP) technologies to provide dynamic links to the Oracle database. Production reports and other documents are created using Microsoft Word, PCTex and Scientific Workplace.

Data Management

The main data management systems have been written in-house and consist of the patient/participant registration and randomization system (WebReg), the patient evaluation system (EVE), and Chart Manager, for the creation, manipulation and viewing of electronic patient charts. These programs all run on Windows, and are being re-written in .NET.

WebReg (Patient Registration)

The patient registration and randomization program is very complex, since it is a generalized program that facilitates study set-up and patient registration. It is designed to handle the wide variety of requirements for SWOG studies. It is a thin client application designed to be used in-house or via the Internet by a CRA. It is written in HTML with one C++ routine to interface with the Oracle database. Where this system is unavailable, registrations are also performed by telephoning a data coordinator in the Statistical Center who is online to the WebReg program. Randomization is typically accomplished using a dynamic balancing algorithm, which uses current accrual counts by stratification variables directly from the database. The Southwest Oncology Group Web registration program is also able to perform direct Internet patient registrations to intergroup studies coordinated by ECOG.

EVE (Patient Evaluation System)

The patient evaluation system allows data coordinators to update patient data such as toxicity, response, and vital status. Its design also allows the capture of data unique to a particular protocol. This system is written in Visual Basic and runs on Windows. A major function of EVE is to provide cross-field edit checks for evaluation data fields.

Chart Manager (Image Management)

Several years ago, the Statistical Center made the decision to move towards electronic data image management for our therapeutic studies. Because we did not want to maintain

both paper and electronic systems into the future, we undertook the scanning of 80,000 paper charts (2.2 million images), now completed. We now create, maintain and view electronic charts using an in-house application (Chart Manager), which is fully integrated with MS Windows and Oracle. All Teleform[®] patient file images and future data entry documents are included in the document management system (as TIFFs).

All SWOG data coordinators have client query tools for viewing patient charts, query and annotation utilities, and redaction (for HIPAA compliance). WebReg and EVE integrate directly with Chart Manager. Future enhancements will include the ability to view electronic data in other than TIFF format, as is being developed for the Selenium and Vitamin E Cancer Prevention Trial (SELECT).

Database

The Southwest Oncology Group Statistical Center manages its database using Oracle[®], a relational database management system. There is a production database that stores data that are reflective of real events, and a test database that is used for ongoing development and testing of applications. Each database is organized into two schemas, one for staging tables and one for active tables. Active tables hold data that are included in evaluations and analysis. Staging tables are used to store submission attempts from electronic data entry (EDC) applications, attempts that may violate pre-determined business rules. Once the business rules are cleared, data from the staging tables are promoted to the active tables.

The table structure in each database is organized into five main components:

1. **Common Patient:** This component contains patient-related data items, which are common to all Group studies including patient characteristics; identification of investigators and institutions to which a patient is associated; registration date; stratification; treatment; common evaluation items; and adverse events. Many of these items are accessed frequently in the day-to-day operations of the Statistical Center.
2. **Study Characteristics:** These describe the Group studies and are available to the data operations software, which must modify its behavior depending on the study being processed.
3. **Membership:** The membership component describes the investigators, clinical research associates, institutions, pharmacies, labs, radiation therapy facilities, bone marrow transplant facilities, and the relationships between people and sites. We further describe members by the web site and registration permissions they have for each site affiliation.
4. **Detailed Patient:** Detailed committee or study specific patient data items are regarded as a separate database component. These are not accessed as frequently as the common patient data items.
5. **Quality Control:** A major aspect of quality control data is our generalized tracking system, which stores data Expectations, and tracks submission of the required study information. We also maintain data from various review processes (pathology, surgery and radiation therapy). In addition, the data operations software forces quality control standards and collects quality control data.

Following are descriptions of some of the major tables in the database. The indications of size are given as rounded numbers from June 2004. These descriptions are in relational database management terminology. The relational model can be thought of as organizing data into tables that can be linked to other tables using key variables. For example, patient number is a key variable used to link prestudy data, adverse event data, and patient registration information (all residing in separate tables) for a single patient.

PAT table

The PAT table contains data on the patients enrolled into Group studies. At present, PAT contains data on over 160,000 patients. The columns include the following identifiers: patient number, birth date, sex, race/ethnicity, date of last contact or death, vital status and current follow-up investigator and institution. Confidential patient data, such as name and social security number are stored in this table under an extra layer of security, to which only approved personnel have access.

REG table

The table REG has one record for each registration. A patient may be registered to more than one study and a study may have more than one registration step. At present, there are over 225,000 registration records in REG. The columns include registration date, the institution and investigator credited with the registration, assigned treatment arm, and indicators for various review requirements (surgical review, for example). In addition, REG is the place where common patient evaluation data are stored for most studies. These include eligibility status indicators, treatment dates, response and relapse data, treatment deviation indicators, adverse event evaluability data, and treatment status indicators.

TOX table

The table TOX contains one row for each adverse event reported. At present, TOX has over 500,000 rows. Columns include cycle identifiers, adverse event, severity, and attribution.

STUDY table

The table STUDY contains one row for each possible type of registration. At present, STUDY contains data on over 950 studies. Approximately 115 registration types from 88 studies are open to accrual. The columns of STUDY include study and registration identification, study name, activation status, phase, study characteristics, review requirement indicators, and dates opened and closed (if applicable). The table STUDY is part of a large complex of tables, which describe various parts of studies, such as treatment and stratification information.

ROSTER table

The table ROSTER consists of one row for each investigator, clinical research associate, affiliate or individual for whom the Statistical Center requires contact information. There are over 23,000 rows. The columns include name, address, phone number, fax number, and email. ROSTER is the central table in a large collection of tables, which provide lists of in-

investigator subsets. For example, table INVEST provides the list of investigators and their current Group affiliations.

EXPECT table

Each row of table EXPECT contains data on an expectation for forms or material submission from an institution for a patient with respect to a particular study. EXPECT is a very large and active table, currently with over 5.8 million rows. The columns include the posting date, the due date and the resolution date.

AUDIT TRAIL table

In October 2000, a comprehensive audit trail was implemented for the entire production database. Every insert, update, and deletion for data from the production database is recorded in the audit trail. The audit trail records the user name, terminal name, date and time of the change, table name, unique identifier for the row, column name, and the old and new value.

Other tables

The detailed patient data component of the database consists of a collection of over 700 tables. Many are study specific, or are applicable to a class of studies (breast cancer studies, for example). These tables generally have many more columns than the tables in the other database components; several have over 200. Many contain prestudy (baseline) data, but others are used for special detailed event data or detailed pathology data.

Data Operations

Patient Registration

All patients are registered using a computer program (WebReg), which interacts with the database during the registration. (Well-defined manual registration procedures exist for those times when the computer system is not available.) Registrations are performed either directly via the Web by institutional CRAs or by data operations staff using WebReg during the telephone conversation with the representative from the institution registering the patient. The main advantage of using WebReg is that the registration procedures are strictly enforced in a uniform manner for all registrations. Some of the other advantages and features of WebReg include:

1. The database is instantaneously updated with each registration.
2. There is validation of the status of the investigator and institution registering the patient. The approval status of the facility to perform radiation therapy (if applicable) is validated.
3. There is a review of previous registrations in cases where the patient is being registered to additional protocols or subsequent parts of a protocol.
4. Blinded treatment assignment is possible.
5. Registrations requiring randomization are dynamically balanced using the current database stratification counts. Post-registration modifications to stratification data instantaneously adjust the dynamic balancing algorithm.
6. Relevant study characteristics and special notes about the study, including stratum or treatment specific notes, are automatically presented during the registration.
7. A confirmation of registration is generated with copies for all relevant study participants. This confirmation includes a summary of the registration along with a list of expectations for forms submission.

WebReg is driven by a study description file (SDF), which contains the information needed to control the registration process and to load the database study description tables. There are one or more SDFs for each study. These are prepared by the data coordinator for that study with the consultation of the data coordinator's supervisor, the study statistician, and discipline data coordinators and biostatisticians when applicable.

Flow of Data

A single set of data forms for SWOG coordinated trials is mailed, faxed or web-submitted to the Statistical Center from the institutions where the data are collected. For non-SWOG coordinated intergroup studies, institutions submit data directly to that coordinating statistical center. Data forms are processed according to the following conventions:

- Illegible forms or those lacking required identifiers are returned to the institution.
- Faxed or mailed forms are routed to the data control technicians. Study Coordinators are mailed copies on an event-driven basis.
- Slide and/or block information (as applicable) for pathology reviews are entered into the database by the discipline coordinator. Materials are tracked in the database through the review process. All results (eligibility and detailed review results) are also entered in the database.
- Complete sets of radiation therapy materials are sent either directly to the Quality Assurance Review Center or to the central RT reviewer. The review results are entered into the database by the discipline coordinator at the SWOG Data Operations Center.
- Surgical summary forms (as applicable) are forwarded to the surgery study coordinator, along with operative records and pathology review forms.
- Upon completion of data entry, all forms are routed either first to the disease site data coordinator for a quality control review or directly to the electronic patient chart.

Data Entry

All newly activated trials use forms, which can be submitted electronically via the web. Entry of other Southwest Oncology Group data is accomplished in several ways. Most of our current forms were created using the commercial product Teleform[®], which produces forms that can be scanned directly into the database. Data entry of non-Teleforms is also done through the Teleform[®] software. Updates of last contact date, status, off-treatment date and expectation resolution are done through a separate in-house Visual Basic application. Data is sent to the database and a TIFF image of the form is then sent to the electronic patient chart.

Electronic File Organization

The Statistical Center organizes files so that all data for a patient are located in one electronic file.

Data within patients' charts are maintained with the most recent data at the top or on the left-hand side of Chart Manager as an incoming document. Electronic records are saved in a back-up database each evening, the security of which is detailed under data security and disaster recovery in the computing infrastructure section. The electronic filing system of data is stored by patient-specific numbers, then under study identification.

Quality Control

Quality control occurs on several levels: the scientific review of new protocols, the design of new forms, the review of data by study coordinators, and the review of each submitted data form for consistency and completeness.

Protocol Development and Review

The first step to ensure quality control in clinical trials is to develop protocols that are clearly stated and exhaustively inclusive of all criteria and procedures necessary for conduct of the study. The protocols are the standard against which the Group measures the conformance of the investigators. Most importantly, protocols reflect the scientific direction and standards of the group.

The Statistical Center is intimately involved in the development of protocols, from the capsule stage through to activation. The Committee Statistician works closely with the Study Coordinator(s), the Disease Chair, and the Protocol Coordinators to develop the protocol design. Statisticians collaborate with their medical colleagues to determine appropriate scientific questions, correct endpoints, and feasible accrual goals. During the protocol development, the Data Coordinator for each committee may provide critiques and recommendations to eliminate ambiguities. Each protocol is reviewed for protocol consistency via a Protocol Review Committee in the Statistical Center and a Protocol Consistency Checklist in the Operations Office.

These aspects of each protocol have been designed to enhance the quality of the data. The first is the eligibility section, which assures eligibility of all patients at the time of registration. This section must be confirmed by the physician and the clinical research associate prior to the placement of the registration call to the Statistical Center. The second is a study parameter calendar for each arm of a study, detailing the data to be collected at each patient encounter, and treatments to be delivered at each treatment visit. This is not only essential at the institution level to assure conformance to standards, but is useful in quality control review of the submitted record. Finally, study data collection forms are designed to provide efficient data collection and study analysis.

Initial Quality Control Review

The initial forms set (IFS) consisting of the completed eligibility worksheet, and any other forms required by the protocol are reviewed by the data coordinator to ensure protocol compliance. All data are assessed to confirm that prestudy information pertaining to stratification variables is the same as that given at registration; that the patient's body surface is calculated correctly; that the required protocol tests were performed; and that initial treatment was given as per the protocol. Results of quality control are communicated to the registering institution via an evaluation status report. If a documentation error or discrepancy is noted, the institution may correct the error by submitting the appropriate amended form. Evaluation status reports are mailed to institutions regardless of the quality control outcome.

Data Entry

Data entry of prestudy information and other data on data collection forms is accomplished through scanning and entry via Teleform[®]. Extensive range and logical error checks are incorporated.

Patient Evaluations

A summary of the data for each patient is mailed to the study coordinator for review and evaluation. This summary includes preliminary evaluation information, entered by the Statistical Center data coordinator on eligibility, protocol deviations, toxicity grades, and response. These judgments are preliminary until reviewed and confirmed by the study coordinator. In the case of a disagreement, the disease committee statistician may become involved. Continuing disagreements are adjudicated by the disease committee chair or in rare instances, the Group Chair.

Data Coordinator Quality Assurance (DCQA)

The DCQA program was implemented in August 2000 as an internal audit mechanism to monitor and maintain the quality of data evaluations conducted by data coordinators. Each data coordinator is audited every three months by the quality assurance coordinator, with respect to coding of: eligibility; disease status, treatment; reason off treatment, toxicity; notes; other (second primaries, etc). If necessary, further instruction and education is provided to the data coordinator to ensure that errors are not repeated.

Study Coordinator Evaluation Monitoring System

Study Coordinator Evaluation Forms (SCEFs), are generated by the disease site data coordinator at certain key points after patient registration to a Southwest Oncology Group protocol, e.g., following completion of treatment, relapse, and death. For some protocols, it may be appropriate for the study coordinator to evaluate patient data every three months during treatment.

The Evaluation Monitoring System tracks all SCEFs generated after June 1995. Each time a SCEF is generated, the patient number, study number, registration type, and date of generation are automatically recorded as a row in the SCEVAL database table. This row also includes fields indicating the SCEF returned status and the study coordinator evaluation date (generally the date the study coordinator completes and returns the form). Upon return of each SCEF from the study coordinator, any coding changes made by the study coordinator are entered into the database by the data coordinator.

A Study Coordinator Evaluation Report is generated and sent to the Group Chair, the disease committee chairs, and to individual study coordinators. This report contains information on the number of evaluations generated for a particular study, the number of evaluations returned, and the length of time since outstanding evaluations were generated. These and other reports are used to monitor study coordinator workload and to evaluate compliance with study coordinator responsibilities.

Expectation System

The Statistical Center expectation system is the structure within which quality control functions are implemented. The primary focus of the current version is data submission timeliness. The expectation system notifies institutional clinical research associates what and when specific submissions or tasks are due; when a submission becomes overdue, the notifications become insistent. It also generates timeliness data on each institution, which can then be used to provide feedback on performance.

Expectations are entered into the Study Description File (SDF) and are posted automatically at the conclusion of the registration process. Expectations are highly study dependent and may additionally depend on registration-specific factors such as treatment assigned, applicable stratum, registering institution, type of institution, prior treatment, and/or prior pathology reviews. Expectations to be generated are specified in the SDF corresponding to that study registration using a simple procedural language to implement the logic necessary for deciding whether an expectation of a specific type is to be posted. For example, SWOG 9917 is a placebo-controlled study of L-Selenium based chemoprevention of prostate cancer among men with high grade prostatic intraepithelial neoplasia. The study requires a repeat biopsy only for those patients whose initial biopsy included fewer than 10 cores or was performed more than 6 months prior to registration. Materials from the repeat biopsy, if required, must be submitted for central review within 14 days of the repeat biopsy to confirm absence of cancer prior to randomization. Therefore, pathology submission expectations associated with the repeat biopsy are posted only for a subset of patients.

An expectation is resolved when the Statistical Center has received the submission or has been provided evidence that the task has been performed. In most cases, the resolution of an expectation is an automatic process resulting from data entry of the submission. Remote data entry, for forms submitted on-line, includes automatic resolution of the data expectation item. Lists of expectations are posted monthly for the institutions, on the CRA Workbench.

Institutional Performance Review (IPR)

In conjunction with the monthly expectation report, the Southwest Oncology Group summarizes monthly statistics that are used to assess institutional performance. These standards assess data submission based on four criteria: submission of initial forms sets, follow-up for alive patients, submission of pathology materials, and submission of radiation therapy materials. Data items that are overdue on any of these categories are starred on the institutions' expectation reports, to aid in identification of cases requiring immediate attention. The monthly IPR statistics are reported to the institutions, and monitored at the Statistical Center by the Coordinating Statistician. Institutions that are out of compliance for two months in a row receive a warning letter. Any institution that is out of compliance with the standards of the Group for three months in a row will lose registration privileges until the deficiencies are corrected.

Serious Adverse Event Reporting System (SAE)

Serious adverse events are reported in several ways. If an SAE occurs that is reportable per protocol, investigators are expected to call the Operations Office to report it. AdEERS (Adverse Event Expedited Reporting System) may also be required. If the investigator fails to report an SAE, the study coordinator may ask that the event be reported. Certain toxicities, including fatal toxicities for all treatments and life threatening nonhematologic toxicities for investigational treatments, are subject to an automatic Serious Adverse Events Reporting System.

For studies with flow sheets, data coordinators at the Statistical Center grade and enter all toxicities for patients undergoing evaluation. Toxicities for many studies are also entered via toxicity forms submitted by institutions. If a Grade 5 (fatal) toxicity is entered for any patient or a Grade 4 (life threatening) toxicity is entered for patients receiving investigational treatment, the computerized Serious Adverse Events Reporting System checks the database to determine if the event has been reported. If the event has been previously reported, no further action is taken. If the event has not been previously reported, the patient number, study number and registration type will appear on a report that is routed to the SAE coordinator and to the Operations Office SAE coordinator. This allows the SAE coordinator to contact the institution to ask that the event be reported.

Serious adverse events reported to the Operations Office in any of above ways receive an SAE number and are entered into the tracking system by the SAE coordinator. The Operations Office will notify the Statistical Center when a final determination has been made as to whether an SAE is attributable to protocol treatment.

Discipline Committees

The Statistical Center has focused a great deal of effort toward service to three Southwest Oncology Group active discipline committees: pathology, surgery, and radiation therapy. These committees require two kinds of support from the Statistical Center: (1) processing the results from the data reviews performed by the members of these committees, and (2) providing statistical design and analysis services in support of their scientific activities.

The review functions performed by the discipline committees are designed to answer questions of protocol eligibility (pathology and surgery) or protocol compliance (surgery and radiation therapy). The summary data, which result from these reviews, are an important part of the Group patient data evaluation system and are integrated into the patient evaluation forms along with other eligibility and compliance data. In addition, the review processes yield important detailed data regarding pre-treatment status (pathology), procedures performed (surgery and radiation therapy), or outcome (pathology and surgery). The study coordinators are provided with copies of these detailed data forms.

The specific elements of the Statistical Center data operations that are relevant to the support of these discipline committee functions include the following.

1. A master's level biostatistician is assigned as the statistical liaison to each discipline committee. This biostatistician is responsible for seeing that the review processing requirements of the assigned committee are implemented and functioning and for seeing that their respective committees are provided with the necessary summary data on the operation of the review process. These reports include information on the number of reviews performed, the performance of specific institutions, review summary outcomes for specific protocols, etc.
2. The discipline coordinator supports each discipline committee through a variety of functions, depending on the needs of each discipline committee. These include patient registration, data entry management, forms tracking, quality control, protocol compliance enforcement, protocol setup, and review of protocols with respect to discipline committee data processing.
3. The study description file (SDF) corresponding to each study registration that involves discipline committee participation contains specifications of the various review committee options. Most of these are processing steps to be executed at the time of patient registration to that study and others cause flags to be set in the database. The processing steps might include request for additional information, display of special notes, creation of expectations, validation of special eligibility requirements, etc. The database flags include indications that the patient is registered to a study involving one or more review processes, and the setting of these flags may depend on the treatment assigned.
4. The expectation system is used to notify institutions of the need to submit materials and/or forms required by the discipline review processes, and then to track the submission of the materials or forms; it also yields timeliness data for the institutions. The resolution of expectations is done as the forms or materials arrive by the discipline coordinator at the Statistical Center.

5. The primary returns from a review process are the summary results, that is, an indication of eligibility (pathology and surgery) or protocol compliance (surgery and radiation therapy). The data are entered as soon as the review is completed and returned by the respective reviewer. The summary data are entered by the Statistical Center discipline coordinator. The input procedures are highly structured and include the creation of notification of changes in eligibility status, automatic update of overall eligibility (if applicable), and the ability to make annotations.
6. Summary results of the discipline review, including annotations, are reported to the study coordinators in the evaluation forms sent to them for patient evaluation data review. In addition, copies of the detailed discipline review process forms are sent to the study coordinators as they become available.
7. For each Group Meeting there are reports prepared for the Radiation Therapy and Pathology discipline committees on the results of the review processing for that cycle. These reports include information on institutional timeliness, review-specific eligibility/deviation rates, review process timeliness, review process volume, etc.
8. The Statistical Center also provides support for the entry of whatever other discipline review committee data are generated.

Cytogenetics

The Southwest Oncology Group Cytogenetics Business Office was relocated from City of Hope National Medical Center to the SWOG Statistical Center in June 2002. The Business Office is staffed by one data coordinator and one statistician. Responsibilities of the SWOG Cytogenetics Business Office include coordinating the application and approval process for new cytogenetics laboratories. Another primary responsibility is the data entry of cytogenetics forms. Data entry time consumes about 30 minutes per case due to the volume of forms submitted per case. The data coordinator and biostatistician coordinate Cytogenetics Committee Meetings, prepare patient files for central reviews held at each semi-annual Group meeting, and assist committee members as necessary during the central review process. Database queries are generated and submitted as requested by the Cytogenetics Committee Chair and other members of the committee. A future goal for the Cytogenetics Business Office is to convert case report forms and data entry operations to web-based programs whereby participating laboratories will remotely submit cytogenetics data directly into the database. There are currently 4329 specimens in the cytogenetics database from 33 different SWOG trials. We receive 23 samples a month on average.

Training Programs

Clinical Trials Training Course (CTTC)

The Data Operations Department at the Statistical Center, along with assistance from the Operations Office staff and experienced clinical research associates, conducts a one and a half day training course for new clinical research associates. This training course is held prior to each Southwest Oncology Group Meeting. Ninety-eight clinical research associates from throughout the Southwest Oncology Group participated in the Spring 2004 course.

The training course provides an overview of the Southwest Oncology Group as well as information pertaining to topics such as the explanation of clinical trials, phases of studies, quality control, quality assurance, ethics and quality of life. Familiarization with the Southwest Oncology Group registration process, forms, and office procedures are also included. A practicum round-table discussion provides clinical research associates with practice of forms completion, adverse event grading, response assessments, and calculating laboratory values.

An explanation of data flow, the expectation system, patient follow-up, and adverse events are other topics discussed in the training. Volume I of the Clinical Research Manual, which details the administrative procedures of the Group, and Volume II, which describes the forms and coding guidelines of each disease committee, are available via the internet at swog.org. CTTC participants are oriented to the web-based manual during the training course. Copies are mailed to participants who do not have Internet access.

Study Coordinators Workshop

Members of the Statistical Center staff, in conjunction with Operations Office Staff, conduct a yearly, half-day course for researchers planning to become Study Coordinators. This course is held the day prior to the Spring Group Meeting. Each year, approximately 35 investigators participate in this workshop.

The course provides training in the responsibilities of a Study Coordinator, and of the coordinated efforts of the Statistical Center and Operations Office. There are presentations on ethics, protocol development steps, clinical trials design, data flow, and study coordinator responsibilities during development, during study accrual, and during manuscript preparation. The final hour and a half is a practicum session, where participants are asked to review a mock patient chart, to introduce Group coding guidelines, data collection forms, and study coordinator patient evaluation responsibilities.

Young Investigators Training Course

The Statistical Center and Operations Office give an intensive course to new investigators who show promise of providing future leadership to the group. This program, provided in part by support from the HOPE foundation, was established in 2000. Applications for the workshop must include a proposal for a clinical trial, and recommendations from a Southwest Oncology Group Investigator who will serve as a mentor to the Young Investigator.

The sessions consist of a four-day program at the Operations Office, followed several weeks later by four days at the Statistical Center. Participants attend presentations of protocol development, statistical methods in clinical trials, an overview of data collection, data flow and data evaluation. A key goal of the program is to have the Young Investigator produce a complete protocol document that is appropriate for CTEP submission.

Interacting with Other Cooperative Groups

The Southwest Oncology Group participates in NCI-sponsored intergroup protocols. These are protocols in which more than one cooperative group participates. The protocol is coordinated by only one of these groups. The three types of participation are as follows.

Other Group-Coordinated Protocols Not Conducted Through the Cancer Trials Support Unit

For non-SWOG intergroup treatment trials that are not conducted through the NCI funded Cancer Trials Support Unit, the Southwest Oncology Group primarily participates by contributing towards the accrual of patients onto the protocol. Currently, the Southwest Oncology Group participates in 15 of these intergroup protocols. The Statistical Center's responsibilities for all of these trials include cooperating with the coordinating group's statistical center and operations office by performing data monitoring and follow-up tasks as required.

Additionally, except for trials coordinated by ECOG, CALGB, or NCCTG, the Statistical Center's responsibilities also include forwarding data to the coordinating group.

For trials coordinated by ECOG, CALGB, and NCCTG, a Direct Data Submission Initiative has been piloted and proven successful for expedient data transfer to the coordinating group. Via this direct data submission initiative, data are submitted by the institution directly to the coordinating group, bypassing the participating group's data operations center. A nightly secure ftp database transfer from the coordinating group to the participating group updates the patient's survival and last contact information as it is processed by the coordinating group.

Except for trials coordinated by ECOG, the Statistical Center is responsible for conducting patient registrations for intergroup/non-CTSU trials. For ECOG coordinated studies, a computer-based project linking the SWOG and ECOG databases for patient randomizations has proven successful, thus eliminating the need for SWOG data coordinators to contact the ECOG randomization office for every SWOG patient registered to an ECOG trial.

For registrations on non-CTSU protocols managed by another cooperative group besides ECOG, a data coordinator contacts the other cooperative group statistical center and obtains treatment assignment and patient identification numbers. The data coordinator is then responsible for completing the registration by calling the Southwest Oncology Group clinical research associate and relaying the treatment assignment and patient identifying information as indicated by the protocol.

The Statistical Center facilitates its participation in non-CTSU intergroup protocols managed by other groups by recording in the Southwest Oncology Group database the patient number assigned to Southwest Oncology Group patients by other groups. Follow-up information for non-SWOG studies is included in the expectation report and IPR calculations as a performance factor.

Other Group-Coordinated Trials Conducted Through the Cancer Trials Support Unit

The Statistical Center's responsibilities for non-SWOG trials conducted through the Cancer Trials Support Unit differ in that patient enrollment and data submission is routed from institution to CTSU to coordinating group. In addition, the CTSU is responsible for cooperating with the coordinating group's statistical center and operations offices by routing data and follow-up requests as required. Registration information is relayed to the SWOG Statistical Center for recording (for those CTSU studies endorsed by SWOG).

Southwest Oncology Group-Coordinated Protocols

Currently, the Southwest Oncology Group coordinates 15 open intergroup treatment protocols. These studies are available to non-SWOG participants either via the CTSU menu, or directly with specific cooperative groups. For these protocols, a number of innovations have been implemented.

1. The computerized registration system is programmed to recognize intergroup protocols. When the registering institution is from another group, the program follows special instructions with respect to identification of the registering investigator and institution, dynamically balanced treatment assignment, the expectation list, and specific notes directed to the registering groups.
2. The Southwest Oncology Group Statistical Center has designed its investigator roster and the follow-up information in the database to be able to identify the investigator and institution associated with each patient in requests for follow-up data sent to other statistical centers.
3. The expectation reports and requests for follow-up sent to other group statistical centers are specifically designed to facilitate communication with their institutions, and are therefore different in style and content from the analogous reports sent to institutions in the Southwest Oncology Group.
4. The Southwest Oncology Group has made on-line registration available to institutions of two of its intergroup partners (ECOG and ACoSOG), and is developing similar links to other cooperative groups.
5. The Southwest Oncology Group Statistical Center data coordinators and biostatisticians have available to them special reports designed to facilitate the monitoring of intergroup studies. These reports are for monitoring accrual (by group) and identifying data submission or follow-up problems.
6. The Southwest Oncology Group is developing procedures to allow participation with collaborators from Europe and Asia.

Statistical Designs and Study Monitoring Policy

All Phase III studies managed by the Southwest Oncology Group are designed with formal stopping rules in the protocol and are reviewed by a single Data and Safety Monitoring Committee. This committee consists of three members of the Group, a patient advocate, two clinicians and a statistician not involved with the Group, two non-voting representatives from the NCI, and the Group statistician (also non-voting). Detailed interim results are generated at the Statistical Center and are presented only to this Data and Safety Monitoring Committee and not to the Group as a whole. The Committee decides when to close the study and when to report the results, using the stopping rules in the protocol as a guideline. Stopping rules are based on group sequential designs, which preserve the overall error rates but allow for early stopping if extreme results are observed. The result is a decrease in the number of false positive trials caused by repeated significance testing without appropriate adjustment, and a decrease in the number of studies that are closed informally by poor accrual based on inappropriate judgments from interim results. In addition to the usual specification of Type I and Type II errors, a typical design for a Phase III study would call for specification of the number of analyses (generally 3 or 4) and of a small probability of terminating at each interim analysis if the null hypothesis is true. A one-sided test can be supplemented with a similar early stopping rule based on testing the alternative hypothesis. The result is a procedure with virtually the same power and level as the fixed sample size procedure, but one that allows for early termination and permits a final analysis at close to traditional levels (Crowley, Green, Liu and Wolf, 1994; Fleming, Green, and Harrington, 1984).

Phase II studies involving investigational new agents are generally designed with two-stage stopping rules, and protocols are temporarily closed at the end of the first stage to assess response rates. Reporting of early results for Phase II studies is restricted, subject to the discretion of the study coordinator, statistician, and disease committee chair, to avoid informal closing due to premature judgments based on small numbers. The standard Southwest Oncology Group approach to Phase II designs is to test the null hypothesis that the response probability is p_0 , too low to be of interest, vs. the alternative that is p_A , sufficiently high to warrant further study, at a level approximately .05 and power close to .9. Studies are stopped early if the alternative hypothesis is rejected at the .02 level after the first stage of accrual. Otherwise, accrual is completed and the agent is judged promising if the null hypothesis is rejected. This approach has good statistical characteristics and is easily adaptable to the typical case in which the actual attained sample size differs from the planned sample size at the first stage or at the second stage (Green and Dahlberg, 1992).

There is no formal Data and Safety Monitoring Committee for Phase II trials. Toxicity and accrual monitoring are done routinely by the study coordinator, study statistician, and the disease committee chair. Response monitoring is done by the study statistician and study coordinator. Accrual reports are generated weekly, and toxicity reports are generated frequently. In addition, the Statistical Center, Serious Adverse Event Coordinator at the Operations Office, and the Executive Officer monitor toxicities on an ongoing basis.

Data Analyses and Reporting

Data Reporting

All open and most recently closed studies are described in the semi-annual Report of Studies, which forms the basis for much of the discussion in the Disease Committees at the Group meeting. Data are as current as possible, subject to printing deadlines and the need for data review by the study coordinators. A patient-driven evaluation system is utilized to smooth the workflow, rather than relying on batching the reviews every six months. The description of each study includes a summary face sheet, schema, accrual information by arm and by institution, stratification and description factors and toxicity information. Response and survival comparisons are presented only when approved by the Data and Safety Monitoring Committee.

Standard report modules extract data from the database and create these tables and a computerized worksheet program helps keep track of the Report contents. The Report of Studies includes chapters for each disease committee, and the cancer control research committee. An additional section contains information on accrual to all Group studies by institution and study type; data on institutional performance regarding timeliness and completeness of forms is also included.

Analysis of Studies

Additional analyses, generally using SAS® after data have been extracted from the database, are typically necessary before a study can be published. A program to extract data from the Oracle® database and create a SAS® data set was written to assist in this process. SAS® version 6.07 procedure LIFETEST is used for log rank and stratified log rank tests, and the procedure PHREG is used for proportional hazards regression and testing alternative hypotheses. Dichotomous data are analyzed using the procedures FREQ and LOGISTIC. Locally developed analysis software is used when necessary, especially for Cox regression diagnostics and for recursive partitioning. In addition, faculty statisticians are involved in ongoing methodologic research for more efficient trials designs, and improved analytic techniques.

Inclusion of Women and Minorities as Research Subjects

The Southwest Oncology Group is committed to the equitable inclusion of women and minorities in all trials. We have a Committee on Women and Special Populations, one of whose missions is the monitoring of accrual for sex, race/ethnicity, and age, and suggesting interventions if needed. Our work on accrual of women was presented at ASCO (Hutchins et al., 1994); overall, the Group accrues more women to trials than would be suggested by incidence due to our emphasis on breast cancer. A study of accrual to older populations (Hutchins et al., 1999) indicated that there is an under representation of patients 65 years of age or older in cancer treatment trials. In addition, we have investigated sex, race/ethnicity, and age as prognostic factors in several disease sites (Modiano et al., 1996; Dahlberg et al. 1994; Flaherty et al. 1996; Albain et al., 1990, 1991, 1991). All Phase III trials now contain an analysis of treatment effects by race/ethnicity and sex, and have accrual goals in subsets where warranted.

Prostate Cancer Prevention Trial (PCPT)

Overview

The Prostate Cancer Prevention Trial (PCPT) was a Phase III, randomized, double-blind, placebo-controlled trial of finasteride for the prevention of carcinoma of the prostate. A total of 18,882 essentially healthy men, aged 55 and older, were randomized to receive finasteride (5 mg daily) or placebo for seven years. Active participants were monitored quarterly for side effects and adherence to treatment schedule. All participants received annual follow-up for signs or symptoms of prostate cancer; an elevated PSA value and/or an abnormal DRE at this time prompted a recommendation for a prostate biopsy. At the end of seven years of follow-up, all participants were asked to undergo a prostate biopsy.

The PCPT was an intergroup study with participation from 213 study sites in the United States and Canada. The trial was activated on October 13, 1993 and closed to enrollment on December 6, 1996; the last participant was randomized on May 16, 1997. The PCPT was closed on June 24, 2003 and the results of the study were published in the online version of the New England Journal of Medicine and appeared in the print journal on July 17, 2003.

Objective

The primary objective of the PCPT was to test the difference in the histologically proven prevalence of carcinoma of the prostate between these two groups of randomized participants. Other objectives included assessment of the effect of finasteride on the stage and grade of carcinoma at the time of diagnosis, and estimations of the difference between the two groups in (1) total and prostate cancer-specific mortality and (2) the incidence and severity of benign prostatic hyperplasia (BPH). The toxicity and side effects of long-term administration of finasteride were assessed, as well as the effect of finasteride treatment on various quality of life dimensions. The standard screening parameters for prostate cancer (DRE, PSA) were evaluated for their sensitivity, specificity and predictive value and the effect of long-term finasteride treatment on these parameters was also estimated.

Results

The analysis of the data revealed that men in the finasteride group who were evaluated were 24.8% less likely to develop prostate cancer when compared to the men evaluated who were in the placebo group. Although the men taking finasteride had fewer prostate cancers, they had an increased number of high-grade prostate cancers. In the entire group of men assigned to finasteride who were evaluated, 6.4% had high-grade cancers while 5.1% of the men evaluated in the placebo had high-grade cancers.

In order to better determine whether finasteride causes more aggressive tumors to develop or whether finasteride affects the appearance of prostate cancers under the microscope leading to a faulty estimate of tumor grade, new studies are being pursued. A program project (PO1) proposal has been submitted using the specimens collected during the course of the trial (serum, white blood cells, prostate biopsy tissue). This proposal is multi-institutional and will investigate various areas of research including:

- 5 α -Reductase, CYP3A4, and CYP3A5 gene variants (*Reichardt, University of Southern California*)
- Androgen receptor and HSD3B2 gene variants and serum hormones (*Figg, National Cancer Institute*)
- Insulin-like growth factor axis and insulin resistance (*Pollack, McGill*)
- Diet and diet-related factors (*Kristal, Fred Hutchinson Cancer Research Center*)
- Oxidative damage, estrogens, and DNA repair (*Santella, Columbia*)
- Genotypic and phenotypic studies of inflammation and atrophy (PIA) (*Platz, Johns Hopkins*)

A new long-term follow-up study has been proposed by the leadership of the PCPT. The primary objective of this study will be to determine the metastatic prostate cancer risk for men who were diagnosed with prostate cancer on the PCPT. Prostate cancer treatment data will be collected and men will be followed for a diagnosis of metastatic disease and death. This study is in development with plans to be activated in late 2004.

Role of the Statistical Center

The primary role of the Statistical Center in the PCPT was to ensure the successful implementation of the trial design through effective study design, data management, data analysis and overall trial management.

Managing a trial of this size involved substantial coordination and close communication with the staff at the Study Centers and Sites. The data operations staff at the Statistical Center was the primary contact with the staff at the Study Sites. In addition to their responsibilities for data management, which included the data entry, validations and review of the large volume of data received for PCPT, they presented a PCPT Workshop twice yearly in conjunction with the Southwest Oncology Group meetings each spring and fall.

The Statistical Center maintained and updated the PCPT Study Manual to ensure that the Study Centers and Sites were informed of all changes to procedures and forms. The updates were communicated as Study Bulletins, memos, Tip Sheets, and general manual and forms revisions. The most recent update to the PCPT Study manual provided Study Centers and Sites with instructions related to study closure procedures. Although the primary study has closed, the Statistical Center staff continues to contact Study Center and Site staff to resolve problems with the data.

The biostatisticians provided analyses for the semi-annual Data Safety Monitoring Committee meetings, and the primary publication with the results of the trial. They are continuing to analyze the trial data and working on several publications of interest in the areas of trial management, recruitment and adherence, quality of life, pathology, diagnostic test (PSA) as well as other areas of urologic and clinical interest.

The programming staff worked to support the Statistical Center's efforts by providing programming tools and reports that assisted with the timely review of the data, reporting of discrepancies and missing data in QC reports, and resolution of the problems identified. Throughout the course of the PCPT, the data acquisition system used was DataFax™, an integrated image management and database system which used Optical Character Recognition (OCR) technology. The trial generated approximately 4.5 million data form images, which required approximately 150 gigabytes of computer storage space. PCPT used a RAID (redundant array of inexpensive disks) for storage of fax images.

With the recent closure of the primary PCPT study, the Statistical Center is in the process of transferring the study data collected via the DataFax™ system into Oracle database tables and converting the accompanying DataFax™ images into TIF files for long-term access and storage. The new data management system integrates image management with an Oracle database and utilizes HTML forms and *.net* technology.

The Selenium and Vitamin E Cancer Prevention Trial (SELECT)

Overview

The Selenium and Vitamin E Cancer Prevention Trial (SELECT) (S0000) is a Phase III, randomized, double blind, placebo-controlled trial to prevent prostate cancer. SELECT was designed to accrue 32,400 healthy men, ages 55 years and older (African American men ages 50 years and older), with 20% (6,480) of the study participants to be African American. The efficacy of selenium and vitamin E, as single agents and in combination, on reduction of prostate cancer incidence, will be tested in a statistically highly powered 2 x 2 factorial trial design. The planned total study period is 14 years, including 1 year pre-study for ramp-up to accrual (completed), 5 years for accrual, 7 to 12 years of treatment, and 1 year post-study for analyses and publication of results. Participants will be followed twice per year (four times in the first year after randomization) to monitor general health, prostate health, and adherence to the study supplements.

In addition to primary and prespecified secondary endpoints, the SELECT will examine many important tertiary/ancillary endpoints, including dietary/nutrient assessments, pathology and molecular/cellular biomarkers, quality of life, and molecular epidemiology. Development and monitoring of the trial involves the input and scientific expertise of a panel of experts representing several cooperative groups.

The study opened for recruitment on July 25, 2001. Randomizations began on August 22, 2001. The trial was closed to accrual on June 24, 2004 at which point 35,534 participants were randomized to the trial, 15% of whom are African American, at 427 different study sites across the United States, Puerto Rico and Canada.

Role of the Statistical Center

The Statistical Center is responsible for data management for the trial, described in detail below:

1) Study Design and Analysis

The Statistical Center develops methods for analysis of study data, evaluates and modifies the study design as needed, analyzes study data, and provides reports to the Data and Safety Monitoring Committee and to the Steering Committee and its subcommittees.

2) SELECT Workbench

The SELECT Workbench is a secure Web site located on the World Wide Web, administered by the SELECT Statistical Center. The Workbench contains the SELECT protocol, Study Manual, and a variety of materials to assist sites in performing all activities associated with randomization, participant follow-up, and study administration. Regular review of the effectiveness of the Workbench and updates to its content are ongoing.

3) SELECT Study Manual

Version 1.0 of the SELECT Study Manual was released in November 2000. Since that time, numerous revisions have been released. The Study Manual contains procedures and guidelines to augment the protocol. All sections of the Study Manual have been released and undergo regular review and revision as necessary.

4) Study Site Staff Training

Study Site staff members receive training on SELECT procedures at the semi-annual SWOG Group Meetings. This training consists of presentations, small group breakout sessions on specialized topics of interest, and poster sessions. A special session is held for staff members who are new to the trial.

5) Communications

The Statistical Center is available Monday through Friday 7:00 AM PST through 4:00 PM PST for telephone queries. In addition, Study Site staff members are contacted as necessary by their assigned Data Coordinator to resolve recurrent problems in data management or study administration. A Helpdesk function is also available, by both voice and e-mail, for Study Sites to call for assistance with a wide range of study management issues.

6) Document Management and Quality Assurance

Participants were randomized by Study Site staff using Web-based (html) data collection and transmission. Data collection is done using Cardiff™ Teleform®, an electronic document management system featuring both Web-based and fax transmissions. Data collection is also done using Web-based (html) transmission. The Statistical Center provides a dedicated toll-free number for faxed data transmission from the Study Sites. Data are validated by data operations staff and stored in a relational database; electronic documents and digitized images of forms are archived in a disk storage system.

a. Electronic Document Management Update

The electronic document management system is operated by Data Control Technicians and Data Coordinators and is maintained by programming staff. Data Control Technicians initially review each page of faxed data for completeness prior to entry to the database. Data Coordinators also assist with this review at a secondary level. Subsequent data review is also completed by the Data Coordinators. This includes evaluating routine data forms and data from narrative summaries for accuracy, completeness, consistency, and compliance to data collection requirements and follow-up procedures. The results of these reviews are noted in an evaluation program.

b. Data Collection Forms

The forms developed for this protocol have been designed to optimize their processing and management using an electronic document management system that features both Web-based and fax transmissions of data from the Study Sites. In general, forms that are completed by the participant are faxed to the Statistical Center; CRA completed forms are submitted via the SELECT Workbench web data entry process.

With the exception of the 17-page Dietary Supplement and Food Questionnaire, which is available in three languages, the Study Sites download data collection forms from the Sta-

tistical Center via the SELECT Workbench. A set of reference forms, with instructions for the completion of each form, is included as part of the Study Manual.

c. Routine Reports

Study Sites have the ability to access a variety of reports via the SELECT Workbench. These reports are developed and maintained by the Statistical Center and are available to assist with identifying time requirements of individual participant follow-up activities, provide updated accrual and data collection and submission information for each Study Site.

d. Quality Control

The Southwest Oncology Group Operations Office conducts Study Site Quality Assurance Audits. The Statistical Center provides data reports for the audit team.

Quality control review begins with the initial review of each incoming electronic document for data entry errors (e.g., missing data items, illegible entries) including those resulting from inaccuracies in image-to-data translation. Translation errors are corrected at this time. Other errors are annotated for correction or clarification; these notes are included in query reports accessible to the Study Sites dynamically on the SELECT Workbench.

Quality control review is completed by the Data Coordinators. This includes evaluating routine data forms for accuracy (the correct form was used to document a specific protocol requirement); completeness (all required data items have been addressed); consistency (data recorded on one form is supported by and/or matches that recorded on another form); and with protocol requirements (all required study parameters have been addressed and documented within the specified time frames).

A system of edit checks implemented at the time of Web-based data transmission from the Study Sites and during review at the Statistical Center contributes to the overall quality control review. All deficiencies, discrepancies, and other quality-related problems identified by the Statistical Center staff are included in the query reports. Study Sites are directed to correct or clarify data items in response to the Statistical Center's requests, and resubmit the amended version of the original document. The results of these reviews are noted in an evaluation program.

All electronic documents received are logged into generalized tracking system (Expectation System) designed to identify those documents not received and participant follow-up not completed per protocol requirements. Reports listing these missing documents and instances of incomplete participant follow-up are available to the Study Sites via the SELECT Workbench

Following study completion each Study Site will be expected to archive the materials collected during the study for storage for a minimum of 5 years.

e. Institutional Performance Review: Monitoring Study Center Performance

Data submission by the Study Centers is monitored monthly, using a number of specific Institutional Performance Review (IPR) measures. These measures include the timeliness of follow-up and the completeness of annual visit data. IPR reports for all SELECT Study Centers are posted to the Workbench.

Any Study Center with a rating of 10% or greater in any measure will be strongly encouraged to take actions to move into compliance. To help an institution improve its IPR rating,

the Statistical Center will schedule a mentoring visit, by members of the Statistical Center staff and/or a CRA, to the Study Center. The mentoring visit is designed to provide a spectrum of techniques, tools and procedures that can be implemented to develop a system that will help a SELECT institution meet the follow-up and documentation requirements of the SELECT.

Statistical Applications and Research

Most of the faculty at the Statistical Center are supported for a fraction of their time on an R01, Statistical Methods for Clinical Studies. Dr. LeBlanc is the PI. The emphases currently are design and analysis strategies for Phase III trials and survival analysis (particularly graphical and other exploratory methods). Methodological publications are given below.

Methodological Publications

1998

1. Moinpour CM, Lovato LC. Ensuring the quality of quality of life data: the Southwest Oncology Group experience. *Stat Med* 17:641-651.
2. Troxel AB. A comparative analysis of quality of life data from a Southwest Oncology Group randomized trial of advanced colorectal cancer. *Stat Med* 17:767-779.
3. Troxel AB, Fairclough DL, Curran D and Hahn EA. Statistical analysis of quality of life data in cancer clinical trials. *Stat Med* 17:653-666.
4. Zhang H, Crowley J, Sox H, Olshen RA. Tree structured statistical methods. In: Armitage, P; Colton, T, ed. *Encyclopedia of Biostatistics*, New York: Wiley: p. 4561-4573.
5. Crowley, J: Interpreting clinical trials 2. Factorial designs, subset analyses, and intent to treat. Thirty-Fourth Annual Meeting, Los Angeles Convention Center. American Society of Clinical Oncology Educational Book, 253-259.

1999

1. Gooley TA, Leisenring W, Crowley J, et al. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Stat Med* 18:695-706.
2. LeBlanc M, Crowley J. Using the bootstrap for estimation in group sequential design: an application to a clinical trial for nasopharyngeal cancer. *Stat Med* 18:2635-2644.
3. LeBlanc M, Crowley J. Adaptive regression splines in the Cox model. *Biometrics* 55(1):204-213.
4. Tangen CM, Koch GG. Complementary nonparametric analysis of covariance for logistic regression in a randomized clinical trial. *J Biopharm Stat* 9(1):45-66.
5. Tangen CM, Koch GG. Nonparametric analysis of covariance for hypothesis testing with logrank and Wilcoxon scores and survival rate estimation in a randomized clinical trial. *J Biopharm Stat* 9(2):307-338.

6. Koch GG, Tangen CM. Nonparametric analysis of covariance and its role in non-inferiority clinical trials. *Drug Information Journal* 33:1145-1159.
7. Liu PY, Tsai WY. A modified logrank test for censored survival data under order restrictions. *Statistics and Probability Letters* 41:57-63.
8. Liu PY, LeBlanc M, Desai M. False positive rates of randomized phase II designs. *Control Clin Trials* 20:343-352.
9. Laird NM, Pauler DK. Discussion of "Adjusting for nonignorable drop-out using semiparametric nonresponse models" by DO Scharfstein, A Rotnitzky and JM Robins. *Journal of the American Statistical Association* 94:1133-1135.
10. Pauler DK, Wakefield JC, Kass RE. Bayes factors for variance component models. *Journal of the American Statistical Association* 94:1242-1253.
11. Pauler DK. Discussion of "Quantifying surprise in the data and model verification" by MJ Bayarri and JO Berger. In *Bayesian Statistics 6*, eds. Bernardo JM, Berger JO, Dawid AP and Smith AFM. Oxford: Clarendon Press pp.70-72.
12. LeBlanc, M. Resampling quantile estimating functions and density estimation. *Computational Statistics and Analysis* 29:163-176.

2000

1. Tangen CM, Koch GG. Non-parametric covariance methods for incidence density analyses of time-to-event data from a randomized clinical trial and their complementary roles to proportional hazards regression. *Stat Med* 19:1039-1059.
2. Mick R, Crowley JJ, Carroll RG. Phase II clinical trial design for noncytotoxic anti-cancer agents for which time to disease progression is the primary endpoint. *Control Clin Trials* 21(4):343-359.
3. Liu PY, Anderson G, Crowley J. Observational studies and randomized trials: Letter to the editor. *N Engl J Med* 343(16):1195.
4. Pauler DK, Laird NM. A mixture model for longitudinal data with application to assessment of noncompliance. *Biometrics* 56:464-472.
5. Pauler DK, Wakefield JC. Modeling and Implementation Issues in Bayesian Meta-Analysis. In: Stangl, D; Berry D, ed. *Bayesian Meta-Analysis*. New York: Marcel-Dekker; p.205-230.

2001

1. Liu, PY. 'Phase II selection designs". In: Crowley J, ed. *Handbook of Statistics in Clinical Oncology*. New York: Marcel-Dekker; p. 119-127.
2. Skates SJ, Pauler DK and Jacobs IJ. Screening based on the risk of cancer calculation from Bayesian hierarchical change-point models of longitudinal markers. *Journal of the American Statistical Association* 96:429-439.

3. Tangen CM, Koch GG. Nonparametric analysis of covariance for confirmatory randomized clinical trials to evaluate dose-response relationships. *Stat Med* 20:2585-2607.
4. Green SJ. Factorial Designs for Time to Event Endpoints. In: Crowley J, ed. *Handbook of Statistics in Clinical Oncology*. New York: Marcel-Dekker
5. Gooley TA, Leisenring W, Crowley J, Storer BE. Why kaplan-meier fails and cumulative incidence succeeds when estimating failure probabilities in the presence of competing risks. In: Crowley J, ed. *Handbook of Statistics in Clinical Oncology*. New York: Marcel-Dekker; p. 513-523.
6. Crowley J, Jacobson J, Alexanian R. Standard-dose therapy for multiple myeloma: the Southwest Oncology Group experience. *Semin Hematol* 38(3):203-208.
7. Hoque A, Albanes D, Lippman SM, Spitz D, Taylor PR, Klein E, Thompson IM, Goodman PJ, Stanford JL, Crowley JJ, Coltman CA, Santella RM. Molecular epidemiologic studies within the selenium and vitamin E cancer prevention trial (SELECT). *Cancer Causes Control* 12:627-633.
8. Faraggi D, LeBlanc M, Crowley J. Understanding neural networks using regression trees: an application to multiple myeloma survival data. *Stat in Med* 20:2965-2976.
9. Crowley J, ed. *Handbook of Statistics in Clinical oncology*. New York: Marcel-Dekker.
10. Kooperberg C, Ruczinski I, LeBlanc M, Hsu L. Sequence analysis using logic regression. *Genet Epidemiol* 21:626-631.
11. LeBlanc M: Tree-based methods for prognostic stratification. In: Crowley, J, ed. *Handbook of Statistics in Clinical Oncology*. New York: Marcel-Dekker; p.457-472.

2002

1. LeBlanc M, Jacobson J, Crowley J. Partitioning and peeling for constructing patient outcome groups. *Stat Methods Med Res* 11:1-28.
2. LeBlanc M: Regression Trees. In: El-Shaarawi, A; Piegorsch, W, ed. *Encyclopedia of Environmetrics*. Wiley; p. 1734-1738.
3. LeBlanc M: Tree structured methods. In: El-Shaarawi, A; Piegorsch, W, ed. *Encyclopedia of Environmetrics*. Wiley; p. 2242-2243.
4. Pauler DK, Laird, NM. Nonlinear hierarchical methods for monitoring compliance. *Stat Med* 21(2):219-220.
5. Green S, Benedetti J, Crowley J. *Clinical Trials in Oncology*, second edition. Chapman and Hall/CRC, London.
6. Green S, Liu PY, O'Sullivan J. Factorial design considerations. *J Clin Oncol* 20(16):3424-3430, 2002.

7. Pauler DK, Finkelstein DM. Predicting time to prostate cancer recurrence based on joint models for non-linear longitudinal biomarkers and event time outcomes. *Stat Med* 21:3897-3911, 2002.
8. Pauler DK, Gower KB, Goodman PJ, Crowley JJ, Thompson IM. Biomarker-based methods for determining noncompliance in a prevention trial. *Control Clin Trials* 23:675-685, 2002.
9. Gatsonis C, Carriquiry A, Gelman A, Higdon D, Kass RE, Pauler DK, Verdinelli I, (eds). *Case Studies in Bayesian Statistics Vol. VI. Lecture Notes in Statistics 167*. New York:Springer, 2002.
10. Abdoell M, LeBlanc M, Stephens D, Harrison RV. Binary partitioning for continuous longitudinal data: categorizing a prognostic variable. *Stat in Med* 21:3395-3409, 2002.
11. Ruczinski I, Kooperberg C, LeBlanc M. Logic regression methods and software. In proceedings of the MRSI workshop, Workshop on nonlinear estimation and classification, Springer, 333-344, 2002.
12. Kooperberg C, Sipione S, LeBlanc M, Strand A, Cattaneo E, Olson JM. Evaluating test statistics to select interesting genes in microarray experiments. *Hum Mol Genet* 11(19):1-10, 2002.
13. Donaldson GW, Moinpour CM. Individual differences in quality-of-life treatment response. *Med Care* 40(6 Suppl):III-39—III-53, 2002.

2003

1. Pauler DK, McCoy S, Moinpour C. Pattern mixture models for longitudinal quality of life studies in advanced stage disease. *Stat Med* 22:795-809, 2003.
2. Unger JM, LeBlanc M, Crowley JJ, et al. Estimating the impact of new clinical trial proven cancer therapy and cancer chemoprevention on population mortality: The Karnofsky Memorial Lecture. *J Clin Oncol* 21(23s):246-252, 2003.

2004

1. Tangen CM, Goodman PJ, Crowley JJ, Thompson IM. Statistical design issues and other practical considerations for conducting phase III prostate cancer prevention trials. *J Urol* 171:S64-S67, 2004.
2. Liu PY, Sondak VK. In: Machin D, Day S and Green S, *Textbook of Clinical Trials*. Wiley; 2004. Melanoma chapter; p. 149-158.
3. Green S, Pauler DK. Statistics in Clinical Trials, *Current Oncology Reports* 6(1):36—41, 2004.
4. Green S. Do estimates of long-term survival tell us whether patients diagnosed with breast cancer before age 50 years are ever cured? Editorial. *J Clin Oncol* 22(3):392-394, 2004.

5. LeBlanc M, Moon J, and Crowley J. Directed peeling and covering of patient rules, recent advances and trends in nonparametric statistics, edited by Michael J. Akritas and Dimitris N. Politis, 393-407, 2004.
6. Ruczinski I, Kooperberg C, LeBlanc ML. Exploring interactions in high-dimensional genomic data: an overview of logic regression, with applications. *Journal of Multivariate Analysis* 90:178-195, 2004.
7. Pauler DK, Hardin J, Faulkner J, Leblanc M, and Crowley JJ. Survival Analysis with Gene Expression Arrays. In *Handbook of Statistics Vol. 23*, eds. CR Rao and N Balakrishnan, Elsevier North Holland: Amsterdam, pp. 675—688, 2004.
8. Hardin J, Waddell M, Page CD, Zhan F, Barlogie B, Shaughnessy J Jr, Crowley J. Evaluation of multiple models to distinguish closely related forms of disease using DNA microarray data: an application to multiple myeloma. *Stat Appl Genet Mol Biol* 3(1):1-21, 2004.

Statisticians usually give presentations at the Group Meeting on some aspect of the design, conduct, or analysis of cancer clinical trials. The titles of these presentations since 1998 are:

- Brent Blumenstein: Some Statistical Thoughts about Tumor Markers and Related Things. Spring 1998.
- Joseph Unger: Southwest Oncology Group Accrual by Sex, Race and Age, Compared to U.S. Population Rates. Spring 1998.
- John Crowley: Informatics: A Status Report. Fall 1998.
- Stephanie Green: Factorial Designs. Spring 1999.
- PY Liu: Cisplatin, 5-Fluorouracil Plus Radiation Therapy are Superior to Radiation Therapy as Adjunct Therapy in High-Risk, Early-Stage Carcinoma of the Cervix after Radical Hysterectomy and Pelvic Lymphadenectomy: Report of a Phase III Intergroup Study. Spring 1999.
- Phyllis Goodman: Design of the Selenium and Vitamin E Cancer Prevention Trial (SELECT—S0000). Spring 2000.
- PY Liu: False Positive Rates of Randomized Phase II Designs. Fall 2000.
- John Crowley: Data and Safety Monitoring Committees: Theory and Practice. Spring 2001.
- Joe Unger (and Charles Coltman Jr., M.D.): Is there a future for federally funded clinical trials? Fall 2001.
- John Crowley: Analysis of Microarray Data. Fall 2002.

- Joseph Unger: African Americans Have Worse Survival in Hormone-Related Cancers. Spring 2003.
- Phyllis Goodman: Design of the Prostate Cancer Prevention Trial (PCPT). Fall 2003.
- Joseph Unger: Accrual Patterns of the Southwest Oncology Group by Sex, Race/Ethnicity and Age: Updated and Expanded Analysis. Spring 2004.
- Cathryn Rankin: Dose Effect of Imatinib in Patients with Metastatic GIST-Phase III Sarcoma Group Study S0033. Spring 2004.

Literature Cited

1. Albain KS, Crowley JJ, LeBlanc M, Livingston RB: Determinants of improved outcome in small cell lung cancer: An analysis of the 2,580-patient Southwest Oncology Group database. *J Clin Oncol* 8:1563-1574, 1990.
2. Albain KS, Crowley JJ, Livingston RB: Long-term survival and toxicity in small cell lung cancer: Expanded Southwest Oncology Group experience. *Chest* 99:1425-1432, 1991.
3. Albain KS, Crowley JJ, LeBlanc M, Livingston RB: Survival determinants in extensive stage non-small cell lung cancer: The Southwest Oncology Group experience. *J Clin Oncol* 9:1618-1626, 1991.
4. Crowley J, Green S, Liu PY, Wolf M: Data monitoring committees and early stopping guidelines: The Southwest Oncology Group experience. *Stat Med* 13:1391-1399, 1994.
5. Dahlberg S, Fisher RI, Coltman CA, Jr.: African Americans treated on SWOG NHL studies: similar incidence, received dose intensity, and outcome to Caucasian patients. *ASCO Proceedings* 13:370(#1251), 1994.
6. Flaherty L, Unger J, Liu PY, Albain KS, Sondak V: Gender differences and independent predictors of survival in patients with metastatic malignant melanoma. *ASCO Proceedings* 15:433, 1996.
7. Fleming T, Green S, Harrington D: Considerations for monitoring and evaluating treatment effects in clinical trials. *Control Clin Trials* 5:55-66, 1984.
8. Green S, Dahlberg S: Planned vs attained design for Phase II clinical trials. *Stat Med* 11:853-862, 1992.
9. Hutchins L, Rector D, Stotts C, Dahlberg S, Coltman CA Jr., Albain K: Analysis of Southwest Oncology Group (SWOG) accrual for sex ratios and associated socioeconomic status (SES) factors. *ASCO Proceedings* 13:188 (#539), 1994. Abstract
10. Modiano M, Villar-Werstler P, Crowley J, Salmon S: Evaluation of race as a prognostic factor in multiple myeloma: An ancillary of Southwest Oncology Group study 8229. *J Clin Oncol* 14:974-977, 1996.
11. Hutchins LF, Unger JM, Crowley JJ, Coltman CA Jr, Albain KS: Underrepresentation of patients 65 years of age or older in cancer treatment trials. *N Engl J Med* 341(27):2061-2067, 1999.