Technology Assessment Program

Report No. 10

Positron Emission Tomography

Descriptive Analysis of Experience with PET in VA

A Systematic Review Update of FDG-PET as a Diagnostic Test in Cancer and Alzheimer's Disease

Authors: Elizabeth Adams, R.R.T., M.P.H., *Management & Program Analyst*
Karen Flynn, D.D.S., M.S., *Manager, MDRC Technology Assessment Program*

Contributors: Elaine Alligood, M.L.S., *MDRC Information Center Librarian*
Jennifer Cheslog, *MDRC Data Assistant*

Report Date: December 1998

HSR&D
Health Services Research & Development Service
MDRC
MANAGEMENT DECISION AND RESEARCH CENTER

MTA 98-032

The Health Services Research and Development Service (HSR&D) is a program within the Veterans Health Administration's Office of Research and Development. HSR&D provides expertise in health services research, a field that examines the effects of organization, financing and management on a wide range of problems in health care delivery, quality of care, access, cost and patient outcomes. Its programs span the continuum of health care research and delivery, from basic research to the dissemination of research results, and ultimately to the application of these findings to clinical, managerial and policy decisions.

Technology Assessment Program
Management Decision and Research Center (152M)
Health Services Research and Development Service
Office of Research and Development
VA Medical Center
150 South Huntington Avenue
Boston, MA 02130

Tel: (617) 278-4469 FTS: (700) 839-4469 Fax: (617) 278-4438
Elizabeth.Adams@med.va.gov
g.mdrc-ta@forum.va.gov

Released July 1999

ACKNOWLEDGEMENTS

The contributions of the following reviewers are gratefully acknowledged. The MDRC takes full responsibility for the views expressed herein. Participation as a reviewer does not imply endorsement.

Valerie A. Lawrence, M.D.	Medical Consultant, MDRC TA Program Associate Professor, Dept of Medicine University of Texas Health Science Center Audie L. Murphy VA Medical Center San Antonio, Texas
Dr. Matthias Perleth, MPH	Dept. of Epidemiology, Social Medicine and Health System Research Hannover Medical School Hannover Germany
David Hailey, M.D.	Director, Health Technology Assessment Unit Alberta Heritage Foundation for Medical Research Edmonton, Alberta Canada
Paula Corabian	Research Associate, Health Technology Assessment Unit Alberta Heritage Foundation for Medical Research Edmonton, Alberta Canada
Mitchell Burken, M.D.	Medical Officer Coverage and Analysis Group Office of Clinical Standards and Quality Health Care Financing Administration Baltimore, Maryland
Ron Milhorn	Senior Health Insurance Specialist Coverage and Analysis Group Office of Clinical Standards and Quality Health Care Financing Administration Baltimore, Maryland
Captain R.K. Leedham, Jr., R.Ph., M.S., BCNP	Chief Project Management Staff Division of Medical Imaging and Radiopharmaceuticals Center for Drug Evaluation and Research Food and Drug Administration
David Pfister, M.D	Associate Attending Physician Memorial Sloan-Kettering Hospital Associate Professor of Medicine Cornell University Medical College New York, New York
M. Elizabeth Glenn, M.D.	Radiologist Women's Health Center

	Memphis, Tennessee
Lisa Hammond, M.D.	Clinical Investigator, Institute for Drug Development Cancer Therapy and Research Center San Antonio Cancer Institute Staff Oncologist Audie L. Murphy VA Medical Center San Antonio, Texas
James K. Stoller, M.D.	Vice Chairman, Division of Medicine Head, Section of Respiratory Therapy Dept. of Pulmonary and Critical Care Medicine The Cleveland Clinic Foundation Cleveland, Ohio
Dawn Provenzale, M.D.	Director, GI Outcomes Research Duke University Medical Center Senior Research Associate Center for Health Services Research in Primary Care VA Medical Center Durham, North Carolina
Ladislav Volicer, M.D., Ph.D.	Professor of Pharmacology and Psychiatry Boston University School of Medicine Clinical Director/GRECC VA Medical Center Bedford, Massachusetts

The MDRC Technology Assessment Program wishes to thank Julie Lowery, Ph.D., Bonnie Bootsmiller, Ph.D., and Andrew Behler, M.P.H. of the VA Center for Practice Management and Outcomes Research in Ann Arbor for contributing VA registry data to the report.

The MDRC Technology Assessment Program also wishes to thank Maria Fonseca, Diane Hanks, Matthew Eberle, Kevin Rys, and the staff of the Information Dissemination Program for their help with the report.

Positron Emission Tomography

PREFACE

The Veterans Health Administration (VHA) has 10 positron emission tomography (PET) imaging facilities and shares ownership and operations with some of its academic affiliates and one with the Department of Defense. Significant resource commitments are associated with the acquisition and operation of these facilities.

In 1996, the MDRC Technology Assessment Program produced a technology assessment report in response to a request from the Office of the Under Secretary for Health for information on VHA's experience with PET. The Advisory Committee for the project provided guidance on the scope and content of the report. The assessment reported the results of: 1) systematic reviews of clinical applications of PET using 2-[F-18]-2-deoxy-D-glucose (FDG) in selected cancers (head and neck, lung cancer staging, solitary pulmonary nodules, breast, and colorectal) and Alzheimer's disease, representing conditions of importance to the veteran population, and 2) surveys of and site visits to VHA PET Centers on PET utilization, center operations, and research activities.

The MDRC found that research into the clinical utility of PET for the selected oncology conditions was in its preliminary stages. Methodological weaknesses in the published literature seriously limited the validity of the available evidence on the accuracy of PET as a diagnostic test, and PET's contribution to improving outcomes had not been systematically assessed. The lack of epidemiological information in these studies made extrapolation of study results to defined VHA populations, and subsequent planning for these populations, difficult.

PET is an accurate diagnostic test for dementia of the Alzheimer's type. Studies to determine whether this accuracy extends to confirmed Alzheimer's disease are under way in Europe. Nonetheless, lack of valid estimates of the positive predictive value of PET, parallel developments in other tests, and limited treatment options for Alzheimer's disease argue for continued use of PET primarily as a research tool. ***Accordingly, the evidence as of September 1996 did not support widespread incorporation of PET studies into routine diagnostic strategies for the applications included in the assessment.***

The site visits and surveys confirmed that VHA has made a substantial resource commitment to its PET facilities and that VHA researchers regard PET as an important research tool. Site investigators identified a wide range of research and clinical activities in VHA PET centers, but noted that these activities remained largely uncoordinated. ***The MDRC concluded that VHA should maximize the value of its existing commitment, rather than establish additional PET centers.*** *This could include:*

- *coordinating activities of VHA PET facilities and their academic affiliates to comply with FDA regulations, to identify research areas of interest to VHA, and to design multi-center studies of high methodologic quality;*
- *implementing a VHA PET registry for systematic data collection and for tracking the utility of PET in selected conditions;*
- *supporting rigorous, prospectively designed clinical research that expands the body of PET literature in a methodologically sound manner; and*
- *submitting currently unpublished data from studies of high methodological quality for peer review.*

Positron Emission Tomography
1998 Update

EXECUTIVE SUMMARY

Purpose
After the delivery of the original assessment report, the Under Secretary for Health directed the Office of Patient Care Services to implement the assessment recommendations. VHA PET centers collaborated on the design of the implementation process, which included initiating a multi-center VHA PET registry, supporting prospective research, and this updated systematic review.

To produce this report the MDRC Technology Assessment (TA) Program surveyed VHA PET facilities, used registry data, and conducted systematic reviews of the published PET literature from September 1996 through December 1998 for selected cancers and Alzheimer's disease. This report includes studies using positron emitting coincidence imaging with the radiopharmaceutical FDG to study cellular glucose metabolism.

Background
PET is a minimally invasive nuclear medicine imaging modality that uses the principle of *coincidence detection* to measure biochemical processes within tissues. PET may complement or supplant other imaging modalities, such as radiography, computed tomography (CT), or magnetic resonance imaging (MRI), which rely on predominantly anatomic definitions of disease.

Conventional positron emission coincidence imaging is accomplished using cameras specifically designed, or "dedicated," for imaging positron-emitting radioisotopes. Dual-headed gamma cameras are being adapted for coincidence imaging positron emitters (called "camera-based PET") as a lower cost and more accessible alternative to dedicated PET. Both PET systems have whole body scanning capability.

Key Findings

Cost and Reimbursement
A dedicated PET system costs from $800,000 to $2.5 million, and a cyclotron costs from $1.2 million to $1.7 million, in addition to the costs of installation, construction, and operation. Camera-based PET systems sell for about $850,000. Annual operating costs vary considerably. The charge for a PET scan will depend on these cost factors, as well as the clinical indication, the radiopharmaceutical used, and caseload.

Effective January 1, 1998 Medicare began offering interim provisional coverage for FDG-PET scans using either dedicated or camera-based PET for characterizing solitary pulmonary nodules and initial staging of suspected metastatic non-small cell lung cancer. On or after July 1, 1999 Medicare expanded coverage to include detecting and localizing recurrent colorectal cancer with a rising carcinoembryonic antigen, staging and characterizing Hodgkin's and non-Hodgkin's lymphoma in place of a gallium scan or lymphangiogram, and identifying metastases in melanoma recurrence in place of gallium studies.

The national average payment is $1,980 per scan, excluding the professional component. HCFA will collect and analyze claims data and data from other sources to determine the medical effectiveness of PET in managing these conditions, after which HCFA will decide the extent to which it should modify the coverage policy.

Regulation

Recent changes in FDA regulation now permit PET imaging facilities that manufacture radiopharmaceuticals on-site to continue in accordance with the positron emission compounding standards and the official monographs of the United States Pharmacopoeia. FDA has either approved or cleared for marketing both PET systems to image radionuclides in the body.

Experience in VHA

- VHA continues its moratorium on adding more dedicated PET facilities within the system. Many VA medical centers are modifying dual-headed gamma cameras for coincidence detection.
- A survey of active funded research at VHA PET sites underscores the importance of PET as a basic research tool. Most of the research is in neurology and cardiology and is funded by a range of private and public VA and non-VA sources.
- There has been an increase in the number of diagnostic PET scans, particularly in oncology. Lung cancer staging was the most common oncology indication among VHA PET sites in FY 1998.
- VHA is maximizing its investment in PET by developing a PET registry to collect critical patient information, funding rigorous, prospectively designed clinical research, and tracking the published peer-reviewed PET literature available in the public domain.
- The MDRC TA Program is coordinating a joint project with other members of the International Network of Agencies for Health Technology Assessment (INAHTA) to produce a report on the use of PET among countries represented by INAHTA members.

Evidence of effectiveness

The existing evidence argues against routine clinical use of PET for diagnosing Alzheimer's disease until more effective treatments and risk modification interventions for Alzheimer's disease are developed, and until meaningful and robust predictive values are obtained from an ongoing European multicenter PET study. The systematic reviews indicate that the data supporting the use of either dedicated or camera-based PET system with FDG in managing patients with selected cancers are deficient.

- The evidence for using camera-based PET in oncology is limited to one small preliminary study in the tertiary-care setting, comparing camera-based PET to dedicated PET using no suitable reference standard. Accordingly, it did not meet the inclusion criteria for this review.
- Included studies assessed dedicated PET as a complement to or replacement for anatomic imaging modalities, as a noninvasive alternative to invasive procedures, or as a method for increasing the diagnostic certainty for performing an invasive procedure. Studies focused on the technical feasibility of using dedicated PET and on defining diagnostic accuracy in the tertiary care setting.

- Studies generally enrolled highly selected patients and failed to adequately describe the previous work up or the size or composition of the referral base from which the patient sample was drawn. All had at least one of the methodologic biases often found in diagnostic imaging test evaluations, and their presence will tend to inflate estimates of diagnostic accuracy. Methods for defining disease on PET imaging have not been standardized and may limit the generalizability of findings across institutions.
- The few studies reporting the influence of PET on changes in diagnostic certainty and/or treatment planning were usually retrospective case series that were not originally designed to document these changes and were not systematically conducted or reported as such. Some authors used likelihood ratios and predictive values to define PET's clinical usefulness, but proper interpretation of these estimates is conditioned on what was known about the patient before the test and on deriving PET results independently of other test results. None of the studies met both conditions, and the influence of PET on diagnostic certainty and subsequent treatment planning could not be determined.

Conclusions/Recommendations

- VHA continues its commitment to delivering high quality patient care and to rational resource management through its support of VHA PET centers, carefully appraising the PET literature to identify areas in need of research, and funding rigorous, prospective clinical research.
- The prevailing evidence does not support the use of either dedicated or modified camera-based PET as a diagnostic test for the applications in this review. The TA Program identified several methodologically rigorous studies of other diagnostic imaging modalities that could serve as models for designing higher quality PET research.
- Systematic reviews from other technology assessment agencies, which used methods similar to VHA's, derived similar conclusions. As in VHA, patients with cancer constitute a considerable burden to the health systems represented by these agencies, and there is growing support for assessing either PET modality in the work up of these patients. Accordingly, agencies identified the uses for PET in oncology, particularly staging non-small cell lung cancer, as major topics for research.
- Several cooperative trials, including a VHA Cooperative Study of PET in solitary pulmonary nodules, are ongoing or planned. Clinicians should await the results of these efforts before incorporating PET into routine diagnostic strategies.
- Individuals interested in clinical PET would benefit from an accessible central repository containing information on existing and proposed rigorously designed cooperative trials of PET. This source could help guide the diffusion of PET into clinical care, as its usefulness and contribution to improved patient outcomes are appropriately evaluated.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

I. INTRODUCTION

VHA is committed to improved quality of care and outcomes for veterans and to rational resource management. As health care decision making transitions from a rationale based on resources and opinions to a rationale based on evidence from research, VHA uses technology assessment (TA) processes and information to guide evidence-based decisions. Health Services Research and Development Service, through the Management Decision and Research Center (MDRC), produces and disseminates TA information in the form of systematic reviews of the literature. VHA uses these reviews to support clinical policy and focus on areas in need of further research.

For example, after delivery of the original MDRC PET technology assessment (Flynn, 1996), the Under Secretary for Health directed the Office of Patient Care Services to implement the assessment findings and recommendations. As a result, VHA continued its moratorium on adding more dedicated PET scanners to its system. A new VHA cooperative study incorporated study design suggestions from the initial assessment. VHA PET Center Directors were instrumental in designing the implementation strategies, which included initiating a multi-center VHA PET registry, completing a rigorous single-site outcome study, and updating the 1996 MDRC PET systematic review.

In this update, the MDRC used evidence-based medicine frameworks and methodology to produce systematic reviews of the peer-reviewed PET literature from September 1996 through December 1998. It reviews the performance of dedicated PET systems and gamma camera systems with coincidence detection capabilities in selected cancers of the head and neck, breast, and colo-rectum, lung cancer staging, solitary pulmonary nodules, as well as Alzheimer's disease. The report also contains:

- clinical and research experience across VHA PET facilities;
- VHA implementation strategies for recommendations made in the first report;
- ongoing multi-site clinical trials of PET for the indications reviewed in the report;
- findings and recommendations from reviews of PET conducted by other technology assessment agencies; and
- a description of an international collaboration studying PET use among countries represented by the collaboration.

II. DESCRIPTION OF TECHNOLOGY

A. Instrumentation

Positron Emission Tomography (PET) is a minimally invasive nuclear medicine imaging modality that uses radiopharmaceuticals to capture and measure biochemical processes within tissues. PET, like other nuclear medicine techniques, defines disease in terms of quantifiably abnormal regional chemistry. PET may complement other imaging modalities, such as radiography, computed tomography (CT), or magnetic resonance imaging (MRI), which rely on predominantly anatomic definitions of disease.

PET imaging employs radioactive isotopes that decay by emitting a positively charged electron, called a positron, from the nucleus. The positron collides with a negatively

charged electron resulting in two high energy (511 keV) photons that travel in opposite directions. PET uses the principle of *coincidence detection* to form the raw image. That is, radiation detectors are arranged in a ring around the patient to allow for simultaneous (coincidence) detection of the two photons. The exact site of origin is recorded, and a cross-sectional image is displayed.

Dedicated PET systems are optimized for high energy dual photon coincidence detection. Two modified forms of single photon emission computed tomography (SPECT) are now available for imaging positron emitters and may be a less costly alternative to dedicated PET (Jarrit and Acton, 1996):

- dual-headed SPECT cameras adapted for coincidence detection, called "camera-based" PET, or
- multi-headed SPECT cameras adapted with special collimators for high energy (511keV) photon absorption.

Both Jarrit and Acton (1996) and Coleman (1997) emphasized that neither modified SPECT system is optimized for clinical use, particularly in oncology. Lower sensitivity restricts their use to studies using isotopes with longer half-lives, and performance and cost data comparing either system to dedicated PET are limited. These authors caution against the premature use of these systems, which could be detrimental to the future acceptance of both dedicated PET and modified PET systems.

In light of recent federal regulatory changes (See Section 111-Regulation and Reimbursement) this report will address only dual-headed gamma cameras adapted for coincidence imaging ("camera-based" PET) and dedicated PET systems.

B. Radiopharmaceutical

The most widely used radiopharmaceutical in PET imaging is the cyclotron-produced FDG. FDG is a D-glucose analog used to study cellular glucose metabolism. Since many diagnostic PET studies rely on FDG, its availability is critical to a facility that wishes to conduct clinical studies using either dedicated or camera-based PET systems.

C. Data analysis

PET and other nuclear medicine image patterns represent spatial and temporal arrangements of the physiological or biochemical process under investigation. There are many ways to detect and compare these patterns such as visual analysis of metabolic patterns, region of interest (ROI) analysis where the regions are hand-drawn or placed (sometimes with coregistration with anatomic images), and neural networks. PET data may be managed by using absolute metabolic values or by normalizing to a reference value to generate metabolic ratios.

D. Potential roles for PET

Flynn (1996) summarized the general rationale for the use of PET in oncology. PET may detect abnormalities in tissue biochemical and physiological processes caused by many forms of cancer. Reliance on tumor histology and anatomy limits the oncologist's tools

for selecting optimal treatment, and adding metabolic data from PET may expand the oncologist's ability to optimize treatment. Finally, monitoring metabolic responses to treatment could allow early redirection of therapy. Several potential applications for PET in oncology were noted:

- Detecting tumors (which may employ coregistration techniques that combine PET and anatomic imaging into a single image);
- Staging (particularly using whole-body imaging methods) although there is a lower limit to the size of metastases that can be detected by PET;
- Detecting local disease recurrence, since anatomically-based imaging is often limited by the effects of treatment;
- Predicting tumor response to chemotherapy; and
- Monitoring treatment.

Studies of Alzheimer's disease and other neurologic and psychiatric conditions predate studies of PET for other diagnostic applications and are prevalent in the PET literature. PET allows qualitative and quantitative evaluation of cerebral physiology and exploration of the biochemical bases for clinical diseases. FDG PET brain studies have been used for many research and clinical purposes related to the central nervous system. These include (Hoffman, 1993):

- defining the magnitude and distribution of normal local cerebral glucose metabolism, and the effects of age and sex on metabolism;
- locating seizure foci in patients with partial complex seizures who are potential surgical candidates for temporal lobectomy;
- assessing brain tumors for degree of malignancy at diagnosis, persistent post operative tumor, differentiating high- from low-grade tumors and radiation necrosis from persistent tumor;
- evaluating schizophrenia, affective disorders, obsessive-compulsive disorder;
- studying cerebral metabolism in cerebrovascular disease; and
- defining regions of altered glucose metabolism in various forms of dementia such as Alzheimer's disease, Pick's disease, and Huntington's disease.

Expanded roles for PET in selected applications will be discussed in Section VIII Published Findings for each application.

III. REGULATION AND REIMBURSEMENT

A. The Food and Drug Administration (FDA)

FDA has either approved or cleared for marketing dedicated PET scanners and coincident imaging gamma cameras to image radionuclides in the body. To date, the FDA has approved two PET radiopharmaceuticals for clinical use:

- Rubidium (^{82}Rb), limited to rest alone or rest with pharmacologic stress PET scans, is used for noninvasive imaging of the perfusion of the heart for the diagnosis and management of patients with known or suspected coronary artery disease.

- FDG indicated for identifying regions of abnormal glucose hypometabolism associated with foci of epileptic seizure. Approval for use is restricted to The Methodist Medical Center in Peoria, Illinois.

In the Food and Drug Modernization Act, which was signed into law on November 21, 1997, Congress directed the FDA to develop new approval procedures and appropriate current good manufacturing practice requirements for PET drug products. FDA may not require the submission of new or abbreviated new drug applications for PET drug products, which are not adulterated, for a period of 4 years after the date of enactment of the Modernization Act or for 2 years after FDA develops the new procedures, whichever is longer. FDA has begun developing these procedures.

In the meantime, PET drug products may be manufactured for clinical use providing they are produced in accordance with the positron emission compounding standards and the official monographs of the United States Pharmacopoeia. These standards are to assure that PET drug products are safe and have the identity, strength, quality, and purity that they are represented to possess.

B. Health Care Financing Administration (HCFA) and Medicare

A health technology review conducted by the Center for Practice and Technology Assessment (formerly the Office of Health Technology Assessment), Agency for Health Care Policy and Research (1998) provided the basis for Medicare's first coverage policy for PET scans performed on or after March 14, 1995 (HCFA, AB972760):

- PET scans using Rubidium (^{82}Rb), done at rest or with pharmacological stress, for noninvasive imaging of the perfusion of the heart for the diagnosis and management of patients with known or suspected coronary artery disease. Coverage is limited to PET scans used in place of SPECT or following an inconclusive SPECT scan, which provide information deemed necessary to determine treatment intervention.

In an agreement with the Chairman of the Senate Appropriations Committee in late 1997, the Secretary of Health and Human Services committed to expanding Medicare coverage of PET scans on an interim basis to include diagnosing solitary pulmonary nodules and initial lung cancer staging (Stevens, 1997). Effective January 1, 1998, FDG-PET scans will be covered when performed using either dedicated or camera-based PET system to image radionuclides in the body for the following conditions (HCFA, 3b4120):

- characterizing solitary pulmonary nodules (SPNs) for the primary purpose of determining the likelihood of malignancy to plan treatment. Coverage is limited to claims that include evidence of the initial detection of a primary lung tumor, usually by CT.

- initial staging of suspected metastatic non-small cell lung cancer in thoracic (mediastinal) lymph nodes in patients with pathologically confirmed primary lung tumor, but whose extent of disease has not yet been established. Coverage is limited to claims that include evidence of confirmed primary tumor, concurrent CT, and follow-up lymph node biopsy.

The use of routine biopsy following a negative PET scan is considered inappropriate in these conditions, and payment for biopsy will be denied unless the claim is supported by evidence explaining the medical necessity of the biopsy.

After an expedited review of scientific information presented at a town hall meeting in January 20-21, 1999, HCFA agreed to expand coverage for PET scans performed on or after July 1, 1999 to diagnose and manage the following three indications:

- detecting and localizing recurrent colorectal cancer with rising carcinoembryonic antigen (CEA);
- staging and characterizing both Hodgkin's and non-Hodgkin's lymphoma in place of a gallium scan or lymphangiogram; and
- identifying metastases in melanoma recurrence in place of gallium studies prior to surgery.

Table 1: Pricing of New PET Scan Indications Approved by HCFA*

HCPCS Codes	Description	National Average Payment for Technical Component**
G0125	PET lung imaging of solitary pulmonary nodules using FDG, following CT	$1,980
G0126	PET lung imaging for initial staging of solitary pulmonary nodules using FDG, following CT or of pathologically diagnosed non-small cell lung cancer	$1,980
G0163	PET, whole body, for recurrence of colorectal or colorectal metastatic cancer	$1,980
G0164	PET, whole body, for staging and characterizing lymphoma	$1,980
G0165	PET, whole body, for recurrence of melanoma or melanoma metastatic cancer	$1,980

*From www.hcfa.gov/pubforms/14%5Fcar/3b4120.htm
**technical component only, including payment for radiotracer, using revenue code 404. Claims for professional component should user modifier 26.

Medicare coverage is conditioned on the ability of PET to affect the management and treatment of patients with these cancers. HCFA will collect and analyze claims data, and data from other sources, to determine the medical effectiveness of PET in managing these conditions. After sufficient claims data have been collected, HCFA will decide the extent to which it should modify the coverage policy.

IV. ACCESS AND COST

The Institute for Clinical PET (1999) reports that there are nearly 147 facilities with coincidence detection capability in the United States. There are 10 dedicated PET facilities in the VHA system, making VHA one of the largest owners of dedicated PET scanners by any single health system in the world.

ECRI (1996) reports that the cost of a PET scanner ranges from $800,000 to $2.5 million, excluding costs associated with installation, construction, and operation, and a cyclotron costs from $1.2 million to $1.7 million. Annual operating costs vary considerably and may include personnel salaries, scanner and cyclotron supplies, service and maintenance contracts, equipment amortization, and other indirect costs. Ultimately, what a PET facility charges for a PET scan will depend on these factors, as well as the clinical indication, the radiopharmaceutical used, and caseload (Flynn, 1996).

Currently, there is a moratorium on adding PET facilities in VHA. Many VHA medical centers without access to PET facilities are adapting gamma cameras for coincidence imaging. The cost of upgrading dual-headed gamma cameras for coincidence imaging is approximately $250,000; dual-headed gamma cameras without the upgrade sells for about $600,000 (ECRI, 1996).

V. EXPERIENCE IN VHA

Table 2 lists VHA PET (dedicated) sites and their sharing partners. In all but two sites, both the camera and cyclotron are in the same location. However, ownership of the camera and cyclotron varies across sites (Flynn, 1996). All sites have access to FDG.

Table 2: VHA PET Facilities and Sharing Partners

VHA PET Facility	VISN	Facility Location	Sharing Partner
VA Connecticut Health Care System VAMC West Haven, Connecticut	1	VAMC	Yale University
VA West New York Health Care System VAMC Buffalo, New York	2	VAMC (cyclotron at sharing partner)	State University of New York at Buffalo
VA Pittsburgh Health Care System VAMC Pittsburgh, Pennsylvania	4	Sharing Partner	UPMC Health Systems- Presbyterian
Richard L. Roudabush VAMC Indianapolis, Indiana	11	Sharing partner	Indiana University
VAMC Ann Arbor, Michigan	11	Sharing Partner	University of Michigan Ann Arbor
VAMC Minneapolis, Minnesota	13	VAMC	None
St. Louis VA Medical Center St. Louis, Missouri	15	Sharing Partner	St. Louis University
VA South Texas Health Care System VAMC San Antonio, Texas	17	Sharing Partner-UTHSC	University of Texas Health Science Center
VA Palo Alto Health Care System VAMC Palo Alto, California	21	VAMC (no cyclotron, FDG purchased from private source)	None
VA Greater Los Angeles Healthcare System VAMC West Los Angeles, California	22	VAMC	Individual investigators

Research continues to constitute considerable activity conducted at VHA PET facilities. All VHA PET facilities were surveyed for a list of active funded research at their site. The results of this survey are listed in Appendix III. Most are multi-year studies with funding from a range of private and public VA and non-VA sponsors. The majority of funded PET research is for the study of neurologic conditions, followed by studies in cardiology.

The VA HSR&D Center for Practice Management and Outcomes Research, Office of Research and Development, provided FY 1998 utilization data from the VHA PET registry for the conditions in this report (See Table 3). Of the subjects that had radiopharmaceutical data available, nearly 70% were scanned using FDG, representing the radiopharmaceutical most often used across VHA PET sites.

Given the significant burden lung cancer represents in both the veteran and general populations, not surprisingly lung cancer was the major oncology diagnosis among VHA PET sites in FY 1998. Alzheimer's disease, colorectal cancer, and head and neck cancer have roughly equivalent numbers of veteran and non-veterans scanned, whereas non-veterans comprise a higher portion of subjects with breast cancer, as expected. The distribution of veterans and non-veterans within and across diagnoses may change as evidence of PET's clinical utility is clarified, or if reimbursement policies in either the public or private sector are altered.

Table 3: Diagnostic-specific Utilization Data Across VHA PET Facilities for FY 1998

Diagnosis	*# Veterans*	*# Non-veterans*	*Total (% of all neurology subjects)*
Alzheimer's disease	11	6	17 (3.4%)

Diagnosis	*# Veterans*	*# Non-veterans*	*Total (% of all oncology subjects)*
Lung cancer	246	192	438 (29.4%)*
Colorectal cancer	63	80	143 (9.5%)**
Breast cancer	1	34	35 (2.3%)
Head & neck cancer	58	52	110 (7.4%)

*excludes 8 patients with unknown veteran status
**excludes 2 patients with unknown veteran status

In the 1996 assessment, the TA Program recommended that VHA maximize the value derived from its existing commitment, rather than invest in additional PET centers, and suggested ways in which PET activities could be coordinated across the VHA system (See Preface). Since then, several suggestions have been implemented:

Develop and maintain a VHA PET registry

The VHA Office of Patient Care Services is providing recurring funding to the HSR&D Center for Practice Management and Outcomes Research in Ann Arbor, Michigan to develop and maintain a VHA PET registry. The Center is collecting annual facility utilization data and subject-specific data from all VHA PET facilities.

Support rigorous, prospectively designed clinical research

- The VHA Office of Patient Care Services is providing funding to the VHA Cooperative Studies Center and to the PET Center in West Haven, Connecticut to complete an outcome analysis. The study addresses clinical utility, cost, utilization of other diagnostic studies, and the impact of PET on treatment planning.

- VHA Cooperative Studies Program is funding a multi-year cooperative trial to evaluate the clinical utility of PET in characterizing solitary pulmonary nodules (See Appendix III, St. Louis). The Palo Alto Cooperative Studies Coordinating Center is monitoring the study. Six VHA PET sites and four non-VHA PET sites with VA affiliation are participating. Patient accrual started in August, 1998.

Results from these studies should clarify the evidence on the utility of FDG-PET in the management of patients with selected clinical conditions.

Conduct regular updates of the PET literature

The VHA Office of Patient Care Services also agreed to fund regular systematic review updates of the 1996 MDRC PET Technology Assessment.

VI. METHODS FOR THE SYSTEMATIC REVIEW

Information about the value of PET scanning in selected cancers and Alzheimer's disease was obtained by conducting a ***systematic review*** of the published literature. A systematic review uses a scientific approach to limit bias and to improve the accuracy of conclusions based on the available data. A systematic review addresses a focused clinical question, uses appropriate and explicit criteria to select studies for inclusion, conducts a comprehensive search, and appraises the validity of the individual studies in a reproducible manner. With respect to the diagnostic test literature, the point of a systematic review can be to examine the ultimate value or benefit derived from the test (Guyatt, 1995).

The MDRC uses a review protocol to guide the inclusion, analysis, and summary of evidence for this review (See Table 4 and Appendix 1). The protocol uses three analytic frameworks to appraise the literature, ensuring that studies are evaluated in a consistent, reproducible manner, and that studies included in the report conform to established scientific standards. These frameworks are critical to understanding the report analysis, conclusions, and recommendations.

Assign to Fryback and Thornbury hierarchical model of diagnostic efficacy

Fryback and Thornbury (1991) note that the localized view of the goal of diagnostic radiology would be that it provides the best images and the most accurate diagnoses possible. A more global view recognizes diagnostic radiology as part of a larger system of medical care whose goal is to treat patients effectively and efficiently. Viewed in this larger context, even high-quality images may not contribute to improved care in some instances, and images of lesser quality may be of great value in others.

Fryback and Thornbury (1991; 1992) present an evolving hierarchical model for assessing the efficacy of diagnostic imaging procedures. Their model, with a list of the types of measures that appear in the literature at each level in the hierarchy, is presented in Appendix I. Using this model, it is possible to follow the development of a diagnostic technology and to align current research efforts with a particular level of development.

Assess the quality of individual studies of diagnostic tests using evidence-based medicine criteria

This assessment has adopted evidence-based medicine criteria as a requirement for assignment of studies to the "diagnostic accuracy" level of the hierarchy. These criteria will be applied to individual studies in the report. If the criteria are not met, the study will generally be considered insufficiently rigorous to provide the basis for patient care decisions. However, such studies often provide useful information on the technical characteristics of a diagnostic test or may provide information necessary to subsequent diagnostic accuracy studies.

Evaluate the strength of the evidence supporting a causal link between the use of the technology and improved outcomes of care

Recommendations about the use of a technology should be linked to the quality of the available evidence, which ultimately depends on the strength of the evidence. The strength of the evidence relates to the overall research design and to the quality of the

implementation and analysis, i.e. how well bias and confounding factors are controlled in the design and conduct of a study. Attributes that strengthen the validity of findings include: ***randomized*** (vs. nonrandomized), ***controlled*** (vs. uncontrolled), ***blinded*** (vs. unblinded), ***prospective*** (vs. retrospective), ***large*** (vs. small), ***multi-site*** (vs. single site), and ***contemporaneous*** (vs. historical) ***controls.***

Table 4: Systematic Review Protocol

1) Conduct search of MEDLINE and other databases. Also search end references from retrieved articles and listings of English language, public domain technology assessments.

2) Apply inclusion criteria to search:
   - English language articles reporting primary data and published in a peer review journal (not abstracts)
   - studies ≥ 12 human subjects (not animal studies) with the disease of interest
   - studies using dedicated PET systems or gamma camera systems adapted with 511 keV coincidence imaging capability
   - studies using the radiopharmaceutical 2-[^{18}F]fluoro-2-D-glucose (FDG)
   - study not duplicated or superseded by subsequent study with the same purpose from the same institution
   - study design and methods clearly described (i.e. sufficient information to judge comparability of case and control groups, details of imaging protocol, whether visual or quantitative analysis of PET data used, or type of PET quantitative data analysis used)

3) Retrieve full text articles meeting inclusion criteria.

4) Review full text articles and assign to level of Fryback and Thornbury (1991) diagnostic efficacy hierarchy.

5) To assess methodologic quality, apply evidence based medicine criteria to studies of diagnostic tests:
   - clearly identified comparison groups, ≥ 1 of which is free of the target disorder.
   - either an objective diagnostic standard (e.g. a machine-produced laboratory result) or a contemporary clinical diagnostic standard (e.g. a venogram for deep venous thrombosis) with demonstrably reproducible criteria for any subjectively interpreted component (e.g., report of better-than-chance agreement among interpreters).
   - interpretation of the test without knowledge of the diagnostic standard result (no test review bias).
   - interpretation of the diagnostic standard without knowledge of the test result (no diagnostic review bias).

6) To further refine judgment of methodological quality, grade **diagnostic accuracy or thinking efficacy** studies:
   Grade A - Studies with broad generalizability to a variety of patients and no significant flaws in research methods
   Grade B - Studies with a narrower spectrum of generalizability, and with only a few flaws that are well described (and impact on conclusions can be assessed)
   Grade C - Studies with several methods flaws, small sample sizes, incomplete reporting or retrospective studies of diagnostic accuracy
   Grade D- Studies with multiple flaws in methods, no credible reference standard for diagnosis, evidence of work up, test review, or diagnostic review bias, or opinions without substantiating data

7) Evaluate quality of studies at each efficacy level; conduct meta analyses if appropriate.
8) Rank the evidence for the degree to which it supports a causal link between technology use and improved outcomes.

Modifications made to the grading system accounted for the degree to which bias could be reasonably minimized in the study design, given the nature of the clinical work up. More

common methods for minimizing the effects of bias are described in Appendix I. If the study provides evidence that the investigators reduced the effects of bias, the methodologic quality grade was advanced to the next highest level.

It should be noted that inclusion criteria could influence report findings. The inclusion criteria chosen for this report permit review of the best evidence available on the clinical use of FDG PET scans for selected conditions. These generally represent larger controlled studies published in the peer-reviewed literature. A limitation of this analysis is the potential language bias owing to including only English language articles. Thus, the reader should keep in mind that the findings and recommendations are based only on evidence that meets criteria for inclusion in the report.

VII. APPRAISAL OF THE LITERATURE

For this update, titles and abstracts of 474 references were screened. Sixty-four references were determined to be relevant, and their full text articles were reviewed for potential inclusion in the systematic review. Additional articles were retrieved to provide background materials about the technology and selected clinical applications.

Forty-seven articles from the database searches and from end references of initially retrieved articles met the inclusion criteria for review. Each included study was classified according to clinical condition and assigned to a diagnostic efficacy level as follows:

Efficacy level*	Head & Neck	Breast	Lung staging	SPN	Colorectal	Alzheimer's
Technical	4	4	7	1	2	8
Diagnostic accuracy	3	6	7	2	3	0
Diagnostic thinking			?	?		
Therapeutic			?		?	
Patient outcome						
Societal						

* Adapted from Fryback and Thornbury, 1991
? Anecdotal data also presented in diagnostic accuracy studies.

In all oncology areas, higher levels of studies in the diagnostic test hierarchy superseded technical efficacy (feasibility) studies, represented the best evidence on the efficacy of FDG PET, and were summarized for this review. Technical efficacy studies are listed in the references. In Alzheimer's disease, only technical efficacy studies met the inclusion criteria for review.

All but one of the included studies were **single-site studies** classified as **case series** (Level V evidence), representing a relatively weak study design that does not provide strong evidence of effectiveness. Case series contain useful information about the clinical course and prognosis of patients, can *suggest* relationships between interventions and outcomes, and can generate ideas for further research. All studies used patients with no disease or with benign disease as **internal controls.**

All included studies used dedicated PET systems. The TA Program identified only one preliminary study using camera-based PET in oncology (Shreve, 1998). These authors compared blinded readings of camera-based PET images, using attenuation-corrected dedicated PET as the standard of reference, in 31 patients with known or suspected tumors. Accordingly, it did not meet criteria for inclusion in this review. The results are summarized below.

Table 5: Summary of the Technical Efficacy of Camera-based PET in 31 Patients with 109 Lesions

Site	*Short-axis diameter (cm) Range, mean*	*# lesions detected on camera-based PET*	*# lesions detected on dedicated PET*
Lung	0.9-4.0, 2.7	13	14
Mediastinum	0.6-1.3, 1.0	5	15
Mediastinum	1.5-3.5, 2.2	15	16
Axilla	1.2-1.5, 1.3	5	9
Head and neck	1.1-2.4, 1.7	5	7
Abdomen	1.2-6.3, 2.8	6	26
Skeleton	Not available, could not be determined	11	22

The authors concluded that camera-based FDG PET could depict many of the lesions depicted with dedicated PET. Detection of lesions using camera-based PET was greatest in the lung and poorest in the abdomen and in all sites, excluding the lungs, for tumors generally less than 1.5 cm in short-axis diameter. The results of this preliminary study require valid estimates of diagnostic accuracy and marginal value using an appropriate reference standard in order to establish camera-based PET as a diagnostic tool.

A. Data Synthesis

This report presents a qualitative overview to synthesize the best available evidence. A quantitative synthesis (meta-analysis) was not attempted. The methodological weaknesses of case series, combined with present differences in design and analysis among the eligible studies, argued against the validity and usefulness of pooling study results (Eysenck, 1994).

VIII. PUBLISHED FINDINGS

Background information on each clinical condition such as risk factors, diagnosis, alternative diagnostic modalities, staging, treatment and survival was described in detail in the first MDRC PET report (Flynn, 1996), and will not be presented here. A brief synopsis of updated epidemiological information and an account of the potential role(s) for PET are presented for each condition in addition to critical evaluation of the literature.

Epidemiological information for oncology conditions in this report is supplied by the American Cancer Society (American Cancer Society, 1998). Data on the veteran population are provided by the 1997 Annual Report of the Secretary of Veterans Affairs (West, 1998).

Results are presented according to the potential role of PET in the management of each disease. Full data abstraction tables of the best evidence of PET for each cancer section are found in Appendix IV.

A. Head and Neck Cancer

This report will define head and neck cancer as the common squamous cell carcinomas of the oral cavity, nasal cavity and paranasal sinuses, pharynx, and larynx. Skin, brain, thyroid, and salivary gland tumors and the rare tumors of other histopathologic types (sarcomas and lymphomas) that can have primary sites in the head and neck will not be discussed.

Approximately 41,400 new cases of head and neck cancer (3% of all incident cases of all types of cancer) and 12,300 deaths (2% of all cancer-related deaths) attributed to head and neck cancer are estimated for the United States in 1998. Within Veterans Health Administration malignant neoplasms of the lip, oral cavity, and pharynx (not larynx) accounted for 2,259 total discharges (0.3% of all discharges), with an average length of stay of 18.5 days, in FY 1997.

Nearly one-third of patients with head and neck cancer has lower stage, confined disease at diagnosis. Most of the remaining patients have locally or regionally advanced disease including spread to lymph nodes in the neck. Less frequent is head and neck cancer that has metastasized beyond the neck region (e.g., brain, lung, bone, or liver), at initial diagnosis. Accordingly, standard therapy emphasizes local and regional approaches (surgery, radiation therapy, or combination) with curative intent.

Chemotherapy is increasingly being added to standard therapy to improve the outcome of patients with locally advanced disease (PDQ®; 1999). For resectable disease neoadjuvant chemotherapy is incorporated into many organ preservation strategies to shrink tumors preoperatively and may improve locoregional control. Organ preservation approaches using concomitant chemotherapy with radiation are advocated in patients with unresectable disease.

Diagnostic tests are used at several points in the initial work up and treatment of head and neck cancer. These include delineating disease at the primary site (including locating unknown primary), identifying early nodal metastases, monitoring results of treatment, and identifying persistent and recurrent disease. CT and MRI have improved detection of occult cervical metastases for patients with head and neck cancer and subsequent management of patients at high risk of cervical metastases.

However, improvements are still needed to define the primary site and in the other points in the work up mentioned above. The ability to assess response to chemotherapy-radiation organ preservation approaches is becoming increasingly more important, since surgical excision would be indicated in the event of treatment failure. The functional information on glucose metabolism in head and neck tumors supplied by FDG PET could be clinically useful.

Table 6 depicts the study elements and Table 7 summarizes the data and quality of individual studies of PET using FDG in head and neck cancer.

Detecting unknown primaries in patients with metastatic cervical nodes

Braams (1997), a small technical feasibility study, detected unknown primaries in 13 patients with various histologic types of cervical metastases (see reference list). They performed whole-body PET followed by endoscopy, after physical exam and MRI and/or CT of the head and neck area failed to detect the primary tumor. PET identified the primary tumor in four (30%) patients and missed one small tumor (4mm) in another. Follow up over 18 to 30 months revealed no primary lesion in the remaining eight patients. The authors suggested that PET may be useful in guiding endoscopic exam and in identifying the primary site to direct more appropriate treatment.

Detecting primary disease

The MDRC Technology Assessment Program was unable to locate any PET studies that met evidenced-based criteria for diagnosing primary disease.

Detecting cervical node metastases

Two studies in Table 6 met some of the evidence-based medicine criteria for diagnostic test evaluations. Wong (1997) evaluated 16 patients, who had neck dissections, from a consecutive case series of 54 patients with known primary disease or with suspected recurrence or residual disease. Data suggest comparable performance of PET to anatomic imaging and improved performance over clinical exam across patients with a range of stages, but a test of statistical significance was not reported. In a small number of patients with occult nodal (N0) disease, PET did not perform as well as in patients with more advanced disease. In addition to small sample size in the subgroup analyses, several aspects of the study design were either unclear or not reported making the efficacy of PET difficult to determine.

In a retrospective evaluation of 14 patients with N0 disease on clinical exam, Myers (1998) reported a trend of increased accuracy of PET, although not statistically significant, over CT. PET combined with CT showed even greater improvement. Data were analyzed by dissected side and not by patient, and important study design elements were not reported.

Monitoring treatment response

Lowe (1997) presented preliminary data on 28 consecutive patients with advanced head and neck cancer, who were enrolled in a neoadjuvant organ-preservation protocol, to assess PET in evaluating tumor response to chemotherapy. The methods were reasonably well described, and the study met all evidence-based medicine criteria for diagnostic test evaluations. The data suggest good face accuracy of PET in distinguishing complete response from residual disease. Wide confidence intervals reflect a small study size, and no comparison data were presented.

The authors commented that while a positive PET scan may be indicative of residual tumor and warrant repeat tissue sampling or resection, a negative PET scan may also call for tissue sampling to rule out false negative results. They also stated that PET may be used in situations when sampling bias is more likely, for example, difficult access, questionable post-therapy biopsy results, or normal, reepithelialized appearance of the tumor site post-therapy.

Detecting recurrent disease

Wong and associates (1997) assessed PET prospectively for detecting both primary site recurrence in 12 patients and nodal recurrence in 13 post-treatment patients. PET showed high sensitivity in detecting recurrence at the primary site, but they presented no comparison data. For detecting nodal recurrence, PET was more sensitive than CT or MRI, was equal to clinical exam, and had superior specificity to both anatomic imaging and clinical exam.

Table 6: Characteristics of Diagnostic Accuracy Studies of FDG-PET in Patients with Head and Neck Cancer

Study characteristics	*Lowe et al. (1997)*	*Wong et al. (1997)*	*Myers et al. (1998)*
Perspective	? prospective	prospective	retrospective
Patient source	Consecutive patients between December 1994 and May 1996 with head and neck cancer: • 28 with stage III/IV who were participating in a neoadjuvant organ-preservation protocol using Taxol and carboplatin	54 consecutive patients who presented to head and neck clinics at two hospitals 31 with primary disease (TI=2 T2=10 T3=9 T4=10), 23 with suspected recurrence or residual disease) • 16 had neck dissections	116 patients diagnosed with head and neck cancer, of which 72 had biopsy-proven SCC and 26 underwent neck dissections: • 14 patients with N0 disease (24 total neck dissections) on clinical exam
Extent of disease (# patients)	Stage III=3 Stage IV =25	N0=8 N1=4 N2a=2 N2b=2	Stage I=1 Stage II=8 Stage III=2 Stage IV=3
Benign conditions	6 patients with pathologic complete response	None reported	None reported
PET criteria for positive result	1,2,or 3 on a 4 point scale	Not reported	Not reported
Contrast CT criteria for positive node	N/A	Standard size and morphological criteria used to assess nodal disease on CT/MRI	Not reported
Interpretation	• Blinded visual consensus using a before and after comparison format • 4-point scale • two readers	Not reported	Not reported
Gold standard determination (# patients)	Pathologic complete response or residual disease based on post therapy biopsies obtained after PET blinded to PET data (28)	• independent biopsy (16) • All suspicious areas of aerodigestive tract were biopsied	Histopathology for number of nodes, presence of malignancy, and extracapsular spread (14)
Data analysis	By patient	By patient	By dissection

Summary/Discussion

Since the 1996 MDRC PET report seven additional studies (three of diagnostic accuracy) of PET in head and neck cancer were published, met the inclusion criteria, and were reviewed. Evaluations of PET in head and neck cancer have focused mainly on detecting cervical node metastases in patients with known primaries, diagnosing disease recurrence, and monitoring response to treatment.

PET has potential uses at several points in the diagnosis and management of head and neck cancer patients. An early step in defining these uses is obtaining estimates of diagnostic accuracy. Only Lowe and associates (1997) met all evidence-based medicine criteria for diagnostic test evaluations, and the methods were reasonably well described. The two other studies did not report blinding of test interpreters and had other methodologic limitations, which affect the validity of the results, and it was unclear whether PET was used in addition to, or as a substitute for, other tests. All of the studies in Table 7 received low methodologic quality scores due to presence of significant bias, insufficient reporting and/or small sample sizes. The diagnostic accuracy estimates from these studies should be interpreted cautiously.

Information from a whole-body PET scan could have important treatment implications for patients with head and neck cancer. For example, identifying the primary tumor site not detected by other modalities could alter treatment planning. If the primary is from the head and neck, it is potentially curable with surgery and/or radiation therapy, whereas if the primary is located elsewhere, less toxic palliative treatment can be given. While there is a lower limit to the size of tumor that can be detected by PET, if validated in larger, rigorous studies, more accurate staging with PET could result in more appropriate treatment.

Minn et al (1997) (see reference list) assessed the feasibility of FDG uptake to predict cancer aggressiveness and survival. The results from 37 patients with primarily advanced Stage III/IV disease suggested a correlation between FDG uptake and prognostic significance on univariate analysis but not on multivariate analysis. Using FDG uptake to identify high-risk patients who would benefit from post-treatment surveillance requires further comparative study. Nonetheless, the wide range of primary sites and stages of head and neck cancer and the associated wide range of site-specific treatment and outcomes would complicate such evaluations of PET.

Accurate diagnosis of disease recurrence is critical to the treating clinician. With the addition of chemotherapy to many organ-sparing protocols, the ability to accurately assess nonsurgical treatment failure becomes increasingly more important to judicious surgical salvage. For patients who become symptomatic or who develop a mass during post-therapy surveillance, PET must be able to distinguish recurrence from treatment-related inflammation or fibrosis.

Goodwin (1998) suggested ways to improve such evaluations of PET that may provide more useful data to the treating physician. A prospective study of these patients, rather than a retrospective study of patients who had PET for various reasons and at various times after treatment, would more appropriately address the clinical issue. Pretreating patients with steroids or antibiotics to reduce inflammation might enhance the positive predictive value of PET. Other considerations include cost-effectiveness and capturing individual patient history, such as the timing of signs and symptoms after completion of therapy.

Controlled, prospective, blinded studies are needed to define the utility of PET (either dedicated or camera-based systems) relative to other imaging modalities in patients with head and neck cancer. Multiple sites may be needed to accrue a sufficient number of patients. Results from this updated literature review confirm the conclusions and recommendations from the first report (see Preface).

Table 7: Summary of Diagnostic Accuracy Studies of FDG-PET in Head and Neck Cancer

H = histology; F = Follow up; S = small size; R = referral bias; W = work up bias; T = test review bias; D = diagnostic review bias (upper case indicates significant limitation; lower case indicates limitation minimized by study design, presence of bias unclear, or small effect on operating characteristics)

Role	*Study*	*N*	*Operating Characteristics*				*Evidence-based Medicine Criteria*			*Study Design Limitations*	*Methodologic Quality Grade*
			PET	*CT*	*MRI*	*Other*	*Comparison group*	*Gold standard*	*blinding*		
Detecting nodal metastases	Wong 1997	12 positive cases 4 negative cases	Se=67%	CT + MRI Se=67%		Clinical exam Se=58%	+	H	—	**S,R,W,t**	**D**
	Myers 1998	9 positive dissections 15 negative dissections	Se=78% Sp=100% PPV=100% NPV=88% Acc=92%**(P=0.11)	Se=57% Sp=90% PPV=80% NPV=75% Acc=76%*		PET + CT Se=86% Sp=100% PPV=100% NPV=91% Acc=95%	+	H	—	**S,r,W,t,d**	**D**
Detecting local recurrence	Wong 1997	10 positive cases 2 negative cases	Se=100%				+	H	—	**S,R,W,t**	**D**
	Lowe 1997	21 positive cases 6 negative cases	Se=90% (77-100%) Sp=83% (53-100%) PPV=95% NPV=71% Accuaracy=89%				+	H	+	**S,R**	**C**
Detecting nodal recurrence	Wong 1997	8 positive cases 5 negative cases	Se=100% Sp=100%	CT + MRI Se=75% Sp=80%		Clinical exam Se=100% Sp=60%	+	H	—	**S,R,W,t**	**D**

B. Breast cancer

The American Cancer Society estimates 180,300 new cases (178,700 women and 1,600 men) of breast cancer will be diagnosed in 1998 in the United States. After a 4% per year increase in the 1980s, breast cancer incidence rates have leveled off in recent years to about 110 cases per 100,000. An estimated 43,500 women and 400 men will die of breast cancer in 1998, making breast cancer the second major cause of cancer death in women. Mortality rates continue to decline, particularly in younger women, likely due to earlier detection and improved treatment.

In FY 1997, there were 1.2 million female veterans (4.8% of all veterans) living in the United States, and the percentage of females in the veteran population is expected to increase. In accordance with the Women Veterans Health Program Act of 1992, Health Services Research and Development supports research to increase outreach and access to health care and to explore health issues that affect many women, including breast cancer (Feussner, 1997). VHA has also established the Mammography Quality Standards Office and has made available a nationwide toll-free mammography information line (888-492-7844) to expand mammography services to female veterans.

Potential applications for PET in breast cancer management were defined previously (Flynn, 1996):

- Non-surgical evaluation of breast disease;
- Staging recurrent disease;
- Quantifying tumor glycolytic rate as a prognostic factor;
- Monitoring response to therapy;
- Patient selection for axillary dissection and for preoperative therapy;
- Screening in subgroups of women (eg, those with breast implants, with prior breast radiotherapy, multiple breast masses and history of negative biopsy results, or severely fibrocystic breasts).

Table 10 summarizes the data and quality of individual studies of PET using FDG in breast cancer. Only studies of dedicated PET for non-surgical diagnosis of breast disease, patient selection for axillary dissection, and staging recurrent/metastatic disease met the inclusion criteria for this review. Three studies evaluated quantitative indices of FDG uptake as an indicator of prognosis. These studies were classified as technical efficacy due to their preliminary nature and will be discussed in the Summary/Discussion section.

Defining unknown primary disease

Palmedo (1997) prospectively compared PET to scintimammography (SMM) using ^{99m}Tc MIBI in the pre-surgical evaluation of 20 patients with 22 suspicious primary lesions detected by clinical exam or mammography. The mean lesion size was 29mm (range 8-53mm), of which only 3 patients had lesions smaller than 9mm. Quantitative analysis of tracer uptake was also performed to characterize disease, but no cut-off value was defined prospectively. Anecdotal data suggested that PET was superior to SMM in detecting axillary lymph involvement, but

neither test could determine extent of disease. The authors stressed that the menstrual cycle and age, which can alter MIBI uptake and FDG uptake, respectively, in normal tissue and the methods used to calculate FDG uptake could affect test accuracy.

Detecting axillary lymph node involvement

The three studies in Table 8 met the inclusion criteria for review. Utech (1996), Crippa (1998), and Adler (1997) compared PET to axillary lymph node dissection (ALND) in patients with either suspected or confirmed breast cancer who were scheduled for axillary staging. Therapeutic decisions at surgery were based on clinical and routine imaging results, including mammography. PET was added in the test sequence after the routine work up as a potential noninvasive method for staging the axilla, the rationale being that a negative PET scan might obviate the need for ALND in selected patients and, thus, decrease the morbidity and costs associated with the procedure.

All were prospective studies, but only Crippa (1998) reported a consecutive series. The evidence for the use of PET in staging the axilla is confined to a select group of patients with a high prevalence of malignancy and few benign conditions. The extent of axillary disease, reported in two studies, was limited to patients with metastases to ipsilateral axillary nodes. Crippa (1998) provided limited evidence from small subgroups on the ability of PET to determine extent of disease, which is an important prognostic indicator; not surprisingly, PET sensitivity improved with more advanced disease.

Two studies used multiple readers to interpret PET images, but neither study assessed interobserver variability. Of note, Adler (1997) used a higher dose of tracer and longer scanning times than were used in other studies. All studies reported some evidence of blinding to the gold standard, but none met strict evidence-based criteria for blinding. Patient and disease characteristics, study design elements, and units of analysis varied across studies, and many study design elements were incompletely described or not reported, making the validity of these results difficult to assess.

Table 8: Characteristics of Prospective Studies of Axillary Lymph Node (N) Staging With FDG-PET in Patients with Potentially Operable Breast Cancer

Note: All studies included primary tumors of mixed histologies, primarily invasive ductal carcinoma.

Study Characteristics	*Utech et al. (1996)*	*Crippa et al. (1998)*	*Adler et al. (1997)*
Patient source	124 patients with newly diagnosed and histologically proven breast cancer prior to therapy • 64 patients with metastatic nodes • 60 w/ surgically negative axilla • ? consecutive series	68 consecutive patients (72 total axilla) with palpable breast nodules scheduled for surgery based on clinical and mammography/ultrasound results • 61 had ALND • no ALND in patients with benign lesions (8) and in situ ductal carcinoma (3)	From a larger prospective study of PET, 50 patients with 52 axillary dissections who met inclusion criteria: • age ≥ 30 years • ≥ 2 ALND within 3 mo. Of PET scan • ≥ 10 nodes dissected • ability to fast ≥ 4 hours • ?consecutive series
Exclusion criteria (# patients)	Hyperglycemic patients	None reported	• History of ipsilateral axillary lymph node dissection • Preoperative systemic therapy • Primary tumor < 5mm • Uninterpretable PET scan (2)
Benign conditions of breast (#patients)	None	• proliferative dysplasia without atypica (6) • focal inflammation (2)	None
Primary tumor size (mean, range)	Reported as: <1cm=16 >1cm=49 >2cm=30 >3cm=29	2.0 cm, 0.4-6.7cm	Reported as: T0=1 T1=31 T2=17 T3=3
Prevalence of confirmed N metastases (# positive patients/total patients)	44/124=35%	27/61=44%	20/52=38% (by axilla)
Extent of N metastases (# patients)	N0=79 N1=43 N2=2 • one with bilateral disease	N0=36 (# axilla) N1a=21 N1b=13 N2=2	Not reported
Axillary node size	Not reported	Not reported	Range <0.1cm-2.5cm
PET criteria for positive node	discrete focal uptake > background	focal uptake > surrounding tissue)	increased FDG uptake and scan quality; scores ≥ 3= positive on a 5-point scale
Interpretation	• 3 radiologists + 1 nuclear medicine • blinded to all data except primary tumor	• # readers not reported • blinded to histopathology, but to other information not reported	• two readers • independent, blinded to all but axilla side • discrepancies resolved by consensus
Gold standard determination (# patients)	• histology (104) • histology + follow up (20) • extensive nodal sampling (average #/patient=19, range 7-46)	• histology (61) • Extensive nodal sampling (average # /axilla=21, range 12-38)	• histology (50) • extensive nodal sampling (average #/patient=17, range not reported)
Data analysis	By patient	By axilla	By axilla

ALND=axillary lymph node dissection

Detecting recurrence and metastases

The two studies in Table 9 presented the best evidence on the use of PET to stage recurrent disease and metastases in breast cancer patients.

Table 9: Characteristics of Studies Using FDG PET to Stage Recurrent Disease and Metastases in Patients with Breast Cancer

Note: Both were retrospective studies.

Study Characteristics	*Bender et al. (1997)*	*Moon et al. (1998)*
Patient source	75 patients with suspected recurrent or with metastatic disease in undecided or equivocal cases • Includes results from CT/MRI • 63 patients had both PET and CT/MRI data available for comparison • ?consecutive series	57 female patients (83 lesion sites) with a clinical suspicion of recurrence not resolved by conventional imaging: • who underwent primary surgery with or without adjuvant chemo- or radiation therapy and • who were referred to the UCLA PET center from October 1990 to October 1995 • ?consecutive series
Exclusion criteria (# patients)	None reported	• patients who underwent chemo-or radiation therapy within 3 mo before PET • lesions that were biopsied • lesions diagnosed with known disease
Benign conditions of breast (#patients)	None	(# sites) • seroma (1) • muscle uptake (5) • thyroiditis (1) • radiation pneumonitis (1) • blood pool of great vessels (2) • osteoarthritis (1) • intestine (1) • unknown (6)
Primary tumor histology	Well-differentiated ductal carcinoma (46) Infiltrating lobular carcinoma (10)	Not reported
Prevalence of confirmed local recurrence (# patients)	14/63=22%	29/57=51%
Prevalence of confirmed N metastases (# positive patients/total patients)	17/63=27%	8/26=31% (reported by lesion site)
Extent of M metastases (# patients)	• Bone (15) • Lung (5) • Liver (2)	• Bone (16) • Lung/Chest wall (7) • Liver (2)
PET criteria for positive lesion	4 point qualitative scale (intense, moderate, low, none) • Positivity criteria not defined	5 point qualitative scale • scores $\geq$ 3=positive
CT/MRI criteria for positive lesion	not defined	N/A
Interpretation	• 2 readers • independent • not blinded to other data	• 3 readers, discrepancies resolved by 4^{h} reader • independent • blinded to histology but aware of suspicion of metastases
Gold standard determination (# patients)	• histology (71) • follow up (4)	• histology • lesion morphology on 2 or more conventional imaging studies • $\geq$ 6 months of clinical and radiographic follow up after PET
Data analysis	By patient	By patient and by lesion

Both studies were retrospective case series of patients with suspected recurrence and/or metastases and equivocal findings after conventional imaging. PET was used as a complement to conventional imaging. It was unclear whether the patients in these studies represented consecutive case series. It should be noted that Bender (1997) presented data on 75 patients, but only 63 patients had information on both PET and CT/MRI for direct comparison. Few benign

conditions were represented in either study. This may be an artifact of the work up, and the benign cases were likely identified prior to inclusion. Both studies had a higher proportion of patients with metastases to the bone than to lung and /or chest wall, or liver. It was difficult to compare other characteristics of the patient population across studies due to incomplete reporting or variations in the units of analysis.

Both studies used qualitative scales to define lesions on imaging and multiple readers to interpret the images. Moon (1998) presented some data on interobserver variability. Moon (1998) met most of the evidence-based medicine criteria for blinding, but Bender (1997) did not blind interpreters to other data.

Summary/Discussion

PET has several potential uses in the management of patients with breast cancer. Since 1996, four technical efficacy and six diagnostic accuracy efficacy studies were published that met inclusion criteria for the review, representing the best evidence supporting the use of PET in breast cancer management to date. No new studies were identified that assessed the role of PET in evaluating response to treatment or screening in subgroups of women, such as women with radiodense breasts or breast implants.

The evidence on the ability of PET to detect unknown primary disease for this report is limited to one small study comprising a select group with a high prevalence of malignancy and few patients with small primary lesions less than 1cm. Limitations in study design and reporting suggest the preliminary nature of this study. The results should be confirmed in a larger group of patients with a range of tumor sizes, benign conditions and stages of disease. Newer PET models with higher resolution and availability of new dedicated breast PET scanners may improve detection of smaller lesions (Wahl, 1998).

The current best evidence, derived exclusively from case series of patients with a high prevalence of malignancy and with few benign conditions, does not support the routine use of PET as the initial test in patient selection for ALND. At face value, the operating characteristics from these studies suggest that PET has a relatively high sensitivity with a lower positive predictive value and a correspondingly lower specificity with a higher negative predictive value as compared to ALND. PET also yielded a fair number of false positives, many of which could not be explained. Some of the more recent studies are larger, but methodologic biases and incomplete reporting justified low methodologic quality scores.

Variations in the characteristics of the study populations, scanning techniques, and in the units of analysis may affect the generalizability of these results, particularly to mammographically tested populations, which typically have a lower prevalence of malignancy. Predictive values and other estimates of diagnostic accuracy should be interpreted with caution.

ALND with histopathology of dissected nodes supplies critical information to treatment management, is currently recommended by the NCI for most patients with Stage 1 or higher disease, but is associated with significant morbidity. Relative to other studies of screening and treatment options, published PET data to date are based on small numbers of patients. Moreover, the lower boundary of resolution limits the ability of current PET modalities to detect tumors less than 1cm in diameter. The consequences of false negative PET results in the absence of ALND in patients for whom effective treatment is available should be avoided.

The potential for PET to visualize the internal mammary nodes (potentially N3 disease) has been reported (Wahl, 1998). An NCI-sponsored multi-center trial is evaluating the accuracy of PET in staging the axilla and will include patients with N3 disease (See Section IX). Clinicians should await the results of this study before incorporating PET into routine clinical practice.

Likewise, the evidence on use of PET in detecting recurrent disease and metastases and defining unknown breast disease is in its early stages. PET was typically part of a testing sequence, but the marginal value of PET in the work up of these patients remains to be determined. The authors emphasized, and the TA Program concurs with, the need for further studies to assess the clinical impact of PET in the management of recurrent breast cancer.

Utech (1996), Crippa (1998), and Oshida (1998) (See technical efficacy list in Reference Section) presented some evidence on the feasibility of using quantitative FDG PET uptake by either the primary tumor or axillary lymph nodes as a prognostic indicator. Any attempt to correlate PET data with survival requires knowledge of the underlying characteristics of the study population and sufficient follow up time to track survival (Laupacis, 1994). The range of disease stages and corresponding treatment options would further confound the results. Large, rigorous studies are needed to define the utility of PET as a prognostic test.

Controlled, prospective, blinded studies are needed to define the utility of PET (either dedicated or camera-based systems) relative to other imaging modalities in patients with breast cancer. Multiple sites may be needed to accrue a sufficient number of patients. Results from this updated literature review confirm the conclusions and recommendations from the first report (see Preface).

Table 10: Summary of Diagnostic Accuracy Studies of PET and Alternatives in Breast Cancer

H = histology; F = Follow up; S = small size; R = referral bias; W = work up bias; T = test review bias; D = diagnostic review bias (upper case indicates significant limitation; lower case indicates limitation minimized by study design, presence of bias unclear, or small effect on operating characteristics)

Role	*Study*	*N*	*Operating Characteristics*			*Evidence-based Medicine Criteria*			*Study Design Limitations*	*Methodologic Quality Grade*
			PET	*SMM*	*CT/MRI*	*Comparison group*	*Gold standard*	*blinding*		
Defining unknown primary disease	Palmedo 1997	13 malignant lesions 7 benign lesions in 20 cases	Se=92% Sp=86%	Se=92% Sp=86%		+	H	+	**S,r,t,d**	**C**
Detecting axillary node involvement	Utech 1996	44 positive cases 80 negative cases	Se=100% Sp=75% PPV=69% NPV=100% Acc=84%			+	H + F	+	**r,d**	**C**
	Crippa 1998	27 positive axilla 45 negative axilla in 61 cases	Se=85% Sp=91% PPV=85% NPV=91% Acc=89% (overall values reported)			+	H	+	**S,r,t,d**	**C**
	Adler 1997	20 positive axilla 32 negative axilla in 50 cases	Se=95% Sp=66% PPV=63% NPV=95% Acc=77% (overall values)			+	H	+	**S,d,r**	**C**
Detecting recurrence or distant metastases	Bender 1997	54 positive cases 9 negative cases	Se=73-100% Sp=93-96% PPV=85-88% NPV=92-100% Acc=90-97%		Se=46-91% Sp=95-98% PPV=88-91% NPV=86-98% Acc=87-97%	+	H+F	—	**S,r,w,T,d**	**D**
	Moon 1998	29 positive cases 28 negative cases	Se=93% Sp=61-79% PPV=82% NPV=92% (overall values)			+	H+F	+	**S,R,t,d**	**C**

SMM = scintimammography with ^{99m}Tc-MIBI

C. Non-Small Cell Lung Cancer

Bronchogenic carcinoma, classified as either small cell or non-small cell, comprises 95% of all primary lung cancers. *This section will address only non-small cell varieties, as they constitute the majority (75%) of all bronchogenic carcinomas and, when localized, have the potential for cure with surgical resection.*

Bronchogenic carcinoma is the leading cause of cancer death in the United States. In 1998 the American Cancer Society estimates 171,500 new cases of lung cancer and 160,100 deaths from lung cancer. Malignant neoplasms of the bronchus and lung accounted for 9,730 discharges (1.5% of all discharges) with an average length of stay of 13.8 days within the Veterans Health Administration in FY 1997.

Non-small cell lung cancers (NSCLC) include adenocarcinoma (including bronchioalveolar), squamous (or epidermoid) cell carcinoma, and large cell (including large cell anaplastic) carcinoma. While 5-15% of NSCLCs are incidental findings on a chest x-ray, the vast majority of patients have symptomatic, advanced disease at clinical presentation.

Initial diagnosis is based on complete history, physical exam, and chest x-ray. If cancer is suspected, then staging is needed to assess the extent of local and distant disease. Stage of disease is the primary predictor of response to treatment and one of the important predictors of survival.

CT is the preferred diagnostic imaging test and is used at several points in the management of a patient with lung cancer: 1) to stage disease; 2) to evaluate treatment response; and 3) to differentiate recurrent disease from fibrosis. Use of other diagnostic imaging technologies to stage lung cancer is circumscribed largely because of technical limitations, availability, and cost.

CT provides morphologic (typically size) detail of the disease site. Accordingly, disease status of mediastinal lymph nodes are classified according to size, with nodes greater than 1 cm in diameter generally indicative of malignancy. This can be problematic, because benign lymph nodes may appear enlarged and micrometastases may appear normal on CT. Consequently, biopsy confirmation of the primary site and metastases is required to determine the most appropriate treatment.

More accurate noninvasive methods for staging NSCLC are needed to minimize the use of invasive procedures for diagnosis and monitoring treatment response. To this end, the metabolic information provided by a PET scan may be useful. Several roles for PET in staging lung cancer have been identified in the literature:

- Defining unknown primary disease;
- Detecting hilar and mediastinal metastases;
- Detecting distant metastases;
- Defining recurrence from fibrosis;
- Analyzing tumor biology;
- Monitoring response to therapy;

- Predicting tumor response by measuring uptake of chemotherapeutic agents.

Tables 11 and 12 depict study characteristics and Table 13 summarizes the data and quality of individual diagnostic accuracy studies of FDG-PET in NSCLC that met the inclusion criteria for this review. Scores were further refined with pluses and minuses to reflect the degree to which investigators minimized the effect of these biases on diagnostic accuracy results.

Defining unknown primary disease

Two studies met the inclusion criteria for the report. Guhlman (1997) and Hagberg (1997) are relatively small retrospective surgical series with a high prevalence of malignancy in their respective cohorts. Both evaluated PET in the test sequence after CT, but only Guhlman (1997) measured PET independently of other tests in all patients. Neither study presented data comparing PET to CT alone. Both studies received low methodologic quality grades due to incomplete reporting of methods and significant biases in study design, which may inflate estimates of diagnostic accuracy.

Detecting hilar/mediastinal adenopathy

Recent evidence on the use of PET in NSCLC emphasizes its staging potential. Six studies meeting the inclusion criteria presented evidence on the diagnostic accuracy of PET in nodal (N) staging and are listed in Table 13. All enrolled patients had suspected or biopsy-proven lung cancer. Data analyses included only biopsy-verified cases, implying a strong presence of work up bias across all studies. All studies assessed the role of PET independently of CT in the work up; Vansteenkiste (1997) also assessed PET as an adjunct to CT.

Guhlman (1997) and Hagberg (1997) were small retrospective studies with several methodologic flaws. The remaining four studies were reported as prospective evaluations of PET. Ambiguous descriptions of study methodology call into question the true, real-time prospective nature of three of them (Steinert, 1997; Vansteenkiste, 1997; Sasaki, 1996). Of these three, Sasaki (1996) was the most methodologically flawed.

Bury (1997) presented the largest and the only discernibly true prospective evaluation of PET in staging patients with NSCLC. Steinert (1997) and Vansteenkiste (1997) also presented notable attributes. These three studies represent the strongest evidence on the use of PET in N staging patients with NSCLC and are presented in Table 11 for comparison.

Table 11: Characteristics of Prospective Studies of Mediastinal Lymph Node (N) Staging With FDG-PET in Patients with Potentially Operable NSCLC

Note: All studies included mixed histologies, primarily squamous cell and adenocarcinoma.

Study Characteristics	*Bury et al. (1997)*	*Steinert et al. (1997)*	*Vansteenkiste et al. (1997)*
Patient source	141 consecutive patients who presented between 9/94-10/96 with new or suspected NSCLC based on sputum cytology, needle biopsy, or flexible bronchoscopy • 109 enrolled	62 surgical candidates with suspected or proven NSCLC who had PET between 2/94 and 3/96 • 47 enrolled	Unknown # patients who presented between 9/95-4/96 with suspected or confirmed NSCLC and who had standard M staging • 50 enrolled
Exclusion criteria (# patients)	• poor physiologic status (22) • poor compliance or no definitive diagnosis (11)	• prior neoadjuvant therapy • diabetes • inadequate CT (2) • distant metastases (8) • inadequate sampling (5)	• inoperable due to distant metastases • diabetes • treatment with oral corticosteroids • ischemic cardiomyopathy • direct mediastinal invasion of primary tumor • obvious bulky mediastinal adenopathies
Prevalence of confirmed N metastases (#N1-N3/# patients)	34/66=52%	29/47=62%	15/50=30%
Extent of N metastases (# patients)	N0=32 N1=20 N2=10 N3=4	N0=18 N1=16 N2=7 N3=6	N0=35 N2=15
Benign conditions	• nonspecific inflammation=2 • pneumonia=1 • multinodular goiter=1 • localized FDG uptake in hepatic-splenic angle of colon=1	none reported	none reported
PET criteria for positive node	• moderate uptake: > 2X uptake in contralateral or reference region • intense uptake: markedly higher than reference region	• FDG uptake ≥ FDG uptake in brain • nodular appearance	Grades 4 and 5 on a 5-point semiquantitative scale
Contrast CT criteria for positive node	short axis diameter > 10 mm	• short axis diameter > 10 mm except: • upper paratracheal nodes > 7mm short axis diameter • infracarinal station > 11 mm short axis diameter	maximal cross-sectional diameter ≥ 1.5 cm
Interpretation	• independent, blind • consensus by 2 radiologists and 2 nuclear medicine	• independent, blind • 1 radiology reader • 1 nuclear medicine reader	• independent, blind • one chest physician, one radiologist • 2 nuclear medicine readers
Gold standard determination (# patients)	• histology from mediastinoscopy (5), thoracotomy (51), both (10) • radiologic follow up based on CT or PET • all accessible nodes at surgery sampled	• extensive nodal sampling at thoracotomy of all identifiable nodes regardless of size on imaging • mediastinoscopy (22) and/or thoracotomy (18)	• nodal sampling at mediastinoscopy (47) and at thoracotomy (49), fine needle aspiration (1) • extent of sampling not reported
Data analysis	correlated by patient	correlated by nodal station	correlated by patient

Variations in study characteristics and units of analyses contributed to the range of reported estimates of diagnostic accuracy and differences in quality scores across studies. All studies had a significant degree of work up bias, which contributed to their low quality scores. All conducted varying degrees of nodal sampling, a means for minimizing diagnostic review bias, but the extent of sampling varied and was not reported with sufficient detail to enable the reader to quantify the effect of this bias on diagnostic accuracy. Bury (1997) and Vansteenkiste (1997) utilized multiple readers for blinded, independent image interpretation, but neither assessed interobserver variability.

Bury (1997) provided the strongest evidence to date on the diagnostic accuracy of PET in N staging NSCLC. A comparison of PET to CT yielded comparable accuracy estimates. The authors presented data on the impact of PET in modifying treatment, but no methods for systematic assessment were described. Bias in the stated methods and in incomplete reporting of other critical design elements hindered evaluation of study validity in the other studies. None of the studies assessed the incremental value of PET in the work up of NSCLC.

Detecting distant metastases

Studies in Table 12 met the inclusion criteria for review. Erasmus (1997) reported on 27 patients diagnosed with bronchogenic carcinoma and adrenal masses detected by CT. Adrenal masses are common in patients with NSCLC, but in the absence of other extrathoracic metastases, they are likely to be benign. Diagnosis of many adrenal masses remains indeterminate after standard anatomic imaging (CT or MRI), and a biopsy is required before treatment can be planned. The rationale for using PET in this case is to improve the noninvasive diagnostic accuracy, thus reducing the need for biopsy. Patients with normal FDG uptake in the adrenals and no evidence of distant metastases might be considered eligible for curative resection.

The findings suggest that, as an adjunct to CT, PET can discern malignant from benign adrenal masses using both visual and semiquantitative analyses. Results from this small preliminary study would need to be confirmed in larger, prospective studies to ascertain valid estimates of diagnostic accuracy and the added value of PET in diagnosing adrenal masses in these patients.

Bury (1997) present the strongest evidence to date on the use of PET for M staging NSCLC. They compared PET independently to conventional imaging (chest CT, abdominal CT, and bone scintigraphy) for M staging 109 patients with new or suspected NSCLC. The results suggest modest improvements in sensitivity and negative predictive value for PET over conventional imaging. The authors reported that PET correctly changed M stage, as determined by conventional imaging, in 14% of the cases and modified therapy in 20% of the patients, but the methods for assessing these changes were not described.

Table 12: Characteristics of Prospective Studies of Distant Metastases (M) Staging With FDG-PET in Patients with NSCLC

Study Characteristics	*Bury et al. (1997)*	*Erasmus et al. (1997)*
Patient source	141 consecutive patients with new or suspected NSCLC who had PET and conventional imaging between September 1994 and October 1996: • 109 patients enrolled in study • 39 patients with 59 sites of confirmed distant metastases	Unknown # consecutive cases presenting to thoracic surgery, oncology, or pulmonary between January 1993 and January 1996 with a diagnosis of bronchogenic carcinoma and an adrenal mass detected by CT • 27 patients with 33 adrenal masses enrolled in study
Exclusion criteria (# patients)	• Poor physiologic status (22) • Poor compliance or no definitive diagnosis (11)	• Inability to obtain informed consent • Poor clinical status • Death
Patient characteristics	• 77 men, 32 women • mean age= 64 yrs (44-83 yrs)	• 19 men, 8 women • mean age= 57 yrs. (39-76 yrs)
Characteristics of metastases (# patients)	• NSCLC (109) • Mean diameter not reported	• NSCLC (24); Small cell (3) • Bilateral masses (6) • Mean diameter=3 cm (1-9cm)
Prevalence of confirmed distant metastases	39 pts /109 pts=36%	23 sites /33 sites=70%
Locations of distant metastases (# sites)	• Adrenal glands(10) • Nonregional lymph nodes (6) • Lung (10); Bone (13); Liver (18) • Pleura (1); Soft tissue (1)	Adrenal glands (27)
Benign conditions (# sites)	• Nonspecific inflammation (2) • Pneumonia (1) • Multinodular goiter (1) • Localized FDG uptake in hepato-splenic angle of colon (1)	Not reported
PET criteria for positive metastases	• Moderate uptake: > 2X uptake in contralateral or reference region • intense uptake: markedly higher than reference region	Positive activity= activity > background
CT criteria for positive metastases	• Nodule characteristics not defined • Presence of clinical disease (symptomatic patient, progression on imaging, abnormal biochemistry) 6 months after imaging negative imaging	Visual detection of mass, characteristics not defined
Interpretation	• Independent, blinded to all data except histology of primary tumor • Consensus by 2 radiologists and 2 nuclear medicine	• Independent, blinded to clinical and biopsy findings • 3 readers
Gold standard determination (# patients or sites)	• Biopsy (21) • Clinical and radiologic follow up (88)	• Percutaneous needle biopsy (11) • Growth characteristics on sequential CT studies (16) • CT attenuation values < 10H (6)
Data analysis	Correlated by patient	correlated by site

Summary/Discussion

Early studies of PET suggested several potential uses for PET in managing NSCLC (Flynn, 1996). Positive trends in Medicare and private sector coverage policies for PET in lung cancer staging continue to fuel interest in the use of dedicated and camera-based PET as diagnostic tools. Since the first report, the TA Program identified 14 additional studies (7 of diagnostic accuracy) using dedicated PET, which met the inclusion criteria for this report. There were three areas in which potential uses for PET in NSCLC were studied: defining unknown primary disease, detecting nodal metastases, and detecting distant metastatic disease.

The best evidence on the diagnostic accuracy of PET in staging NSCLC suggests comparable accuracy of PET to CT in nodal staging and slightly better sensitivity, negative predictive value, and accuracy of PET over conventional imaging in staging distant metastases (Bury, 1997). Significant methodological biases, incomplete reporting of critical design elements, and variations in study characteristics (e.g., lack of uniform criteria for defining positive results on PET) limit the validity of the included studies and warranted low methodologic quality scores.

Appropriate use of the reference standard, or the "truth measure", is among the most challenging aspects of these studies to assess. Diagnostic review bias is often introduced, as biopsy sampling is rarely carried out independently of imaging results (e.g., it would be impractical to blind the surgeon to imaging). Bury (1997) minimized the effect of diagnostic review bias in nodal staging by conducting extensive nodal sampling and in distant staging by confirming disease status in all subjects using radiologic or clinical follow up or other confirmatory tests.

Imaging results are often used to determine which patients receive biopsy verification of mediastinal involvement (work up bias). To improve N staging accuracy several investigators advocated complementing the sensitivity of CT with the high negative predictive value of PET. They reasoned that a negative PET scan following a positive or indeterminate CT scan would exclude mediastinal metastases with a high degree of certainty and might obviate the need for invasive mediastinal evaluation (e.g., mediastinoscopy).

The best evidence for PET's N staging potential is confined to biopsy verified cases who had suspicious nodes on imaging. The size criteria for characterizing disease on CT and the lower detectable limit of resolution with PET may misclassify small tumor involvement, resulting in understaging. Failure to confirm disease status through follow-up in patients with negative CT or PET results may miss false negative results; failure to include the results in the analysis would result in inflated sensitivity and negative predictive values. Accurate, robust negative predictive values from studies that reduce the effect of work up bias are critical to determining the utility of PET in mediastinal staging.

Methodologically rigorous evaluations of diagnostic imaging, which reduced or accounted for the effects of methodologic biases on diagnostic accuracy, have been published (See Appendix II). In particular, Webb (1991) of the Radiologic Diagnostic Oncology Group (RDOG) provides an excellent model for evaluating diagnostic imaging in staging NSCLC. From patient enrollment to data analysis this rigorous evaluation offers extensive, detailed techniques for limiting the many biases inherent in diagnostic imaging studies. Incorporating study design elements from this model would strengthen the current best evidence for staging NSCLC using PET.

The value of diagnostic PET cannot be determined solely on improved accuracy over existing modalities. PET must demonstrate changes in diagnostic certainty and/or treatment planning or lower overall costs of patient management to justify its role in the work up. It can be argued that the metabolic information from PET may complement the information provided by conventional anatomic imaging and improve staging accuracy. More accurate staging may lead to more appropriate treatment planning. Studies included in this review reported anecdotal evidence of changes in treatment planning attributable to PET, but the impact of PET on treatment management was not systematically assessed, or reported as such. Furthermore, the range of stages and histologies of NSCLC and the associated range of treatments and outcomes would confound the effect of PET on outcomes of treatment, many of which are under investigation.

The TA Program concludes that the prevailing evidence does not support the routine use of either dedicated or camera-based PET in lung cancer staging. Data from rigorous, prospective clinical trials are needed to determine the added value of PET in the work up of NSCLC. Methodologically rigorous studies of diagnostic imaging have been published in the peer-reviewed literature. These studies may serve as models for guiding design of future PET research. Review of the more recent evidence confirms the conclusions from the first report.

Table 13: Summary of Diagnostic Accuracy Studies of PET and Alternatives in Staging Lung Cancer

H = histology; F = Follow up; S = small size; R = referral bias; W = work up bias; T = test review bias; D = diagnostic review bias (upper case indicates significant limitation; lower case indicates limitation minimized by study design, presence of bias unclear, or small effect on operating characteristics)

Role *(Some assessed multiple roles)*	***Study***	***N***	***Operating Characteristics***			***Evidence-based Medicine Criteria***			***Study Design Limitations***	***Methodologic Quality Grade***
			PET alone (95%CI)	***PET + CT***	***CT alone (95%CI)***	***comparison group***	***gold standard***	***blinding***		
Defining unknown primary disease	Guhlman 1997	32 malignant cases 14 benign cases	Se=94% Sp=86% Acc=91%		Not reported	+	H	+	**S,r,w,d**	**C**
	Hagberg 1997	44 positive nodules 10 neg nodules (in 49 patients)		Se=93% Sp=70%	Not reported	+	H	—	**S,r,w,T,d**	**D**
Detecting mediastinal/hilar adenopathy	Bury 1997	34 positive cases 32 negative cases	Se=89% (72-96%) Sp=87% (71-97%) PPV=89% (72-96%) NPV=87% (71-96%) Acc=88%		Se=79% Sp=71% PPV=75% NPV=76% Acc=75%	+	H	+	**s,r,W,t,d**	**D+**
	Guhlman 1997	20 positive cases 12 negative cases	Se=80% (56-94%) Sp=100% (73-100%) Acc=87% (71-96%) (p<.02)		Se=50% (27-73%) Sp=75% (43-95%) Acc=59% (41-76%)	+	H	+	**S,R,W,d**	**D-**
	Hagberg 1997	9 positive nodes 9 negative nodes (in 18 patients with N2 disease only)	Se=67% Sp=100%		Se=56% Sp=100%	+	H	+	**S,r,W,d**	**D-**
	Steinert 1997	28 positive nodal stations 84 negative nodal stations (in 47 patients)	Se=89% (P=0.0066) Sp=99% PPV=96% NPV=97% Acc=97%		Se=57% Sp=94% PPV=76% NPV=87% Acc=85%	+	H	+	**S,r,W,d**	**D**
	Vansteenkiste 1997	15 positive cases 35 negative cases	Se=67% Sp=97% PPV=91% NPV=87% Acc=88%	Se=93% Sp=97% PPV=93% NPV=97% Acc=96%	Se=67% Sp=63% PPV=43% NPV=81% Acc=64%	+	H	+	**S,r,W,t,D**	**D**
	Sasaki 1996	17 positive regions 54 negative regions (in unknown # patients)	Se=76% Sp=98% (P<0.05) PPV=93% NPV=93% Acc=93% (P<0.05)		Se=65% Sp=87% (P<0.05) PPV=61% NPV=89% Acc=82% (P<0.05)	+	H	—	**S,R,W,T,d**	**D-**

Table 13 (cont.): Summary of Diagnostic Accuracy Studies of PET and Alternatives in Staging Lung Cancer

H = histology; F = Follow up; S = small size; R = referral bias; W = work up bias; T = test review bias; D = diagnostic review bias (upper case indicates significant limitation; lower case indicates limitation minimized by study design, presence of bias unclear, or small effect on operating characteristics)

Role *(Some assessed multiple roles)*	***Study***	***N***	***Operating Characteristics***			***Evidence-based Medicine Criteria***			***Study Design Limitations***	***Methodologic Quality Grade***
			PET alone (95%CI)	***PET + CT***	***CT alone (95%CI)***	***comparison group***	***gold standard***	***blinding***		
Detecting distant metastases	Bury 1997	39 positive cases 70 negative cases	Se=100% (91-100%) Sp=94% (86-98%) PPV=90% (78-97%) NPV=100% (95-100%) Acc=96% (90-98%)		(conventional imaging*) Se=82% Sp=89% PPV=80% NPV=89% Acc=86%	+	H + F	+	**r,t,**	**B/C**
	Erasmus 1997	23 malignant lesions 10 benign lesions	Se=100% Sp=80%			+ internal	histology + CT follow up	+	**S,r,d**	**C**

* Bury et al 1997 defined conventional imaging as chest CT, abdominal CT, and bone scintigraphy

D. Solitary Pulmonary Nodules

Background information on solitary pulmonary nodules (SPN) is supplied by Lillington and Caskey (1993). A SPN is a single spherical lesion within the lung not associated with hilar enlargement or atelectasis and with a diameter generally less than 4.0 cm. The American Cancer Society reports that SPNs represent approximately 15% of all lung cancer diagnosed and estimates 25,725 new cases of malignant SPNs in the United States in 1998.

The differential diagnoses of a SPN include many malignant and benign processes. The most common malignant forms are bronchogenic carcinomas. Reported prevalence of malignant SPNs range from less than 5% to greater than 70% because of differences in the spectrum and severity of disease within each reported patient series. A malignant SPN represents a clinical stage I lesion, which is potentially curable with resection. Infectious granulomas represent the majority of benign processes and are caused predominately by coccidiomycosis, histoplasmosis, and tuberculosis.

The following risk factors directly correlate with the probability of cancer in patients with a SPN: 1) patient's age; 2) smoking history; 3) antecedent malignancy; 4) stability of lesion size on chest x-ray for 2 years; 5) absence of benign patterns of calcification within the nodule; and 6) nodule morphology (size and edge characteristics on CT). The baseline prevalence of malignancy in the study population may suggest the likelihood of a malignant SPN. Exposure to benign diseases such as tuberculosis or a history of residence in areas endemic for coccidiomycosis or histoplasmosis will suggest a lesser likelihood, but not rule out, malignancy.

Following clinical exam and chest radiography, the standard radiologic method of choice for evaluating SPNs is CT. CT provides information on the location and morphology of the nodule and can be used to guide biopsy procedures. Iodinated contrast material and high resolution CT densitometry may be used to enhance the differential diagnosis. However, limitations in the use of CT have been reported. Many SPNs are classified as "indeterminate" after CT and warrant invasive biopsy confirmation to determine the appropriate therapeutic course.

FDG PET has been proposed as a potential solution for improving the noninvasive differential diagnosis of SPNs, thereby reducing the need for higher risk invasive biopsy sampling and the associated morbidity and costs. Current evidence from this review supports the complementary use of PET after CT in the work up of patients with nodule diameters less than 3 cm or 4 cm, i.e., those nodules most likely to be indeterminate.

Table 14 displays the attributes of each study to highlight the variations in study quality and in criteria relevant to the applicability of the results. Table 15 summarizes the data and quality of individual diagnostic accuracy studies of FDG PET in SPNs.

Characterizing indeterminate solitary pulmonary nodules

Two studies met the inclusion criteria for this report. Dewan (1997) conducted a retrospective single-site study of indeterminate SPNs in 52 consecutive patients, who underwent PET between April 1990 and February 1994. They compared PET with and without standard criteria (clinical and radiologic data) using likelihood ratios[1] in Bayesian analysis to predict the probability of cancer in a SPN. Using sensitivity and specificity derived from this patient group, the authors determined that PET alone was the best predictor of cancer.

However, biases in study design and violation of the assumption of conditional independence between tests in the testing sequence, a requirement of Bayesian analysis, preclude drawing definitive conclusions regarding the accuracy of PET and its contribution to diagnostic certainty in these patients. Moreover, the impact of PET on treatment planning was not assessed. It is also important to note that many of these patients may have been included in studies assessed in the 1996 report.

Lowe (1998) conducted a multi-site study of radiologically indeterminate SPNs in 105 consecutive patients, who underwent imaging between October 1993 and August 1994. The study population included a broader range of benign conditions and nodule sizes compared with other published studies for this indication, reflecting the advantages of multi-site design. The authors presented a very detailed description of their blinding procedures and were the only investigators to calculate interobserver variability in visual analysis. From the stated methods, it is unclear whether they collected patient data in a "real-time" prospective fashion or retrospectively from surgical series.

These authors calculated likelihood ratios overall and for each subgroup. The likelihood of cancer was consistently higher using quantitative analysis over visual analysis. Except for specificity in SPNs ≤ 3cm in diameter, there were no significant differences between visual and quantitative analyses in the other diagnostic accuracy measures across subgroups. Small sample sizes in the subgroups likely contributed to the failure to detect any significant differences. Interobserver variability was very low (kappa=0.95), indicating good reproducibility of image interpretation.

[1] Likelihood ratio, expressed as Sensitivity/1-Specificity, is a measure of accuracy that indicates by how much a diagnostic test result will raise or lower the pretest probability of disease, thereby increasing the certainty about a positive or negative diagnosis.

Table 14: Characteristics of Studies Using FDG-PET of Patients with Radiographically Indeterminate Solitary Pulmonary Nodules

Study Characteristics	*Dewan et al. (1997)*	*Lowe et al. (1998)*
Perspective	Retrospective	Prospective (?not real-time)
Patient source	52 consecutive patients who underwent PET between April 1990 and February 1994 • included 3 with extrathoracic malignancy	Multisite study of 89 of 105 consecutive patients who underwent imaging between October 1993 and August 1994
Exclusion criteria (# patients)	• Cavitary or calcified nodules • Nodule size > 3cm • Age ≤ 30 years • # patients not reported	• no definitive histologic confirmation (8) • 4 not classified as radiographically indeterminate SPN (4) • no available CT scans (2) • nodule size < 0.7cm or > 4.0cm on CT(? # pts.)
Patient demographics	• 43 men (83%) • mean age ± SD=63.6±11.3 years • 41(79%) current smokers • 52% ≥ 20 cigs/day	• 61 men (69%) • mean age ± SD=63±9.5 years • smoking status not reported
Prevalence of malignancy	37/52=71%	60/89=67%
Nodule size in cm (%malig. pts. vs. %benign pts.)	≤ 1.0= 19% vs. 47% 1.1-2.0=51% vs. 40% 2.1-3.0=30% vs. 13%	0.7-1.5= 25% vs. 66% 1.6-3.0=60% vs. 24% 3.1-4.0=15% vs. 10%
Nodule Morphology (%malig. pts vs. % benign pts.)	Edge characteristics reported: • Sharp, smooth=14% vs.20% • Lobulated=30% vs. 40% • Slightly irregular w/ few spiculations=38% vs. 33% • Grossly irregular and spiculated=19% vs. 7%	Not reported
Benign conditions (#pts.)	• histoplasma granuloma with active inflammation (2) • other conditions not reported	• granuloma (7), coccidiomycosis (4), benign cellular debris (4), nonspecific inflammation (3), necrotizing granuloma (3) • fibrosis (1), hemangioma (1), aspergillosis (1), metaplasia (1)
PET criteria for positive node	focal FDG uptake > surrounding lung tissue, but more than mild intensity	• focal uptake > mediastinal blood pool structures (qualitative) • SUV> 2.5 (semiquantitative)
CT criteria for nodule edge	based on 4-type scale to reflect degree of spiculation and irregularity	not specified to image interpreters
Interpretation of PET	• qualitative • 1 reader blinded to histology • blinding to clinical and radiologic information varied	• semiquantitative using SUV • independent qualitative analysis using 2 readers blinded to clinical , imaging, and histopathologic data reached by consensus • readers interpreted studies with which they were not involved to ensure blinding • interobserver variability calculated
Interpretation of CT	• independent • 2 readers blinded to clinical diagnosis • consensus reading	• independent interpretation by > 1 reader blinded to clinical, PET, or histopathologic results • qualitative interpretation as benign or indeterminate
Gold standard determination (# patients)	thoracotomy (36), mediastinoscopy (3), bronchoscopy (3), needle lung biopsy (9), follow-up imaging for > 2 yrs (1)	TTNA (29) or surgery (60)
Data analysis	By patient	By patient

TTNA=Transthoracic Needle Aspiration

Summary/discussion

Since the 1996 report, three additional studies using dedicated PET in diagnosing solitary pulmonary nodules met the inclusion criteria for review. One was a technical feasibility study, and two were of diagnostic accuracy assessing PET in the test sequence after CT but prior to any histologic confirmation of disease. Both had significant biases in study design that warranted low methodologic quality scores and call for caution in generalizing these results to other populations.

Most false negative results reported in the PET literature are caused by small nodules with diameters commonly <1 cm that approach the resolution limits of the camera. Both studies reported false negatives comprising a variety of non-small cell cancers with diameters ranging from 1 cm to 2.5 cm. Moreover, the impact of PET on treatment planning, particularly the decision to proceed to surgery, was not systematically assessed.

One of the deficiencies outlined in the first report is the relatively low number of patients and a correspondingly narrow spectrum of benign conditions represented in the study base. Lowe (1998) presented the largest and only multi-site study of PET in diagnosing SPNs. Multi-site trials have the advantage of recruiting larger numbers of patients with a comprehensive array of malignant and benign conditions that are needed to apply the results to other populations. The detailed description of the blinding procedures used in the study may serve as a model for future studies of PET.

Both studies derived likelihood ratios (LR) to quantify the importance of the PET results in the work up of SPNs. As with predictive values, LRs are more useful accuracy measures to a clinician than sensitivity and specificity. LRs are used to calculate the probability of disease given a test result. They are independent of disease prevalence in most circumstances, but differences in case mix and methodologic biases can influence their validity (Gurney, 1993).

For example, the prevalence of malignancy in SPNs is lower in community hospitals than in most surgical series or in tertiary care facilities, where most PET scanners are found. Areas that experience a higher prevalence of particular benign conditions may encounter more false positive results on PET. A study with too few patients with benign nodules may overestimate specificity and inflate the negative LR; presence of methodologic biases may overestimate sensitivity and inflate the positive LR. In both studies the inclusion criteria favored a higher proportion of patients with malignancies and with too few benign conditions to offset the influence on specificity. Thus, rigorous study of a larger number and range of patients with a mix of diseases is needed to derive valid likelihood ratios for PET in patients with SPNs.

Table 15: Summary of the Diagnostic Accuracy and Diagnostic Thinking Efficacy Studies of PET in Indeterminate Solitary Pulmonary Nodules (SPN)

H = histology; F = Follow up; S = small size; R = referral bias; W = work up bias; T = test review bias; D = diagnostic review bias (upper case indicates significant limitation; lower case indicates limitation minimized by study design, presence of bias unclear, or small effect on operating characteristics)

Role	*Study*	*N*	*Operating Characteristics*		*Evidence-based Medicine Criteria*			*Study Design Limitations*	*Methodologic Quality Grade*
			PET after CT Visual analysis (95%CI)	*PET after CT Semiquantitative analysis (95%CI)*	*comparison group*	*gold standard*	*blinding*		
Defining indeterminate SPN	Dewan 1997	32 malignant cases 14 benign cases	Se=95% Sp=87% Acc=92% LR_{mal}=7.11 (6.36-7.96) LR_{ben}=0.06 (0.05-0.07) *Overall values reported*		+	H + F	varied	**S,R,W,T,d**	**D**
	Lowe 1998	60 malignant cases 29 benign cases	Se=98% (95-100%) Sp=69% (57-81%) Acc=89% LR_{mal}=3.0 LR_{ben}=0.02 *Overall values reported*	Se=92% (82-100%) Sp=90% (79-100%) Acc=91% LR_{mal}=9.0 LR_{ben}=0.09 *Overall values reported*	+	H	+ (for visual analysis only)	**s,r,W,d**	**D**

Once valid LRs are derived, they may be used to estimate the odds that a patient has a cancer, given the PET result. Any attempt to use LRs in evaluating the odds of cancer after PET requires: 1) knowledge of the odds of cancer before PET, and 2) that the PET results were derived independent of the other test results. In neither study were both conditions satisfied, and the influence of PET on diagnostic certainty and subsequent treatment planning could not be determined.

Rigorous studies of patients comprising a range of pre-PET probabilities of malignancies are needed to assess the diagnostic accuracy and contribution of either dedicated or camera-based PET to the work up of solitary pulmonary nodules. Multiple sites may be needed to accrue a sufficient number and array of patients. Results from this review update confirm the conclusions and recommendations from the first report.

The Cooperative Studies Program of the VHA Office of Research and Development has funded a multi-year cooperative study to determine the efficacy of FDG-PET in defining solitary pulmonary nodules (See Section VIII). Results from this study should address the shortcomings of the existing literature.

E. Colorectal Cancer

Colorectal cancer is the third leading cause of death among men and women, representing a significant public health problem in the United States. Colorectal cancers account for approximately 11% of new cancer diagnoses. Death rates from colorectal cancer have fallen 25% for women and 13% for men during the past 20 years, reflecting a decreasing incidence of new cancer cases and increasing survival rates.

An estimated 131,600 cases and 56,500 deaths are attributable to colorectal cancer in the United States in 1998. An estimated 1 million veterans over the age of 50 will develop colorectal cancer over the remainder of their lives and nearly 433,000 will die from it (Wingo, 1995; Brown, 1996). Within the Veterans Health Administration, malignant neoplasms of the digestive organs and peritoneum (which include colorectal cancer) accounted for 8,280 discharges (1.2% of all discharges) with an average length of stay of 15.7 days in FY 1997.

Winawer (1997) reported the following risk factors for colorectal cancer: age over 50 years; a history of adenomatous polyps; a history of curative intent resection of colorectal cancer; inflammatory bowel disease; and familial colorectal cancer, adenomatous polyposis, or hereditary nonpolyposis colorectal cancer.

Nationally, the estimated relative five-year survival rate among veterans is approximately 40%, substantially lower than estimates from the general population of 62% (colon) and 59% (rectum). In VA, the Office of Research and Development (ORD)'s Epidemiologic Research and Information Center in Durham, North Carolina is conducting a four-year initiative to identify factors that may explain the worsened prognosis among veterans, and that may be responsive to intervention (Provenzale, 1998). ORD is also conducting a large prospective study of risk factors and/or detection of altered cell proliferation for

large colonic adenomas in asymptomatic subjects; the results will have important implications for colon cancer screening (Lieberman, 1998).

Data on management of colorectal cancer are from the National Cancer Institute's Physician Desk Query (PDQ) system retrieved in October 1998. The most prevalent histologic type of colorectal cancer is adenocarcinoma. Metastases to the liver, abdominal cavity, and extra-abdominal areas at initial diagnosis are common, as is recurrent disease after surgical resection of the primary tumor. Prognosis and management depends on the depth of tumor penetration into the bowel wall and the presence of both regional lymph node involvement and distant metastases (staging).

Surgery is the primary therapy for colorectal cancer, and for cancers that have not metastasized, it is frequently curative. Many patients with confined recurrent disease or with metastases limited to the liver or lungs may also be amenable to resection. However, the high rate of recurrence and a troubling overall five-year survival rate call for more appropriate selection of patients who may benefit from surgical resection. The morbidity and costs associated with surgery for patients who do not have genuinely resectable recurrent tumor could be avoided by improved methods of tumor detection.

Stotland (1997) reviewed several imaging modalities commonly used to stage and diagnose colorectal cancer. The most common modalities include CT, MRI, endoscopic ultrasonography (EUS), and transabdominal ultrasonography. The popularity of EUS, in particular, has grown in recent years for its ability to image the depth of tumor penetration into the bowel wall and regional lymph node involvement. MR endorectal coils or ultrasound probes may be used to image rectal lesions. However, all structural imaging modalities are circumscribed in their ability to determine the presence and extent of disease and disease recurrence. Information from newer modalities, such as intraoperative ultrasonography, immunoscintigraphy, arterioportography, and PET, may increase the accuracy of staging and detecting recurrence.

Potential roles for PET in colorectal management have been identified in the literature:

- Pre-operative staging, including diagnosing presence and extent of liver metastases, and;
- Post-operative monitoring of recurrent disease.

Five studies met the inclusion criteria for review. Of these, two were technical efficacy studies and are listed in the reference section. Table 16 lists the characteristics of two retrospective case series and one prospective case series of diagnostic accuracy, and Table 17 summarizes the data and quality, representing the best evidence for the use of PET in managing patients with colorectal cancer. All studies presented some anecdotal evidence of therapeutic efficacy.

Preoperative staging of colorectal cancer

The TA Program identified one small uncontrolled, unblinded technical feasibility study of PET for staging initial *primary* colorectal cancer (Abdel-Nabi, 1998).

No diagnostic efficacy studies of staging primary colorectal carcinomas using PET were identified for review.

Four relatively small case series presented evidence on the use of PET in patients with suspected *recurrent* colorectal cancer, of which Ruhlmann (1997) was a retrospective technical feasibility study. The three remaining case series are diagnostic accuracy studies. Ogunbiyi (1997) and Flanagan (1998) are retrospective analyses from the same institution with overlapping study populations. Ogunbiyi (1997) studied 58 patients with a high suspicion for recurrence, including some with advanced primary disease, based on clinical symptoms, elevated plasma carcinoembryonic antigen (CEA) concentration, and/or CT findings. Flanagan (1998) assessed the ability of PET to detect recurrence in 22 asymptomatic patients with a post-operative elevated CEA concentration and normal clinical and radiologic findings.

Delbeke (1997) presented the only prospective comparison of PET to CT and CT arterial portography (CTAP) in detecting liver and extrahepatic metastases in 52 patients with suspected recurrent colorectal cancer. This is likely a continuation of an earlier, smaller study from the same institution (Vitola, 1996), which was reviewed in the previous 1996 MDRC technology assessment.

In all studies PET was performed as an adjunct to the routine clinical and radiologic work up, but the initial work up was not described in detail. Current evidence suggests that, when PET is added to the work up, there is improved sensitivity in distinguishing recurrence from post-surgical changes and documenting the presence and extent of liver and more distant metastases. However, the methodologic shortcomings in these studies limit the validity of these estimates. Predictive values may be subject to considerable referral bias owing to the high suspicion for malignancy in the study population. Lack of documentation of disease severity and underlying condition of the liver, completeness of the work up prior to PET, and blinding further hinders assessment of these results.

Table 16: Characteristics of Studies of Pre-operative Staging With FDG-PET in Patients with Suspected Recurrent Colorectal Cancer

Study Characteristics	*Delbeke et al. (1997)*	*Ogunbiyi et al. (1997)*	*Flanagan et al. (1998)*
Perspective	Prospective	Retrospective	Retrospective
Patient source	52 patients presented on 61 occasions with suspected recurrent carcinoma • Consecutive series	58 patients who had PET between 1/91 and 1/95 with suspected recurrent (n=47) or advanced primary (n=11) disease • ? Consecutive series	22 of 128 patients with history of colorectal cancer, who underwent PET from 6/93 to 6/96 • ? Consecutive series
Inclusion criteria	• Elevated CEA levels or abnormal CT • Abdominal CT (n=48); CTAP (n=40); or both (n=29)	• High clinical suspicion and equivocal or positive CT findings (n=39) • Elevated CEA levels with normal CT (n=19)	• Normal CEA levels after initial resection • Plasma CEA level > 5.0 ng/ml (mean 25 ng/ml), normal imaging studies, endoscopy, and physical exam on routine follow-up
Patient characteristics	• 31 men, 21 women • Mean age 63 ± 11 yrs	• 33 men, 25 women • Mean age 60 yrs. (23-81 yrs)	• 17 men, 5 women • Ages 17-84 • Primary site: colon (9), rectum (10), rectosigmoid (2), appendix (1)
Extent of disease (#patients)	• Liver metastases (45) • Extrahepatic disease (26, including 16 with liver mets)	• Primary disease or local recurrence (21) • Liver metastases (23) • Extrahepatic metastases (20)	• Stage B (10) • Stages C (5), C1 (2), C2 (3), • Stage D (2)
Benign conditions (# patients)	• Normal liver (7) • Post-surgical site (8) • Local fibrosis (2) • Resolving abscess (1), hepatic cyst (1), hematoma (1)	Not reported in reproducible detail	Not reported
PET criteria for positive site	• Not specified for qualitative PET • Cut-off not specified for semiquantitative analysis	• Malignancy=FDG uptake moderately or markedly intense; • Benign=no or mild uptake, or if abnormality identified on other imaging for which no corresponding abnormality was present on PET	Not specified
Contrast CT criteria for positive site	• Not specified for surgical cases • In nonsurgical cases, an increase in lesion volume > 20% on serial scans	Not specified	Not specified
CTAP criteria for positive site	• Not specified for surgical cases • In nonsurgical cases, an increase in lesion volume > 20% on serial scans	N/A	N/A
Interpretation	• 2 readers for PET , 2 readers for CT and CTAP, • Independent, qualitative PET blinded to other imaging results • Semiquantitative PET SUR calculations excluded lesions < 1 cm in diameter	• Qualitative PET interpreted with access to CT results • Two readers • CT interpreted in “routine clinical fashion”	• Qualitative PET scans interpreted with access to CT results • Consensus of at least two readers • CT interpreted in “routine clinical fashion”
Gold standard determination (# patients)	• Clinical or radiologic follow up (n=17) • Histopathology obtained surgically (n=44) • Percutaneous fine needle aspiration (n=2) • Nonresected lesions =surgical exam and intraoperative ultrasound (unknown #)	• Surgery, histology, or both (n=40); • Clinical and radiologic follow up (n=16); autopsy reports(n=2) • All patients followed for at least 12 months after PET or until death	• Pathology (n= 9) • All patients had radiologic and clinical follow up ≥ 6 months
Data analysis	By lesion site	By patient	By patient

Each study presented some evidence on changes in patient management attributable to PET, but the methods for assessment were not reported. The evidence suggests that adding PET to the work up may help optimize treatment (e.g., improve patient selection for curative surgery) by documenting the presence or absence of hepatic or more distant metastases. These data would need to be confirmed in much larger prospective studies designed to systematically assess the incremental value of PET against the many other available imaging modalities used in the work up of colorectal cancer.

Postoperative monitoring recurrent disease

The TA Program did not identify any studies in the published literature that addressed the role of PET in routine postoperative monitoring of patients for recurrent disease.

Summary/Discussion

Since the first report, five additional studies using dedicated PET in the management of colorectal cancer met the inclusion criteria for review. The best evidence to support the use of PET in colorectal cancers are three reported case series of diagnostic accuracy, of which two were retrospective studies from the same institution with overlapping study populations. All assessed the ability of PET as an adjunct to CT and other diagnostic tests to stage potentially operable patients with a high suspicion of recurrent disease; the one prospective case series also included patients with advanced primary disease. No diagnostic accuracy studies of PET to stage early, primary disease were identified.

Current evidence suggests that to further define recurrent disease, PET added after CT may offer improved sensitivity over CT alone. The absolute sensitivity of imaging modalities in detecting hepatic and more distant metastases is difficult to determine (Stark, 1987). ***Work-up bias*** is present when results from PET and/or other imaging tests under evaluation are used to direct biopsies to confirm suspicious liver lesions or to direct the choice of the most appropriate reference measure. Biopsy resection, while not entirely perfect, is a very accurate reference measure.

All authors attempted to offset work up bias by confirming disease in unresected patients using less perfect truth measures, such as clinical and radiologic follow-up, surgical exam and palpation, and intraoperative ultrasound. Although using these truth measures may not adequately identify the number of false negatives, they are reasonable alternatives and are preferred over nothing. The extent to which work up bias can be eliminated in this clinical setting is limited.

All of these studies had significant methodologic biases and insufficient reporting of fundamental design elements that preclude definitive assessment of study validity. The accuracy estimates from these studies should be interpreted with caution.

Table 17: Summary of Diagnostic Accuracy Studies of FDG-PET in Colorectal Cancer

H = histology; F = Follow-up; S = small size; R = referral bias; W = work up bias; T = test review bias; D = diagnostic review bias (upper case indicates significant limitation; lower case indicates limitation minimized by study design, presence of bias unclear, or small effect on operating characteristics)

Role	*Study*	*N*	*Operating Characteristics*			*Evidence-based Medicine Criteria*			*Study Design Limitations*	*Methodologic Quality Grade*
			PET after CT	*CTAP*	*CT alone*	*comparison group*	*gold standard*	*blinding*		
Diagnosing local recurrence	Ogunbiyi (1997)	21 recurrent cases 26 no recurrence	Se=90% (P=0.008) Sp=100% PPV=100% NPV=93% Acc=96%		Se=57% Sp=81% PPV=71% NPV=70% Acc=70%	+	Surg, H & F	—	**S,R,w,T,d**	**D**
	Flanagan (1998)	15 recurrent cases 7 no recurrences	Se=100% Sp=71% PPV=89% NPV=100%			+	H & F	—	**S,r,T,D**	**D**
Detecting liver metastases	Delbeke (1997)	104 malignant lesions 23 benign lesions in 45 patients	Se=91% Sp=95% Acc=92%	Se=97% Sp=5% Acc=80%	Se=81% Sp=60% Acc=78%	+	Surg, intra-operative US, H & F	partial	**S,r,W,t,d**	**D**
	Ogunbiyi (1997)	23 cases with disease 35 no disease	Se=96% (P=0.02) Sp=100% PPV=100% NPV=97% Acc=98%		Se=74% Sp=86% PPV=77% NPV=83% Acc=81%	+	Surg, H, F & autopsy	—	**S,R,w,T,d**	**D**
Detecting extrahepatic metastases	Delbeke et al. (1997)	34 malignant lesions 5 benign lesions in 26 patients	Se=100%		Se=74%	+	H & F	partial	**S,r,w,t,d**	**D**

CTAP=CT with arterial portography

All discussed changes in therapeutic management attributable to PET, but the methods for evaluation, details of the work up, or documentation of disease severity among the cases were not described. To suggest that PET improves the pre-operative staging process for selecting more appropriate patients for resection based on the existing evidence is ill-advised.

The TA Program did not identify any studies evaluating the efficacy of PET in post-operative monitoring. There is no consensus on the benefit of routine intensive follow-up after primary treatment, and the timing, frequency, type, and indications for post-operative follow-up using imaging are not standardized (Stotland, 1997). Any evaluation of PET in this role would be in the context of uncertain benefits of such monitoring.

Appendix II lists two particularly relevant studies for staging colorectal cancer and could serve as models for future PET research. Notable design features are highlighted. Zerhouni (1996) of the Radiology Diagnostic Oncology Group conducted a large, multi-site trial to compare the relative accuracies of CT and MRI in staging primary colorectal cancer. Stark (1987) compared CT and MRI to detect liver metastases, an important aspect of staging colorectal cancer patients. Studies of PET that incorporate these features with the comparable level of detail would provide more robust data on which to more confidently judge the added value of PET in the work up of colorectal cancer.

The TA Program concludes that the prevailing evidence does not support the routine use of either dedicated or camera-based PET in the management of colorectal cancer. Larger, prospective studies of diagnostic accuracy and subsequent therapeutic efficacy of PET in the work up are needed. Methodologically rigorous studies of diagnostic imaging have been published that may serve as models for guiding design of future PET research. Review of the recent evidence confirms the conclusions from the first report.

F. Alzheimer's Disease

This section briefly summarizes Alzheimer's disease (AD) and presents updated epidemiological information and results of a systematic review of the literature evaluating PET using FDG as a diagnostic test in AD. Appendix 8 of the MDRC technology assessment report on PET (Flynn, 1996) provides an expanded discussion of the disease, diagnosis, treatment, methodological and ethical considerations, and alternative neuroimaging technologies and other relevant diagnostic tests used in AD.

Unless otherwise noted, epidemiological information is from a consensus statement of the American Association for Geriatric Psychiatry, the Alzheimer's Association, and the American Geriatrics Society (Small, 1997). AD, a progressive neurodegenerative disorder, is the most common form of dementia and affects an estimated 4 million people in the United States. AD is characterized by steady irreversible decline in cognition, functioning, and behavior with sparing of motor and sensory functions until later stages. The rate of progression is variable, but duration of illness from diagnosis to death is approximately 10 years.

The reported prevalence of AD is approximately 6-8% of all persons 65 years or older. It doubles every 5 years after the age of 60 years, so that about 30% of the population older than 85 years will have AD. By the next century, an estimated 600,000 veterans with severe dementia will require long-term institutional care (ORD Impacts, 1997). The direct and indirect costs for care of AD patients in the United States approach $100 billion annually. The true costs of AD to society is likely much more, as economic assessments frequently underestimate the economic and emotional burden imposed on the caregivers as well as the patients.

Hendrie (1998) recently summarized the achievements in understanding genetic and nongenetic risk factors associated with AD. Genetic risk factors account for about 2% of all AD cases. Both causative (mutations on chromosomes 1, 12, 14, and 21) and associative genes (APOE-*4* allele[2] on chromosome 19) for AD have been identified. In VA, ORD researchers are: 1) studying genetic and environmental factors that contribute to delayed onset of AD in subjects with chromosome 1 mutations (ORD, 1997), and 2) are following subjects with the APOE-*4* allele at higher risk for developing AD to better detect and characterize early stages of this disease (Bondi, 1997).

Diagnostic tests that detect the presence of the APOE-*4* allele for apolipoprotein E, a serum lipoprotein involved in cholesterol transport, are under investigation, but experts differ on its usefulness. Since the APOE-*4* allele is found in many elderly persons without AD and is not always found in patients with AD, the Working Group of the American Medical Genetics/American Society of Human Genetics concluded that predictive testing of APOE-*4* for AD should not be done.

The only nongenetic risk factors consistently associated with risk for AD are age and family history. Other possible risk factors with a predominately positive association include low education, depression, estrogen-replacement therapy, nonsteroidal anti-inflammatory drugs (NSAIDs). Female gender, head injury, hypothyroidism and, to a lesser extent, insulin-dependent diabetes, aluminum exposure and smoking are inconsistently associated with an increased risk for AD. Clinical trials examining the role of estrogen, NSAIDs, and vitamin E in AD are reportedly underway.

The primary role of diagnostic testing is the differential diagnosis of AD from other reversible or treatable dementias. A definitive diagnosis is based on a typical clinical picture and histopathologic sampling of brain tissue at autopsy. In the absence of histologic confirmation, patients with probable AD are often referred to as having dementia of the Alzheimer's type (DAT). Two distinct sets of antemortem clinical criteria from the following may be used to characterize patients with DAT:

- (NINCDS/ADRDA)--National Institute of Neurologic and Communication Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association
- (DSM-IIIR or the more recent DSM-IV)--Diagnostic and Statistical Manual for Mental Disorders, American Psychiatric Association.

[2] In Mendelian genetics, an allele is any alternative form of a gene at a given locus. An allele may express a dominant, a recessive, or an intermediate trait.

While advanced stage AD is usually easier to diagnose, early stage disease can be problematic. There is no cure for AD, but psychosocial techniques for behavioral problems associated with dementia and drug therapies for cognitive impairment have been developed, which can improve quality of life. HSR&D researchers found that two approaches improve quality of care and reduce costs associated with caring for AD patients: 1) simulated presence therapy, which uses selected memories through tape recorded conversations to manage problem behaviors in AD patients (Camberg, 1999); and 2) hospice care for managing AD patients with advanced dementia (Volicer, 1994).

New therapy aimed at slowing disease progression is also available. Since it is most effective if given at the earliest stages of AD, there is a need for obtaining earlier and more accurate antemortem diagnoses. Such information would also help patients and their families better prepare for future challenges. Functional imaging technologies such as PET and SPECT have been used to improve diagnostic certainty and to provide information on the pathophysiologic basis of AD.

Eight studies of technical efficacy using only dedicated PET scanners met the inclusion criteria for review. The TA Program was unable to identify published PET studies at higher levels of the Fryback and Thornbury diagnostic efficacy hierarchy. The following table summarizes information from these studies. All studies used FDG-PET to study regional cerebral glucose metabolic rates; Ishii (1997) also measured cerebellar glucose metabolic rates.

Evidence from recent technical efficacy studies shows a growing interest in the use of PET to better understand the biological mechanisms of neurodegenerative disease. The research suggests a link between cognitive function, functional imaging data, and the neurobiology of dementia. There is also increasing emphasis in these studies on improving methods for detecting early stage AD by improving the measurement of regional brain function. More precisely defined neuroanatomical atlases and methods of analysis may help explain the underlying pathophysiology of AD and the differences between diseases and disease progression.

Results from Imamura (1997) and Vander Borght (1997) underscore the limitations in existing knowledge using PET to diagnose AD. That is, while the temporal and parietal metabolic patterns often differentiate AD from other causes of dementia, AD also shares functional imaging features with other causes.

Table 18: Summary of Recent Technical Efficacy Studies Using FDG PET in Alzheimer's Disease

Study	*Objective*	*Findings suggest...*
Desgranges et al. (1998) N = 19	To study the neuronal basis for memory impairment in AD using Tulving's hierarchical model of memory systems and PET measurement of resting regional cerebral glucose utilization	• Their methodology for mapping neuronal substrates of cognitive impairment are valid and useful.
Higuchi et al. (1997) N = 20	To examine regional cerebral glucose metabolism using PET in AD patients with defined genetic risk factors (APOE-*4*, ACT, and PS-1 genotypes)	• APOE-*4* does not adversely affect the AD process or preserve brain metabolism after clinical onset of AD. • ACT gene has deleterious effects on cerebral glucose metabolism during the clinical stages of AD. • Differences in cerebral regions are influenced by the two genes. • Inheritance pattern of the two alleles may explain divergent patterns of progression in AD.
Imamura et al. (1997) N = 38	To study regional cerebral glucose metabolism in AD vs. dementia with Lewy bodies (DLB)	• There are differences in regional glucose hypometabolism consistent with the pathological and neurochemical differences between DLB and AD. • FDG-PET may help in the clinical discrimination between DLB and AD.
Ishii et al. (1997) N = 81	To study regional cerebral and cerebellar glucose metabolic rates in AD	• There is a significant cerebellar glucose metabolic reduction in severe AD with no apparent cerebellar atrophy. • AD is a global degenerative brain disease in which degeneration is correlated with severity. • Method of analysis using normalization of regional glucose metabolic data to cerebellar values may be liable to err in severe AD patients.
Pietrini et al. (1997) N = 16	To study regional glucose metabolism under stress using an audiovisual paradigm in nondemented adults with trisomy 21 Down's syndrome	• There are no differences in metabolism at rest. • In older subjects had significantly lower glucose metabolic rates in the parietal and temporal cortical areas. • A stress test paradigm can detect metabolic abnormalities in the preclinical stages of AD.
Stein et al. (1998) N = 50	Using a template of Brodmann areas derived from whole brain histological section atlas to analyze glucose metabolic rates in AD patients	• Vulnerability is greatest in cortical areas that are in closer synaptic contact with limbic areas. • Integrating statistical techniques of brain imaging into neuroanatomical atlases and incorporating fine-tuned calibration of neuroanatomical studies into brain-imaging analyses, may increase correlation of findings and a more complete characterization of the pathophysiology of AD.
Vander Borght et al. (1997) N = 27	To study regional cerebral glucose metabolism in AD vs. Parkinson's disease with dementia (PDD)	• AD and PDD may share common features in the patterns of metabolic alterations and also presence of regional metabolic differences in the visual cortex and in the medical temporal cortex. • These differences may help explain different degrees and combinations of disease specific underlying pathological and neurochemical processes.
Yamaguchi et al. (1997) N = 23	To study regional glucose metabolism in hippocampal atrophy in AD	• Morphologic asymmetry of the hippocampus and a metabolic asymmetry of the temporoparieto-occipital were correlated. • These asymmetries are present in early stage AD.

Summary/Discussion

Recent evidence exploits functional imaging technologies such as PET for pathophysiologic information that may be applied toward earlier preclinical diagnoses of AD. Jagust (1996) highlighted the importance and the complexities of obtaining earlier and more accurate diagnoses of AD:

- Earlier diagnosis is important for understanding the biological mechanisms of AD;

- Clinically, early diagnosis becomes more critical, as treatments become available;
- Information from early diagnoses may enable forecasting which elderly persons who experience memory lapses will develop dementia;
- Normal aging processes can complicate early diagnosis; and
- Research should also assess factors key to the production of disease symptoms.

The best evidence demonstrating the accuracy of FDG PET in diagnosing Alzheimer's disease is from four published studies reviewed by Flynn (1996). They are listed in the Alzheimer's disease references (Section XI). Although these studies reported good diagnostic accuracy for PET in AD, the diagnostic utility of PET remains controversial:

- While each set of clinical criteria has different associated sensitivity, specificity, and likelihood ratios, careful application of the clinical criteria does appear to identify most cases of treatable dementia.
- Sources of bias attributed to the spectrum and severity of disease, the use of clinical criteria as the gold standard, and the choice of clinical criteria (NINCDS/ADRDA versus DSM-IIIR or DSM-IV) may have influenced diagnostic accuracy estimates in these studies.
- Few studies applied PET prospectively to large numbers of patients with a spectrum of dementia and disease severity, which would be necessary to define the positive predictive value of PET as a diagnostic test, and followed them until death.

Flynn (1996) reported that a cooperative group of European PET centers is conducting such a study. The study will include patients with NINCDS/ADRDA "possible" AD, the patients in whom there is the greatest uncertainty regarding diagnosis and for whom a more accurate test would most contribute to posttest certainty.

Small (1997) suggested that improved diagnostic information to patients and their families may allow families to better prepare for the challenges ahead and that early and accurate diagnosis may prevent the use of costly medical resources. The TA Program was unable to locate any studies of PET that assessed the impact of PET on the costs associated with caring for patients with AD.

Flynn (1996) concluded that existing evidence argues against routine clinical use of PET for diagnosing AD until more effective treatments and risk modification interventions for AD are developed, and until meaningful and robust predictive values are obtained from an ongoing European multicenter PET study.

The value of improved diagnostic information to AD patients and their families should not be dismissed; however, this value should be quantified in the context of accessibility and accuracy of alternative imaging technologies and of phenotypically or genetically defined subsets of AD. In the absence of effective treatments for AD, an accurate diagnostic test may be needed primarily in research for epidemiologic studies and evaluations of potential therapies.

IX. ONGOING CLINICAL STUDIES AND ON-LINE RESOURCES

Several on-line sources provide useful information about ongoing clinical trials:

- CenterWatch™ Clinical Trials Listing Service [http://www.centerwatch.com]
- NIH Clinical Research Studies [http://clinicalstudies.info.nih.gov/]
- NCI cancerTrials™ PDQ® database search [http://cancertrials.nci.nih.gov/]

These on-line sources were searched in November 1998 for active clinical trials studying the efficacy of FDG PET. Thirty-eight active protocols using FDG PET were retrieved, of which the following six protocols are assessing diagnostic PET for the conditions reviewed in this report.

Table 19: Active NIH Trials of FDG PET in Selected Cancers and Alzheimer's Disease

PROTOCOL	COMMENTS
NCI-94-C-0151 Diagnostic study of PET in patients with stage II-IV or recurrent breast cancer Single-site	**Sponsor-** NCI Start date 1994 Active accrual for at least 3 years
MSKCC-97046, NCI-G97-1308 Comparison of positron emitter Iodine I^{124} Iododeoxyuridine with fludeoxyglucose F 18 (F-18-Fluoro-2-Deoxy-(D)-Glucose) as a tracer for glycolysis on scans and in tumor samples in patients with advanced breast cancer Single site	**Sponsor-** local funding * Start date (1997) Active accrual for about 1 year
NCI-97-C-0068 Phase II study of Anti-CEA antibody immunoscintigraphy and PET in the localization of recurrent colorectal carcinoma in patients with rising serum CEA levels in the absence of imageable disease by conventional modalities Single site	**Sponsor-** NCI Start date (1997) Active accrual for 3 years
MSKCC-96079, NCI-G97-1334 Phase II/III Diagnostic Study of Whole Body PET to measure the response to induction chemotherapy of potentially resectable lung and esophageal carcinomas Single site	**Sponsor-** local funding * Start date (1997) Active accrual open
NCI-98-C-0163 The use of PET and MRI to assess the effects of anti-neoplastic therapy on tumor associated vasculature Unknown	**Sponsor-** NCI Start date 1998 Accrual pending
81-N-0010 Study of regional cerebral utilization of glucose in organic dementia and Down syndrome by the Laboratory of Neurosciences of The National Institute on Aging Unknown	**Sponsor-**National Institute of Neurological Disorders and Stroke (NINDS) Start date 1981 Active accrual

* Personal communication: Dr. Steven Larson, Memorial Sloan Kettering Cancer Center, New York

Since there is no central repository for locating active clinical trials of PET, these sources may not provide a complete listing of all multi-site studies evaluating PET as a clinical test.

Consequently, individuals actively involved in the use and evaluation of PET were queried for their knowledge of other relevant cooperative trials.

- NCI is funding a multi-center trial of FDG PET in staging **breast cancer**. The primary goal is to assess the accuracy of PET for detecting the presence, absence, and extent of axillary nodal metastases in women with newly diagnosed breast cancer; a secondary endpoint will evaluate PET for detecting internal mammary nodal disease as a prognostic indicator (personal communication: Dr. Barry Siegel, Washington University, St. Louis, Missouri).

- NCI is sponsoring a new cooperative group within the American College of Surgeons called the American College of Surgeons Oncology Group (ACoSOG) (NIH, 1998). The ACoSOG will design and conduct cooperative trials in surgical oncology. The primary goal of the ACoSOG is to evaluate surgical approaches for diagnosis and treatment of patients with malignant solid tumors. Patients with the most common cancers of the breast, lung, and colo-rectum will be studied initially. Completion of two protocols comparing the incremental value of PET to conventional staging in potentially operable patients with **lung cancer** and esophageal cancer is imminent (personal communication: Dr. Barry Siegel)**.**

- The Southwest Oncology Group (SWOG) is developing a companion study within a Phase III cooperative trial comparing surgery and pre-operative chemotherapy for patients with **lung cancer**. The companion study will evaluate PET in assessing tumor response to chemotherapy. Both studies will be activated in 1999 (personal communication: Suzan Myers, SWOG).

X. OTHER SYSTEMATIC REVIEWS OF PET

Since 1996 several organizations have conducted assessments to support evidence-based recommendations for the use of PET as a diagnostic test (See Appendix V). The majority of assessments were qualitative systematic reviews of dedicated PET used in neurology to diagnose and manage patients with medically refractory partial seizures, central nervous system tumors, and cerebrovascular disease. Recent systematic reviews reflect an increasing interest in PET and in other positron imaging modalities to manage patients with non-central nervous system cancers, emphasizing staging non-small cell lung cancer.

For the indications in this review, the findings of assessments with either full text or abstracts in English in the public domain, or otherwise available to the MDRC, are summarized below:

- **There is general agreement that the evidence on FDG-PET for diagnosing, staging or monitoring treatment of primary cancers outside the lung is not firmly established.**

- **There is general agreement that the effect of PET on the management of patients with primary lung cancers is not known.**

  The Agencia de Evaluación de Tecnologías Sanitarias (AETS) in Spain, the Committee for Evaluation and Diffusion of Innovative Technologies (CEDIT) in France, and the NHS Health Technology Assessment Programme (NHS HTAP) in the United Kingdom recommend comparative studies of effectiveness and of the diagnostic contribution of

dedicated PET (and, in some cases, coincidence imaging gamma cameras) in patients with lung cancer.

- **Assessment findings and recommendations are mixed regarding the use of PET to diagnose and stage non-small cell lung cancer and solitary pulmonary nodules (SPNs).**

  Two agencies, AETS and the NHS HTAP, used VA review methods and frameworks to update and/or expand the first VA PET report (Flynn, 1996). Both reports confirmed VA's original findings that the evidence for the diagnostic efficacy of PET in managing patients with lung cancer was insufficient. Blue Cross/Blue Shield Association found that FDG-PET imaging meet their quality assessment criteria for staging mediastinal lymph nodes and characterizing radiographically indeterminate SPNs, provided the test results could change medical management (HCFA, 1997).

  An ECRI quantitative analysis determined that for both lung cancer indications PET is cost-effective when used to confirm resectability, but that PET is not cost-effective when used earlier in the diagnostic algorithm. A SPN strategy using CT for initial diagnosis, needle biopsy to confirm positive results, and PET to confirm negative results attained the greatest life expectancy (Mitchell, 1998).

  There are several possible reasons for the discrepancies across these assessments. Variations in criteria for including published studies and for judging the quality of the included studies, in analytical methods, in the rationale for the assessment, and in the focus of the report are likely causes. Often, assessments must be purchased or may require language translation to be systematically evaluated. For this review, the MDRC considered information available only in the public domain in English or with English translation. Proprietary or non-translated reports may have derived different conclusions. Valid comparisons of technology assessments that address similar topics are critical to health care organizations wishing to establish policies based on the best available evidence.

  Increasingly, agencies are using quantitative analyses, (e.g., decision analyses, meta-analyses, and cost-effectiveness analyses) to quantify the utility of clinical PET. Many analyses extrapolate existing diagnostic accuracy estimates to population impact, or pool accuracy results from multiple studies. It is important to note that the validity of the studies that are the source of these estimates is an essential consideration when evaluating the robustness of the results (Petitti, 1994).

- **Until recently, agencies considered only dedicated PET scanners, but now are asked to review other positron imaging modalities**.

  An expert panel at CEDIT considered coincidence imaging gamma cameras and dedicated PET in their recommendations. The NHS HTAP report will include evaluations of partial ring PET, coincidence imaging gamma cameras, and collimated 511 keV imaging.

PET is a topic for a joint project of the International Network of Agencies for Health Technology Assessment (INAHTA), to which the TA Program belongs. The TA Program is coordinating the project with members from Spain and the Agency for Health Care Policy and Research. Member agencies are collaborating to synthesize their assessments of clinical PET applications into a

single, broadly applicable document. The report will also include a description of the evolution of PET use in the United States and current indications and coverage policies of PET among countries represented by INAHTA members. The report will be available in 1999 on the INAHTA web site at [http://www.inahta.org].

XI. CONCLUSIONS

A. Experience in VA

VHA continues to make a substantial resource commitment to its PET imaging facilities. This commitment has the potential to help support two parts of VHA's mission: research and clinical care. The medical community regards PET as an important basic research tool. A survey of active funded research at VHA PET sites underscores this importance, with the vast majority of basic research activity in neurology and cardiology. VHA is maximizing its investment in PET by supporting high quality outcomes research and systematic collection of utilization data.

All VHA PET sites have access to FDG, enabling them to conduct glucose metabolic studies for various clinical applications. The number of PET oncology studies conducted across VHA PET facilities from FY 1994 to FY 1998 has nearly quadrupled, likely reflecting the positive changes in Medicare and private sector reimbursement and changes in practitioners' attitudes. Since VHA continues its moratorium on adding dedicated PET centers to its system, many VA medical centers without access to dedicated PET scanners are adapting existing dual-headed gamma cameras for coincidence detection.

B. Systematic reviews

The prevailing evidence does not support the use of either dedicated or gamma cameras modified for coincidence detection (camera-based PET) as a diagnostic test for the applications in this review. All studies were subject to considerable bias, which will have resulted in overestimating accuracy and clinical value. Several studies presented anecdotal data on the influence of PET on changing diagnostic certainty and treatment planning, but the methods for assessing these changes were not described, and the systematic nature could not be determined.

Caution must be exercised to not apply accuracy estimates from dedicated PET to camera-based PET systems. Whereas dedicated PET scanners are limited primarily to tertiary care institutions, dual-headed gamma camera systems are more widely employed. Technical differences between the two systems and potential differences in the study populations represented across different health facilities emphasize the need for large, rigorous studies of diagnostic efficacy to define the clinical role of camera-based PET.

The TA Program identified several methodologically rigorous studies of other diagnostic imaging modalities that could serve as models for designing future PET research (Appendix II). Incorporating aspects from these studies would correct the methodologic shortcomings of the existing literature and strengthen the evidence on which to base future patient care decisions.

Qualitative systematic reviews produced by other technology assessment agencies, which used methods similar to the VA PET report, reached similar conclusions. Most agencies agree that the effect of positron imaging on managing patients with cancer needs further study. Several cooperative trials and other data collection efforts are ongoing or are being proposed that may address many unanswered questions regarding the utility of FDG PET in the work up of patients with cancer and Alzheimer's disease. **Clinicians should await the results of these efforts before incorporating PET into routine diagnostic strategies.** Nonetheless, variations across studies in study populations, imaging protocols, threshold values, and formulae for calculating quantitative uptake values may limit the generalizability of the findings to other institutions and populations. **Review of recent evidence confirms the conclusions from the original VA PET assessment (Flynn, 1996).**

Information on some of the cooperative trials can be accessed through on-line data sources. **Advocates of clinical PET and decision makers interested in its clinical utility would benefit from an accessible central repository containing information on existing and proposed rigorously designed cooperative trials of PET.** This source could help guide the diffusion of PET into clinical care, as its usefulness and contribution to improved patient outcomes are appropriately evaluated.

XII. REFERENCES

Background

Positron Emission Tomography. In: ECRI, Ed. *Health Technology Forecast*. Plymouth Meeting, PA: ECRI; 1996.

Food and Drug Modernization Act of 1997. s830. 1997, 105th Congress: Washington, DC.

Health Care Financing Administration. Health Care Financing Administration. (1999). *Carriers manual part 3. Chapter IV- Claims review and adjudication procedures.* [Web site]. Available: **http://www.hcfa.gov/pubforms/14%5Fcar/3b4120.htm,** [July 20, 1999].

Illumina Interactive, Boston, MA. (1998). *CenterWatch Clinical Trials Listing Service*^{TM}. [Web Site]. Available: **http://www.centerwatch.com/,** [May 27, 1998].

^{18}F-Labeled 2-Deoxy-2-Fluoro-D-Glucose Positron-Emission Tomography Scans for the Localization of the Epileptogenic Foci. Health Technology Assessment Number 12. AHCPR Pub. No. 98-0044. 1998, U.S. Department for Health and Human Services. Agency for Health Care Policy and Research: Rockville, MD.

American Cancer Society. *Cancer facts & figures - 1998.* 1998, American Cancer Society: Atlanta, GA.

Balch CM, Singletary SE, Bland KI. Clinical decision-making in early breast cancer [see comments]. *Ann Surg* 1993; 217(3): 207-225.

Begg CB. Biases in the assessment of diagnostic tests. *Stat Med* 1987; 6(4): 411-423.

Bombardieri E, Crippa F, Maffioli L, Greco M. Nuclear medicine techniques for the study of breast cancer. *Eur J Nucl Med* 1997; 24(7): 809-824.

Bondi MW. U.S. Department of Veterans Affairs. Medical Research. (April 5, 1997). *Cognitive and Imaging Studies of Older Adults with the ApoE-E4 Allele.* [Web site]. Available: **http://www.va.gov/research/files/VA011308.HTM,** [April 28, 1999].

Brown J. *Annual report of the Secretary of Veterans Affairs.* 1997, Department of Veterans Affairs: Washington, DC.

Buscombe JR, Cwikla JB, Thakrar DS, Hilson AJ. Scintigraphic imaging of breast cancer: a review. *Nucl Med Commun* 1997; 18(8): 698-709.

Camberg L, Woods P, Ooi WL, Hurley A, Volicer L, Ashley J, et al. Evaluation of simulated presence: a personalized approach to enhance well-being in persons with Alzheimer's disease. *J Am Geriatr Soc* 1999; 47(4): 446-452.

Coleman RE. Camera-based PET: the best is yet to come [editorial]. *J Nucl Med* 1997; 38(11): 1796-1797.

Cook DJ, Guyatt G, Laupacis A, Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents [published erratum appears in Chest 1994 Feb;105(2):647]. *Chest* 1992; 102(4 Suppl): 305S-311S.

ECRI. Diagnostic imaging. *Health Technology Forecast* 1996; 7-16.

Eysenck HJ. Meta-analysis and its problems. *Bmj* 1994; 309(6957): 789-792.

Fahey FH. State of the art in emission tomography equipment. *Radiographics* 1996; 16(2): 409-420.

Feussner JR. *Refining Research Priorities: New Initiatives Meeting Veterans Needs.* 1997, Veterans Health Administration. Office of Research & Development: Washington, DC.

Flynn K, Adams E, Anderson D, et. al. *Positron emission tomography: descriptive analysis of experience with PET in VA and systematic reviews: FDG-PET as a diagnostic test for cancer and Alzheimer's disease.* 1996, U.S. Department of Veterans Affairs. Management Decision & Research Center: Washington, DC.

Food and Drug Administration. Revocation of certain guidance documents on Positron Emission Tomography drug products. *Federal Register* 1997; 62(244): 66636.

Food and Drug Administration. Revocation of regulation on Positron Emission Tomography drug products. *Federal Register* 1997; 62(244): 66522.

Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991; 11(2): 88-94.

Goodwin WJ. PET and recurrent squamous cell carcinoma of the head and neck: a surgeon's view. *AJNR: Am J Neuroradiol* 1998; 19(7): 1197.

Gurney JW. Determining the likelihood of malignancy in solitary pulmonary nodules with Bayesian analysis. Part I. Theory. *Radiology* 1993; 186(2): 405-413.

Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ. Users' guides to the medical literature. IX. A method for grading health care recommendations. Evidence-Based Medicine Working Group [published erratum appears in JAMA 1996 Apr 24;275(16):1232]. *JAMA* 1995; 274(22): 1800-1804.

Haynes RB, Sackett D. Purpose and procedure (abbreviated). *Evidence-based Medicine* 1995; 1 2.

Health Care Financing Administration. Health Care Financing Administration. (1998). *Transmittal No. AB-97-27. Implementing Instructions - PET Scans for Characterizing Solitary Pulmonary Nodules or Staging Lung Cancer Performed on or After January 1, 1998.* [Web Site]. Available: **httP//www.hcfa.org/pubforms/transmit/AB972760.htm/,** [January 13, 1999].

Health Care Financing Administration. Health Care Financing Administration. (1998). *Transmittal No. A-98-9. Positron Emission Tomography(PET) Scans-ACTION.* [Web Site]. Available: **http://www.hcfa.gov/pubforms/transmit/a98960.htm/,** [March 23].

Hendrie HC. Epidemiology of dementia and Alzheimer's disease. *Am J Geriatr Psychiatry* 1998; 6(2 Suppl 1): S3-18.

Hoffman JM, Hanson MW, Coleman RE. Clinical positron emission tomography imaging. *Radiology Clinics of North America* 1993; 31 935.

Ibrahim MA. *Epidemiology and health policy.* Rockville, Md.: Aspen Systems Corp; 1985.

Institute for Clinical PET. Institute for Clinical PET, Foothill Ranch, CA. (Feb. 2, 1998). *U.S. PET Centers.* [Web Site]. Available: **http://www.icppet.org,** [May 3, 1999].

International Network of Agencies for Health Technology Assessment. International Network of Agencies for Health Technology Assessment. (1999). *Homepage.* [Web site]. Available: **http://www.inahta.org/,** [April 28, 1999].

Ioannidis JP, Cappelleri JC, Lau J. Issues in comparisons between meta-analyses and large trials. *JAMA* 1998; 279(14): 1089-1093.

Jagust WJ. Functional imaging patterns in Alzheimer's disease. Relationships to neurobiology. *Ann N Y Acad Sci* 1996; 777 30-36.

Jarritt PH, Acton PD. PET imaging using gamma camera systems: a review. *Nucl Med Commun* 1996; 17(9): 758-766.

Kent DL, Larson EB. Magnetic resonance imaging of the brain and spine. Is clinical efficacy established after the first decade? [published erratum appears in Ann Intern Med 1988 Sep 1;109(5):438]. *Ann Intern Med* 1988; 108(3): 402-424.

Kent DL, Larson EB. Disease, level of impact, and quality of research methods. Three dimensions of clinical efficacy assessment applied to magnetic resonance imaging. *Invest Radiol* 1992; 27(3): 245-254.

Kent DL, Haynor DR, Longstreth WT, Larson EB. The clinical efficacy of magnetic resonance imaging in neuroimaging [see comments]. *Ann Intern Med* 1994; 120(10): 856-871.

Larson S. Personal communication. November 19, 1998.
Laupacis A, Wells G, Richardson WS, Tugwell P. Users' guides to the medical literature. V. How to use an article about prognosis. Evidence-Based Medicine Working Group. *JAMA* 1994; 272(3): 234-237.

Lieberman DA. U.S. Department of Veterans Affairs. Cooperatives Studies Program. (1998). *Prospective Evaluation of Risk Factors for Large Colonic Adenomas in Asymptomatic Subjects.* [Web Site]. Available: **http://www.va.gov/research/files/VA011064.HTM,** [April 28, 1999].

Lillington GA, CI. C. Evaluation and management of solitary and multiple pulmonary nodules. *Clinics in Chest Medicine* 1993; 14(1): 111-119.

Maublant J. Scintigraphic imaging of breast tumors. *Eur J Radiol* 1997; 24(1): 2-10.

Mitchell MD, Turkelson C, Doggett D. *Cost-effectiveness analysis of PET in lung cancer diagnosis and staging*. Annual Meeting of the International Society for Technology Assessment in Health Care, *Abstracts.* Vol. 14; 33, 1998.

Mushlin AI, Detsky AS, Phelps CE, O'Connor PW, Kido DK, Kucharczyk W. The accuracy of magnetic resonance imaging in patients with suspected multiple sclerosis. The Rochester-Toronto Magnetic Resonance Imaging Study Group. *JAMA* 1993; 269(24): 3146-3151.

National Institute of Health. Warren Grant Magnuson Clinical Center, NIH. (1997). *NIH Clinical research studies protocol database.* [Web Site]. Available: **http://clinicalstudies.info.nih.gov/**, [July 7, 1997].

National Institutes of Health. National Cancer Institute. National Institutes of Health,. (March 27, 1998). *American College of Surgeons and National Cancer Institute Announce Funding of Clinical Trials Project.* [Web site]. Available: **http://www.nih.gov/news/pr/mar98/nci-27.htm**, [April 28, 1999].

PDQ. National Cancer Institute. (1998). *PDQ clinical trials search.* [Web Site]. Available: **http://cancernet.nci.nih.gov/prot/patsrch.shtml/**, [December 1998].

PDQ. National Cancer Institute. (1999). *PDQ Treatment Health Professionals: Lip and oral cavity cancer.* [Web Site]. Available: **http://cancernet.nci.nih.gov/,** [April 30, 1999].

PDQ. National Cancer Institute. (1999). *PDQ Treatment Health Professionals: Colon Cancer.* [Web Site]. Available: **http://cancernet.nci.nih.gov/,** [April 30, 1999].

PDQ. National Cancer Institute. (1999). *PDQ Treatment Health Professionals: Laryngeal Cancer.* [Web Site]. Available: **http://cancernet.nci.nih.gov/,** [April 30, 1999].

PDQ. National Cancer Institute. (1999). *PDQ Treatment Health Professionals: Hypopharyngeal cancer.* [Web Site]. Available: **http://cancernet.nci.nih.gov/,** [April 30, 1999].

PDQ. National Cancer Institute. (April, 1999). *PDQ Treatment Health Professionals: Oropharyngeal cancer.* [Web Site]. Available: **http://cancernet.nci.nih.gov/,** [April 30, 1999].

Petitti DB. *Meta-Analysis, Decision-Analysis, and Cost-Effectiveness Analysis: methods for quantitative synthesis in medicine.* New York, NY: Oxford University Press; 1994.

Provenzale D. *Colorectal Cancer: Risk Factors for Advanced Disease - Comparison of Stage at Diagnosis with SEER Cancer Statistics*. Veterans Affairs Health Services Research &

Development 17th Annual Meeting, *Health Services Research at the Interface. Submitted Abstracts.* 131-132, February 24-26, 1999.

Rifkin MD, Zerhouni EA, Gatsonis CA, Quint LE, Paushter DM, Epstein JI. Comparison of magnetic resonance imaging and ultrasonography in staging early prostate cancer. Results of a multi-institutional cooperative trial [see comments]. *N Engl J Med* 1990; 323(10): 621-626.

Sackett D. *Clinical epidemiology: a basic science for clinical medicine*. Boston: Little Brown; 1991.

Shreve PD, Steventon RS, Deters EC, Kison PV, Gross MD, Wahl RL. Oncologic diagnosis with 2-[fluorine-18]fluoro-2-deoxy-D-glucose imaging: dual-head coincidence gamma camera versus positron emission tomographic scanner. *Radiology* 1998; 207(2): 431-437.

Small GW, Rabins PV, Barry PP, Buckholtz NS, DeKosky ST, Ferris SH, et al. Diagnosis and treatment of Alzheimer disease and related disorders. Consensus statement of the American Association for Geriatric Psychiatry, the Alzheimer's Association, and the American Geriatrics Society [see comments]. *JAMA* 1997; 278(16): 1363-1371.

Stark DD, Wittenberg J, Butch RJ, Ferrucci JT. Hepatic metastases: randomized, controlled comparison of detection with MR imaging and CT. *Radiology* 1987; 165(2): 399-406.

Stevens T. (1997). *News Release 11/6/97- Sen. Stevens announces new Medicare reimbursement policy for PET scans used for screening and staging of lung cancer.* [Press Release]. Available: **http://www.dana.org/dana/pr_110697.html/,** [December 17, 1997].

Stotland BR, Siegelman ES, Morris JB, Kochman ML. Preoperative and postoperative imaging for colorectal cancer. *Hematology/Oncology Clinics of North America* 1997; 11(4): 635-654.

Thornbury JR, Kido DK, Mushlin AI, Phelps CE, Mooney C, Fryback DG. Increasing the scientific quality of clinical efficacy studies of magnetic resonance imaging. *Invest Radiol* 1991; 26(9): 829-835.

Thornbury JR, Fryback DG. Technology assessment--an American view. *Eur J Radiol* 1992; 14(2): 147-156.

Thornbury JR, Fryback DG, Turski PA, Javid MJ, McDonald JV, Beinlich BR. Disk-caused nerve compression in patients with acute low-back pain: diagnosis with MR, CT myelography, and plain CT [published erratum appears in Radiology 1993 Jun;187(3):880] [see comments]. *Radiology* 1993; 186(3): 731-738.

U.S. Department of Health and Human Services. Health Care Financing Administration. Medicare Program; Open Town Hall Meeting to Discuss the Positron Emission Tomography. *Federal Register* 1999; 64(1): 170-171.

U.S. Department of Health and Human Services. Health Care Financing Administration. U.S. Department of Health and Human Services. Health Care Financing Administration. (April 25, 1999). *Coverage Issues Manual - Diagnostic Services.* [Web site]. Available: [April 28, 1999].

U.S. Department of Health and Human Services. Health Care Financing Administration. (March 8, 1999). *Positron Emission Tomography (PET) Scan Coverage Expanded.* [Web site]. Available: **http://www.hcfa.gov/news/pr1999/n990308.htm,** [May 3, 1999].

U.S. Department of Health and Human Services. Health Care Financing Administration. Technology Advisory Committee. (September 29, 1997). *Technology Advisory Committee Minutes, August 5 and 6, 1997.* [Web site]. Available: **http://www.hcfa.gov/events/0897tmin.htm,** [May 3, 1999].

U.S. Department of Veterans Affairs. Veterans Health Administration. Office of Research and Development. *Improving Health Care for Veterans: 1997 Annual Report.* 1998, U.S. Department of Veterans Affairs. Veterans Health Administration. Office of Research and Development: Washington, D.C.

U.S. Secretary of Veterans Affairs. *Annual Report Fiscal Year 1995.* 1996, U.S. Department of Veterans Affairs: Washington, D.C.

Vitola JV, Delbeke D, Sandler MP, Campbell MG, Powers TA, Wright Jk. Positron emission tomography to stage suspected metastatic colorectal carcinoma to the liver. *Am J Surg* 1996; 171(1): 21-26.

Volicer L, Collard A, Hurley A, Bishop C, Kern D, Karon S. Impact of special care unit for patients with advanced Alzheimer's disease on patients' discomfort and costs [see comments]. *J Am Geriatr Soc* 1994; 42(6): 597-603.

Wahl RL. Overview of the current status of PET in breast cancer imaging. *Quarterly Journal of Nuclear Medicine* 1998; 42(1): 1-7.

Webb WR, Gatsonis C, Zerhouni EA, Heelan RT, Glazer GM, Francis IR. CT and MR imaging in staging non-small cell bronchogenic carcinoma: report of the Radiologic Diagnostic Oncology Group. *Radiology* 1991; 178(3): 705-713.

Winawer SJ, Fletcher RH, Miller L, Godlee F, Stolar MH, Mulrow CD, et al. Colorectal cancer screening: clinical guidelines and rationale [see comments] [published errata appear in Gastroenterology 1997 Mar;112(3):1060 and 1998 Mar;114(3):625]. *Gastroenterology* 1997; 112(2): 594-642.

Wingo P, Tong T, Bolden S. Cancer Statistics 1995. *CA Cancer J Clin* 1996; 65 5-27.

Zerhouni EA, Rutter C, Hamilton SR, Balfe DM, Megibow AJ, Francis IR, et al. CT and MR imaging in the staging of colorectal carcinoma: report of the Radiology Diagnostic Oncology Group II. *Radiology* 1996; 200(2): 443-451.

Head and Neck

Included Studies

Diagnostic Accuracy

Lowe VJ, Dunphy FR, Varvares M, Kim H, Wittry M, Dunphy CH, et al. Evaluation of chemotherapy response in patients with advanced head and neck cancer using [F-18] fluorodeoxyglucose positron emission tomography. *Head Neck* 1997; 19(8): 666-674.

Myers LL, Wax MK, Nabi H, Simpson GT, Lamonica D. Positron emission tomography in the evaluation of the N0 neck. *Laryngoscope* 1998; 108(2): 232-236.

Wong WL, Chevretton EB, McGurk M, Hussain K, Davis J, Beaney R, et al. A prospective study of PET-FDG imaging for the assessment of head and neck squamous cell carcinoma. *Clin Otolaryngol* 1997; 22(3): 209-214.

Technical Efficacy

Braams JW, Pruim J, Kole AC, Nikkels PG, Vaalburg W, Vermey A, et al. Detection of unknown primary head and neck tumors by positron emission tomography. *Int J Oral Maxillofac Surg* 1997; 26(2): 112-115.

Kuhn GD, Reisser C, DimitrakopoulouStrauss A, Oberdorfer F, Strauss LG. PET studies of perfusion and glucose metabolism in patients with untreated head and neck tumours. *Onkologie* 1997; 20(3): 226-230.

Minn H, Lapela M, Klemi PJ, Grenman R, Leskinen S, Lindholm P, et al. Prediction of survival with fluorine-18-fluorodeoxyglucose and PET in head and neck cancer. *Journal Of Nuclear Medicine* 1997; 38(12): 1907-1911.

Sakamoto H, Nakai Y, Ohashi Y, Okamura T, Ochi H. Positron emission tomographic imaging of head and neck lesions. *Eur Arch Otorhinolaryngol Suppl* 1997; 1 pS123-6:

Excluded Studies

Changlai SP, Kao CH, Chieng PU. 18F-2-fluoro-2-deoxy-D-glucose positron emission tomography of head and neck in patients with nasopharyngeal carcinomas. *Oncology Reports* 1997; 4(6): 1331-1334.

McGuirt WF, Greven K, Williams D, 3rd, Keyes JW, Jr., Watson N, Cappellari JO, et al. PET scanning in head and neck oncology: a review. *Head Neck* 1998; 20(3): 208-215.

Breast Cancer

Included Studies

Diagnostic Accuracy

Adler LP, Faulhaber PF, Schnur KC, Al-Kasi NL, Shenk RR. Axillary lymph node metastases: Screening with (F-18)2-deoxy-2-fluoro- D-glucose (FDG) PET. *Radiology* 1997; 203(2): 323-327.

Bender H, Kirst J, Palmedo H, Schomburg A, Wagner U, Ruhlmann J, et al. Value of 18 fluoro-deoxyglucose positron emission tomography in the staging of recurrent breast carcinoma. *Anticancer Res* 1997; 17(3b): 1687-1692.

Crippa F, Agresti R, Seregni E, Greco M, Pascali C, Bogni A, et al. Prospective evaluation of fluorine-18-FDG PET in presurgical staging of the axilla in breast cancer. *J Nucl Med* 1998; 39(1): 4-8.

Moon DH, Maddahi J, Silverman DH, Glaspy JA, Phelps ME, Hoh CK. Accuracy of whole-body fluorine-18-FDG PET for the detection of recurrent or metastatic breast carcinoma. *J Nucl Med* 1998; 39(3): 431-435.

Palmedo H, Bender H, Grunwald F, Mallmann P, Zamora P, Krebs D, et al. Comparison of fluorine-18 fluorodeoxyglucose positron emission tomography and technetium-99m methoxyisobutylisonitrile scintimammography in the detection of breast tumours. *Eur J Nucl Med* 1997; 24(9): 1138-1145.

Utech CI, Young CS, Winter PF. Prospective evaluation of fluorine-18 fluorodeoxyglucose positron emission tomography in breast cancer for staging of the axilla related to surgery and immunocytochemistry. *Eur J Nucl Med* 1996; 23(12): 1588-1593.

Technical Efficacy

Avril N, Bense S, Ziegler SI, Dose J, Weber W, Laubenbacher C, et al. Breast imaging with fluorine-18-FDG PET: quantitative image analysis. *J Nucl Med* 1997; 38(8): 1186-1191.

Noh DY, Yun IJ, Kim JS, Kang HS, Lee DS, Chung JK, et al. Diagnostic value of positron emission tomography for detecting breast cancer. *World J Surg* 1998; 22(3): 223-227; discussion 227-228.

Oshida M, Uno K, Suzuki M, Nagashima T, Hashimoto H, Yagata H, et al. Predicting the prognoses of breast carcinoma patients with positron emission tomography using 2-deoxy-2-fluoro[18F]-D-glucose. *Cancer* 1998; 82(11): 2227-2234.

Torizuka T, Zasadny KR, Recker B, Wahl RL. Untreated primary lung and breast cancers: correlation between F-18 FDG kinetic rate constants and findings of in vitro studies. *Radiology* 1998; 207(3): 767-774.

Excluded Studies

Crippa F, Agresti R, Donne VD, Pascali C, Bogni A, Chiesa C, et al. The contribution of positron emission tomography (PET) with 18F-fluorodeoxyglucose (FDG) in the preoperative detection of axillary metastases of breast cancer: the experience of the National Cancer Institute of Milan. *Tumori* 1997; 83(2): 542-543.

Holle LH, Trampert L, Lung-Kurt S, Villena-Heinsen CE, Puschel W, Schmidt S, et al. Investigations of breast tumors with fluorine-18-fluorodeoxyglucose and SPECT. *J Nucl Med* 1996; 37(4): 615-622.

Katzenellenbogen JA, Welch MJ, Dehdashti F. The development of estrogen and progestin radiopharmaceuticals for imaging breast cancer. *Anticancer Res* 1997; 17(3b): 1573-1576.

Kole AC, Nieweg OE, Pruim J, Paans AM, Plukker JT, Hoekstra HJ, et al. Standardized uptake value and quantification of metabolism for breast cancer imaging with FDG and L-[1-11C]tyrosine PET. *J Nucl Med* 1997; 38(5): 692-696.

Petren-Mallmin M, Andreasson I, Ljunggren O, Ahlstrom H, Bergh J, Antoni G, et al. Skeletal metastases from breast cancer: uptake of 18F-fluoride measured with positron emission tomography in correlation with CT. *Skeletal Radiol* 1998; 27(2): 72-76.

Lung Cancer

Included Studies

Diagnostic Accuracy

Bury T, Dowlati A, Paulus P, Corhay JL, Hustinx R, Ghaye B, et al. Whole-body (18)FDG positron emission tomography in the staging of non-small cell lung cancer. *European Respiratory Journal* 1997; 10(11): 2529-2534.

Erasmus JJ, Patz EF, Jr., McAdams HP, Murray JG, Herndon J, Coleman RE, et al. Evaluation of adrenal masses in patients with bronchogenic carcinoma using 18F-fluorodeoxyglucose positron emission tomography [see comments]. *AJR Am J Roentgenol* 1997; 168(5): 1357-1360.

Guhlmann A, Storck M, Kotzerke J, Moog F, Sunder-Plassmann L, Reske SN. Lymph node staging in non-small cell lung cancer: evaluation by [18F] FDG positron emission tomography (PET). *Thorax* 1997; 52(5): 438-441.

Sasaki M, Ichiya Y, Kuwabara Y, Akashi Y, Yoshida T, Fukumura T, et al. The usefulness of FDG positron emission tomography for the detection of mediastinal lymph node metastases in patients with non-small cell lung cancer: a comparative study with X-ray computed tomography. *Eur J Nucl Med* 1996; 23(7): 741-747.

Vansteenkiste JF, Stroobants SG, De Leyn PR, Dupont PJ, Verschakelen JA, Nackaerts KL, et al. Mediastinal lymph node staging with FDG-PET scan in patients with

potentially operable non-small cell lung cancer: a prospective analysis of 50 cases. Leuven Lung Cancer Group. *Chest* 1997; 112(6): 1480-1486.

Technical Efficacy

Bury T, Paulus P, Dowlati A, Corhay JL, Rigo P, Radermecker MF. Evaluation of pleural diseases with FDG-PET imaging: Preliminary report. *Thorax* 1997; 52(2): 187-189.

Ferlin G, Rubello D, Chierichetti F, Zanco P, Bergamin R, Trento P, et al. The role of fluorine-18-deoxyglucose (FDG) positron emission tomography (PET) whole body scan (WBS) in the staging and follow-up of cancer patients: Our first experience. *Tumori* 1997; 83(3): 679-684.

Higashi K, Nishikawa T, Seki H, Oguchi M, Nambu Y, Ueda Y, et al. Comparison of fluorine-18-FDG PET and thallium-201 SPECT in evaluation of lung cancer. *Journal Of Nuclear Medicine* 1998; 39(1): 9-15.

Kim BT, Kim Y, Lee KS, Yoon SB, Cheon EM, Kwon OJ, et al. Localized form of bronchioloalveolar carcinoma: FDG PET findings. *AJR Am J Roentgenol* 1998; 170(4): 935-939.

Nettelbladt OS, Sundin AE, Valind SO, Gustafsson GR, Lamberg K, Langstrom B, et al. Combined fluorine-18-FDG and carbon-11-methionine PET for diagnosis of tumors in lung and mediastinum. *J Nucl Med* 1998; 39(4): 640-647.

Torizuka T, Zasadny KR, Recker B, Wahl RL. Untreated primary lung and breast cancers: correlation between F-18 FDG kinetic rate constants and findings of in vitro studies. *Radiology* 1998; 207(3): 767-774.

Wang H, Maurea S, Mainolfi C, Fiore F, Gravina A, Panico MR, et al. Tc-99m MIBI scintigraphy in patients with lung cancer. Comparison with CT and fluorine-18 FDG PET imaging. *Clin Nucl Med* 1997; 22(4): 243-249.

Excluded Studies

Erasmus JJ, McAdams HP, Patz EF, Jr., Coleman RE, Ahuja V, Goodman PC. Evaluation of primary pulmonary carcinoid tumors using FDG PET. *AJR Am J Roentgenol* 1998; 170(5): 1369-1373.

Erdi YE, Mawlawi O, Larson SM, Imbriaco M, Yeung H, Finn R, et al. Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding. *Cancer* 1997; 80(12 Suppl): 2505-2509.

SPN

Included Studies

Diagnostic Accuracy

Dewan NA, Shehan CJ, Reeb SD, Gobar LS, Scott WJ, Ryschon K. Likelihood of malignancy in a solitary pulmonary nodule: comparison of Bayesian analysis and results of FDG-PET scan. *Chest* 1997; 112(2): 416-422.

Lowe VJ, Fletcher JW, Gobar L, Lawson M, Kirchner P, Valk P, et al. Prospective investigation of positron emission tomography in lung nodules. *J Clin Oncol* 1998; 16(3): 1075-1084.

Excluded Studies

Lowe VJ, Duhaylongsod FG, Patz EF, Delong DM, Hoffman JM, Wolfe WG, et al. Pulmonary abnormalities and PET data analysis: a retrospective study [see comments]. *Radiology* 1997; 202(2): 435-439.

Colorectal Cancer

Included Studies

Diagnostic Accuracy

Delbeke D, Vitola JV, Sandler MP, Arildsen RC, Powers TA, Wright JK, Jr., et al. Staging recurrent metastatic colorectal carcinoma with PET. *J Nucl Med* 1997; 38(8): 1196-1201.

Flanagan FL, Dehdashti F, Ogunbiyi OA, Kodner IJ, Siegel BA. Utility of FDG-PET for investigating unexplained plasma CEA elevation in patients with colorectal cancer [see comments]. *Ann Surg* 1998; 227(3): 319-323.

Ogunbiyi OA, Flanagan FL, Dehdashti F, Siegel BA, Trask DD, Birnbaum EH, et al. Detection of recurrent and metastatic colorectal cancer: Comparison of position emission tomography and computed tomography. *Annals of Surgical Oncology* 1997; 4(8): 613-620.

Technical Efficacy

Abdel-Nabi H, Doerr RJ, Lamonica DM, Cronin VR, Galantowicz PJ, Carbone GM, et al. Staging of primary colorectal carcinomas with fluorine-18 fluorodeoxyglucose whole-body PET: correlation with histopathologic and CT findings. *Radiology* 1998; 206(3): 755-760.

Ruhlmann J, Schomburg A, Bender H, Oehr P, Robertz-Vaupel GM, Vaupel H, et al. Fluorodeoxyglucose whole-body positron emission tomography in colorectal cancer patients studied in routine daily practice. *Dis Colon Rectum* 1997; 40(10): 1195-1204.

Alzheimer's Disease

Included Studies

Diagnostic Accuracy (reviewed in 1996 report)

Burdette JH, Minoshima S, Vander Borght T, Tran DD, Kuhl DE. Alzheimer disease: improved visual interpretation of PET images by using three-dimensional stereotaxic surface projections. *Radiology* 1996; 198(3): 837-843.

Herholz K, Perani D, Salmon E, Franck G, Fazio F, Heiss WD, et al. Comparability of FDG PET studies in probable Alzheimer's disease. *J Nucl Med* 1993; 34(9): 1460-1466.

Kippenhan JS, Barker WW, Nagel J, Grady C, Duara R. Neural-network classification of normal and Alzheimer's disease subjects using high-resolution and low-resolution PET cameras [see comments]. *J Nucl Med* 1994; 35(1): 7-15.

Salmon E, Sadzot B, Maquet P, Degueldre C, Lemaire C, Rigo P, et al. Differential diagnosis of Alzheimer's disease with PET. *J Nucl Med* 1994; 35(3): 391-398.

Technical Efficacy

Desgranges B, Baron JC, de la Sayette V, Petit-Taboue MC, Benali K, Landeau B, et al. The neural substrates of memory systems impairment in Alzheimer's disease: A PET study of resting brain glucose utilization. *Brain* 1998; 121(Pt 4): 611-631.

Higuchi M, Arai H, Nakagawa T, Higuchi S, Muramatsu T, Matsushita S, et al. Regional cerebral glucose utilization is modulated by the dosage of apolipoprotein E type 4 allele and alpha1-antichymotrypsin type A allele in Alzheimer's disease. *Neuroreport* 1997; 8(12): 2639-2643.

Imamura T, Ishii K, Sasaki M, Kitagaki H, Yamaji S, Hirono N, et al. Regional cerebral glucose metabolism in dementia with Lewy bodies and Alzheimer's disease: A comparative study using positron emission tomography. *Neuroscience Letters* 1997; 235(1-2): 49-52.

Ishii K, Sasaki M, Kitagaki H, Yamaji S, Sakamoto S, Matsuda K, et al. Reduction of cerebellar glucose metabolism in advanced Alzheimer's disease. *J Nucl Med* 1997; 38(6): 925-928.

Pietrini P, Dani A, Furey ML, Alexander GE, Freo U, Grady CL, et al. Low glucose metabolism during brain stimulation in older Down's syndrome subjects at risk for Alzheimer's disease prior to dementia. *Am J Psychiatry* 1997; 154(8): 1063-1069.

Stein DJ, Buchsbaum MS, Hof PR, Siegel BV, Jr., Shihabuddin L. Greater metabolic rate decreases in hippocampal formation and proisocortex than in neocortex in Alzheimer's disease. *Neuropsychobiology* 1998; 37(1): 10-19.

Vander Borght T, Minoshima S, Giordani B, Foster NL, Frey KA, Berent S, et al. Cerebral metabolic differences in Parkinson's and Alzheimer's diseases matched for dementia severity. *J Nucl Med* 1997; 38(5): 797-802.

Yamaguchi S, Meguro K, Itoh M, Hayasaka C, Shimada M, Yamazaki H, et al. Decreased cortical glucose metabolism correlates with hippocampal atrophy in Alzheimer's disease as shown by MRI and PET. *J Neurol Neurosurg Psychiatry* 1997; 62(6): 596-600.

Excluded Studies

Albin RL, Minoshima S, D'Amato CJ, Frey KA, Kuhl DA, Sima AA. Fluoro-deoxyglucose positron emission tomography in diffuse Lewy body disease. *Neurology* 1996; 47(2): 462-466.

Bergman H, Chertkow H, Wolfson C, Stern J, Rush C, Whitehead V, et al. HM-PAO (CERETEC) SPECT brain scanning in the diagnosis of Alzheimer's disease. *Journal Of the American Geriatrics Society* 1997; 45(1): 15-20.

de Leon MJ, McRae T, Rusinek H, Convit A, De Santi S, Tarshish C, et al. Cortisol reduces hippocampal glucose metabolism in normal elderly, but not in Alzheimer's disease. *J Clin Endocrinol Metab* 1997; 82(10): 3251-3259.

Kondoh Y, Nagata K, Sasaki H, Hatazawa J. Dynamic FDG-PET study in probable Alzheimer's disease. *Ann N Y Acad Sci* 1997; 826: 406-409.

Marcus DL, Freedman ML. Decreased brain glucose metabolism in microvessels from patients with Alzheimer's disease. *Ann N Y Acad Sci* 1997; 826: 248-253.

Mega MS, Chen SS, Thompson PM, Woods RP, Karaca TJ, Tiwari A, et al. Mapping histology to metabolism: coregistration of stained whole-brain sections to premortem PET in Alzheimer's disease. *Neuroimage* 1997; 5(2): 147-153.

Meguro K, Yamaguchi S, Itoh M, Fujiwara T, Yamadori A. Striatal dopamine metabolism correlated with frontotemporal glucose utilization in Alzheimer's disease: a double-tracer PET study. *Neurology* 1997; 49(4): 941-945.

Minoshima S, Giordani B, Berent S, Frey KA, Foster NL, Kuhl DE. Metabolic reduction in the posterior cingulate cortex in very early Alzheimer's disease. *Ann Neurol* 1997; 42(1): 85-94.

XIII. EPILOGUE

On January 28, 1999 the TA Program conducted a final update of the literature by searching the literature published from July 6, 1998 through December 31, 1998 using the same search and appraisal strategies described in Appendix 1. Titles and abstracts of 346 citations were screened. Forty-one were determined to be relevant, and their full text articles were reviewed for potential inclusion in the review.

Thirty articles from the database searches and from end references of initially retrieved articles met inclusion criteria for review. Each included study was classified according to clinical condition and assigned to a diagnostic efficacy level as follows:

Efficacy level*	Head & Neck	Breast	Lung staging	SPN	Colorectal	Alzheimer's
Technical	3	1	6	0	0	7*
Diagnostic accuracy	5	1	5**	1	0	0
Diagnostic thinking			†			
Therapeutic						
Patient outcome				1		
Societal						

*includes 6 overlapping studies from same institution
**includes 3 overlapping studies from same institution (Vansteenkiste, 1998a,b,c)
†diagnostic thinking data and diagnostic accuracy data provided from one study (Vansteenkiste, 1998a)

All of the studies represented are single-site case series. All studies used dedicated PET systems. PET was usually added in the work up to complement anatomic imaging data, and most were retrospective analyses.

As in the main report, recent studies of FDG PET in Alzheimer's disease explore the relationships between regional glucose metabolism and cognitive function and are classified as technical efficacy studies. Several studies of diagnostic PET in oncology met inclusion for review and could be classified at higher levels of diagnostic efficacy. Five studies in lung cancer staging, three from the same institution (Vansteenkiste, 1998a; 1998b; 1998c) were continuations of studies reviewed in the main report with overlapping study populations (Bury, 1998; Weder, 1998)

The diagnostic accuracy studies were further appraised for study quality and content. None of the studies met strict evidence-based medicine criteria for evaluations of diagnostic tests, as the extent of blinding was either not clearly reported or was incomplete. However, two met most of the criteria and had reasonably well reported and designed studies, despite their small sizes (Smith, 1998; Präuer, 1998). All studies used patients with no metastases or with benign diseases as internal controls, and all reported using an objective gold standard. Expanded criteria for methodologic quality of diagnostic accuracy studies used by the American College of Physicians yielded the following quality scores:

Table 20: Methodologic Quality of Diagnostic Accuracy Studies of FDG PET in Selected Cancers

Methodologic Quality Grade*	Head & Neck	Breast	Lung staging	SPN	Colorectal
A					
B					
C		1	3	1	
D	5		2		

Studies received overall quality scores of "D" if the presence of referral bias and methodologic biases related to the association between test interpretation and gold standard diagnosis were not minimized in the study. Studies received a "C" because of small study sizes, incomplete reporting of critical study design elements, and/or a study design that minimized the effect of methodologic biases. Several asserted the potential for PET to directly affect patient management, but this was not systematically assessed in any study.

Two studies were classified either as diagnostic thinking efficacy (Vansteenkiste, 1998a) or patient outcome efficacy (Gambhir, 1998). Expanding on their study reviewed in the main report, Vansteenkiste (1998a) used ROC analysis with PET to calculate optimal accuracy and likelihood ratios (LR) for estimating the probability of nodal metastases in 690 lymph node stations in 68 patients with non-small cell lung cancer. For their study population, a cut-off SUV of 4.40 provided optimal accuracy. Based on these data, the authors suggested that positive LRs for SUVs <3.5 or >4.5 offered high diagnostic value and recommended the following:

- The high negative predictive value of mediastinal CT+PET is sufficient to exclude N2/N3 disease, to exclude malignancy in individual node stations and, therefore, to omit invasive mediastinal staging.
- Despite the high positive predictive value of CT+PET, mediastinoscopy is still advised in patients with a positive mediastinal PET to ensure that no patient with N0 or N1 disease is denied curative resection based on a false positive PET.

LRs can vary with severity of disease in the case mix and positivity criteria (different threshold values) used for interpretation of both imaging tests. There were few benign conditions that may contribute to false positive diagnoses on CT and PET, and only four patients had confirmed N3 disease. The authors calculated positive LRs for both CT and quantitative PET but did not report the probability of nodal metastases before CT. In the absence of knowing the pre-test probability of malignancy, LRs are inconclusive for assessing the impact of the test on diagnosis or treatment planning, and these findings should be interpreted cautiously.

Gambhir (1998) conducted a cost-effectiveness analysis to compare various strategies for diagnosing and managing SPNs. Expanding on a decision analysis by Cummings (1986), the authors incorporated PET into a CT-based strategy for patients with noncalcified solitary pulmonary nodules < 3cm in diameter. They concluded that a CT-plus-PET strategy was the most cost-effective over a wide range of pre-CT probabilities of malignancy (0.12 to 0.69), and offered cost savings over the CT-alone strategy ranging from $91 to $2,200 per patient.

The assumptions upon which the analysis is based may affect the stability of the conclusions. PET sensitivity and specificity estimates were based on data from one abstract and biased estimates from three peer-reviewed studies, which were reviewed in the first VA PET report (Flynn, 1996). The model did not account for the possibility of an indeterminate PET scan. Payment and charge data used in the analysis may not adequately reflect true costs or be sufficiently comprehensive to reflect the true work-up of these patients.

The MDRC agrees with the authors' statement that "this analysis is not a substitute for clinical trials, but a guide to the design of clinical trials." The MDRC does not agree with the authors' statement that "there is significant savings when using a PET-based strategy. This warrants a more widespread dissemination of the technology." Given the preliminary nature of the assumptions, a more widespread dissemination of the technology based on the results of this cost-effectiveness analysis would be premature.

Conclusion

Recent studies from 1998 do not provide conclusive evidence to support the use of PET in the work up of patients with the cancers assessed in this report. Prospective, rigorously designed studies with a sufficient spectrum of patients are needed to assess the incremental value of PET in these patients. The impact of PET results on treatment planning has been alleged, but further research designed to assess impact on treatment management and associated costs is needed. The findings from recent 1998 studies confirm the conclusions and recommendations in the main report.

References

Background

Cummings SR, Lillington GA, Richard RJ. Managing solitary pulmonary nodules. The choice of strategy is a "close call." *Am Rev Respir Dis* 1986; 134(3): 453-460.

Laupacis A, Wells G, Richardson WS, Tugwell P. Users' guides to the medical literature. V. How to use an article about prognosis. Evidence-Based Medicine Working Group. *JAMA* 1994; 272(3): 234-237.

Phelps CE. Good technologies gone bad: how and why the cost-effectiveness of a medical intervention changes for different populations. *Med Decis Making* 1997; 17(1): 107-117.

Head and Neck

Included Studies

Diagnostic Accuracy

Adams S, Baum RP, Stuckensen T, Bitter K, Hor G. Prospective comparison of 18F-FDG PET with conventional imaging modalities (CT, MRI, US) in lymph node staging of head and neck cancer. *Eur J Nucl Med* 1998; 25(9): 1255-1260.

Fischbein NJ, Aassar OS, Caputo GR, Kaplan MJ, Singer MI, Price DC, et al. Clinical utility of positron emission tomography with F-18-fluorodeoxyglucose in detecting residual/recurrent squamous cell carcinoma of the head and neck. *American Journal Of Neuroradiology* 1998; 19(7): 1189-1196.

Kao CH, ChangLai SP, Chieng PU, Yen RF, Yen TC. Detection of recurrent or persistent nasopharyngeal carcinomas after radiotherapy with 18-fluoro-2-deoxyglucose positron emission tomography and comparison with computed tomography. *J Clin Oncol* 1998; 16(11): 3550-3555.

Manolidis S, Donald PJ, Volk P, Pounds TR. The use of positron emission tomography scanning in occult and recurrent head and neck cancer. *Acta Otolaryngol Suppl* 1998; 534(11): 1-11.

Paulus P, Sambon A, Vivegnis D, Hustinx R, Moreau P, Collignon J, et al. 18FDG-PET for the assessment of primary head and neck tumors: clinical, computed tomography, and histopathological correlation in 38 patients. *Laryngoscope* 1998; 108(10): 1578-1583.

Technical Efficacy

Collins BT, Gardner LJ, Verma AK, Lowe VJ, Dunphy FR, Boyd JH. Correlation of fine needle aspiration biopsy and fluoride-18 fluorodeoxyglucose positron emission tomography in the assessment of locally recurrent and metastatic head and neck neoplasia. *Acta Cytol* 1998; 42(6): 1325-1329.

Kole AC, Nieweg OE, Pruim J, Hoekstra HJ, Koops HS, Roodenburg JL, et al. Detection of unknown occult primary tumors using positron emission tomography. *Cancer* 1998; 82(6): 1160-1166.

Uematsu H, Sadato N, Yonekura Y, Tsuchida T, Nakamura S, Sugimoto K, et al. Coregistration of FDG PET and MRI of the head and neck using normal distribution of FDG. *J Nucl Med* 1998; 39(12): 2121-2127.

Excluded Studies

Davis JP, Maisey MN, Chevretton EB. Positron emission tomography--a useful imaging technique for otolaryngology, head and neck surgery? [editorial]. *J Laryngol Otol* 1998; 112(2): 125-127.

Breast Cancer

Included Studies

Diagnostic Accuracy

Smith IC, Ogston KN, Whitford P, Smith FW, Sharp P, Norton M, et al. Staging of the axilla in breast cancer: accurate in vivo assessment using positron emission tomography with 2-(fluorine-18)-fluoro-2-deoxy-D-glucose. *Ann Surg* 1998; 228(2): 220-227.

Technical Efficacy

Cook GJ, Houston S, Rubens R, Maisey MN, Fogelman I. Detection of bone metastases in breast cancer by 18FDG PET: differing metabolic activity in osteoblastic and osteolytic lesions. *J Clin Oncol* 1998; 16(10): 3375-3379.

Excluded Studies

Dehdashti F, Flanagan FL, Mortimer JE, Katzenellenbogen JA, Welch MJ, Siegel BA. Positron emission tomographic assessment of "metabolic flare" to predict response of metastatic breast cancer to antiestrogen therapy. *Eur J Nucl Med* 1999; 26(1): 51-56.

Flanagan FL, Dehdashti F, Siegel BA. PET in breast cancer. *Seminars in Nuclear Medicine* 1998.

Greco M, Agresti R, Giovanazzi R. Impact of the diagnostic methods on the therapeutic strategies. *Q J Nucl Med* 1998; 42(1): 66-80.

Paganelli G, De Cicco C, Cremonesi M, Prisco G, Calza P, Luini A, et al. Optimized sentinel node scintigraphy in breast cancer. *Q J Nucl Med* 1998; 42(1): 49-53.

Lung Cancer

Included Studies

Diagnostic Accuracy

Bury T, Barreto A, Daenen F, Barthelemy N, Ghaye B, Rigo P. Fluorine-18 deoxyglucose positron emission tomography for the detection of bone metastases in patients with non-small cell lung cancer. *Eur J Nucl Med* 1998; 25(9): 1244-1247.

Vansteenkiste JF, Stroobants SG, De Leyn PR, Dupont PJ, Verbeken EK, Potential use of FDG-PET scan after induction chemotherapy in surgically staged IIIa-N-2 non-small-cell lung cancer: A prospective pilot study. *Annals Of Oncology* 1998; 9(11): 1193-1198.

Vansteenkiste JF, Stroobants SG, Dupont PJ, De Leyn PR, De Wever WF, Verbeken EK, et al. FDG-PET scan in potentially operable non-small cell lung cancer: do

anatometabolic PET-CT fusion images improve the localisation of regional lymph node metastases? The Leuven Lung Cancer Group. *Eur J Nucl Med* 1998; 25(11): 1495-1501.

Vansteenkiste JF, Stroobants SG, De Leyn PR, Dupont PJ, Bogaert J, Maes A, et al. Lymph node staging in non-small-cell lung cancer with FDG-PET scan: a prospective study on 690 lymph node stations from 68 patients. *J Clin Oncol* 1998; 16(6): 2142-2149.

Weder W, Schmid RA, Bruchhaus H, Hillinger S, von Schulthess GK, Steinert HC. Detection of extrathoracic metastases by positron emission tomography in lung cancer. *Ann Thorac Surg* 1998; 66(3): 886-892; discussion 892-883.

Technical Efficacy

Ahuja V, Coleman RE, Herndon J, Patz EF, Jr. The prognostic significance of fluorodeoxyglucose positron emission tomography imaging for patients with non-small cell lung carcinoma. *Cancer* 1998; 83(5): 918-924.

Gupta N, Gill H, Graeber G, Bishop H, Hurst J, Stephens T. Dynamic positron emission tomography with F-18 fluorodeoxyglucose imaging in differentiation of benign from malignant lung/mediastinal lesions. *Chest* 1998; 114(4): 1105-1111.

Higashi K, Ueda Y, Seki H, Yuasa K, Oguchi M, Noguchi T, et al. Fluorine-18-FDG PET imaging is negative in bronchioloalveolar lung carcinoma. *J Nucl Med* 1998; 39(6): 1016-1020.

Imran MB, Kubota K, Yamada S, Fukuda H, Yamada K, Fujiwara T, et al. Lesion-to-background ratio in nonattenuation-corrected whole-body FDG PET images. *J Nucl Med* 1998; 39(7): 1219-1223.

Kiffer JD, Berlangieri SU, Scott AM, Quong G, Feigen M, Schumer W, et al. The contribution of 18F-fluoro-2-deoxy-glucose positron emission tomographic imaging to radiotherapy planning in lung cancer. *Lung Cancer* 1998; 19(3): 167-177.

Kutlu CA, Pastorino U, Maisey M, Goldstraw P. Selective use of PET scan in the preoperative staging of NSCLC. *Lung Cancer* 1998; 21(3): 177-184.

Excluded Studies

Erasmus JJ, McAdams HP, Patz EF, Jr., Coleman RE, Ahuja V, Goodman PC. Evaluation of primary pulmonary carcinoid tumors using FDG PET. *AJR Am J Roentgenol* 1998; 170(5): 1369-1373.

Laymon CM, Turkington TG, Coleman RE. Attenuation effects in gamma-camera coincidence imaging. *IEEE Transactions on Nuclear Science* 1998; 45(6): 3115-3121.

Lowe VJ, Naunheim KS. Positron emission tomography in lung cancer. *Ann Thorac Surg* 1998; 65(6): 1821-1829.

SPN

Included Studies

Diagnostic Accuracy

Prauer HW, Weber WA, Romer W, Treumann T, Ziegler SI, Schwaiger M. Controlled prospective study of positron emission tomography using the glucose analogue [18f] fluorodeoxyglucose in the evaluation of pulmonary nodules. *Br J Surg* 1998; 85(11): 1506-1511.

Patient Outcome Efficacy

Gambhir SS, Shepherd JE, Shah BD, Hart E, Hoh CK, Valk PE, et al. Analytical decision model for the cost-effective management of solitary pulmonary nodules. *J Clin Oncol* 1998; 16(6): 2113-2125.

Colorectal Cancer

Excluded Studies

Miraldi F, Vesselle H, Faulhaber PF, Adler LP, Leisure GP. Elimination of artifactual accumulation of FDG in PET imaging of colorectal cancer. *Clin Nucl Med* 1998; 23(1): 3-7.

Alzheimer's Disease

Included Studies

Technical Efficacy

Hirono N, Mori E, Ishii K, Imamura T, Shimomura T, Tanimukai S, et al. Regional metabolism: associations with dyscalculia in Alzheimer's disease. *J Neurol Neurosurg Psychiatry* 1998; 65(6): 913-916.

Hirono N, Mori E, Ishii K, Ikejiri Y, Imamura T, Shimomura T, et al. Hypofunction in the posterior cingulate gyrus correlates with disorientation for time and place in Alzheimer's disease. *J Neurol Neurosurg Psychiatry* 1998; 64(4): 552-554.

Hirono N, Mori E, Ishii K, Ikejiri Y, Imamura T, Shimomura T, et al. Regional hypometabolism related to language disturbance in Alzheimer's disease. *Dement Geriatr Cogn Disord* 1998; 9(2): 68-73.

Hirono N, Mori E, Ishii K, Ikejiri Y, Imamura T, Shimomura T, et al. Frontal lobe hypometabolism and depression in Alzheimer's disease. *Neurology* 1998; 50(2): 380-383.

Ibanez V, Pietrini P, Alexander GE, Furey ML, Teichberg D, Rajapakse JC, et al. Regional glucose metabolic abnormalities are not the result of atrophy in Alzheimer's disease. *Neurology* 1998; 50(6): 1585-1593.

Ishii K, Sasaki M, Yamaji S, Sakamoto S, Kitagaki H, Mori E. Relatively preserved hippocampal glucose metabolism in mild Alzheimer's disease. *Dement Geriatr Cogn Disord* 1998; 9(6): 317-322.

Ishii K, Sakamoto S, Sasaki M, Kitagaki H, Yamaji S, Hashimoto M, et al. Cerebral glucose metabolism in patients with frontotemporal dementia. *J Nucl Med* 1998; 39(11): 1875-1878.

Excluded Studies

Alavi A, Clark C, Fazekas F. Cerebral ischemia and Alzheimer's disease: critical role of PET and implications for therapeutic intervention. *J Nucl Med* 1998; 39(8): 1363-1365.

Frey KA, Minoshima S, Kuhl DE. Neurochemical imaging of Alzheimer's disease and other degenerative dementias. *Q J Nucl Med* 1998; 42(3): 166-178.

Small GW, Leiter F. Neuroimaging for diagnosis of dementia. *J Clin Psychiatry* 1998; 59(11): 4-7.

XIV. APPENDIX 1

Methods for the Systematic Review

The MDRC performed a ***systematic review*** of the published literature to address the diagnostic efficacy of PET in selected cancer applications and Alzheimer's disease. A systematic review differs from a traditional narrative literature review in that it uses a rigorous scientific approach to limit bias and to improve the accuracy of conclusions based on the available data (Guyatt, 1995). A systematic review addresses a focused clinical question, uses appropriate and explicit criteria to select studies for inclusion, conducts a comprehensive search, and appraises the validity of the individual studies in a reproducible manner.

Consistent with established methods for conducting a systematic review, the MDRC developed criteria to select studies for inclusion, conducted a comprehensive search, and appraised the validity of the individual studies in a reproducible fashion using the analytic frameworks presented below.

Search Strategy

An update of the literature was carried out by thoroughly searching the literature published from September 1996 through July 6, 1998. MEDLINE®, HealthSTAR®, EMBASE®, Current Contents®, and BIOSIS® were searched using a range of descriptors: tomography, emission computed; positron emission tomography; gamma camera; PET; and other synonyms. These were combined with the descriptors for Alzheimer's, colorectal neoplasms, breast neoplasms, head and neck neoplasms, and lung neoplasms. Over 400 citations were retrieved.

Inclusion Criteria

All published studies included in this report met the following inclusion criteria:

- English language articles reporting primary data and published in a peer review journal (not abstracts);
- studies $\geq$ 12 human subjects (not animal studies) with the disease of interest;
- studies using positron emission transverse tomography or positron emission coincidence imaging;
- studies using the radiopharmaceutical 2-[^{18}F]fluoro-2-D-glucose (FDG);
- study not duplicated or superseded by later study with the same purpose from the same institution; and
- study design and methods clearly described (i.e. sufficient information to judge comparability of case and control groups, details of imaging protocol, whether visual or quantitative analysis of PET data used, or type of PET quantitative data analysis used).

Methodologic standards for studies

The purpose of appraising the literature using clearly defined methodologic criteria is to ensure that studies are evaluated in a consistent, reproducible manner, and that studies included in the report conform to established scientific standards. Studies reviewed for possible inclusion in this report were classified according to the strength of the evidence they provided, and the strongest available evidence for each application was summarized. The strength of a study is based on the overall research design and on the quality of the implementation and analysis. The methodologic standards and the types of studies to which they were applied are summarized below. The standards are also discussed in the MDRC report *Assessing Diagnostic Technologies* (Flynn, 1996).

1. Assign to level of diagnostic efficacy hierarchy

Accurate estimation of the characteristics of a diagnostic test is one of the early steps in the assessment of that test. However, a complete assessment requires further research.

Fryback and Thornbury (1991) note that the localized view of the goal of diagnostic radiology would be that it provides the best images and the most accurate diagnoses possible. A more global view recognizes diagnostic radiology as part of a larger system of medical care whose goal is to treat patients effectively and efficiently. Viewed in this larger context, even high-quality images may not contribute to improved care in some instances, and images of lesser quality may be of great value in others. The point of the systematic view may be to examine the ultimate value or benefit that is derived from any particular diagnostic examination.

Fryback and Thornbury (1991; 1992) present the most recent manifestation of an evolving hierarchical model for assessing the efficacy of diagnostic imaging procedures. Their model, with a list of the types of measures that appear in the literature at each level in the hierarchy, is presented in the next table. The table progresses from the micro, or local level, at which the concern is the physical imaging process itself, to the societal efficacy level. The model stipulates that for a procedure to be efficacious at a higher level in the hierarchy it must be efficacious at the lower levels, but the reverse is not true; this asymmetry is often lost in research reports at Levels 1 and 2. Using this model, it is possible to follow the development of a diagnostic technology, and to align current research efforts with a particular level of development.

A Hierarchical Model of Efficacy for Diagnostic Imaging

Level	*Typical Measures of Analysis*	*Comments*
1. Technical efficacy	• Resolution of line pairs • Modulation transfer function • Gray scale range • Amount of mottle • Sharpness	• Physical parameters describing technical imaging quality
2. Diagnostic accuracy efficacy	• Yield of abnormal or normal diagnoses in a case series • Diagnostic accuracy (% of correct diagnoses in case series) • Sensitivity and specificity in a defined clinical setting • Measures of ROC curve height *(d')* or area under the curve A_z	• Joint function of images and observer • Also a function of clinician who requests diagnostic procedure, since selection controls specificity of test in clinical practice and sensitivity to the extent that it varies with the spectrum of the disease
3. Diagnostic thinking efficacy	• Number (%) of cases in series in which image judged "helpful" to making diagnosis • Entropy change in differential diagnosis probability distribution • Difference in clinicians' subjectively estimated diagnosis probabilities pre-to posttest information • Empirical subjective log-likelihood ratio for test positive and negative in a case series	• Inducing change in clinicians' diagnostic thinking is a necessary prerequisite to impact on patients. • Clinicians may value results which reassure them, but which do not change treatment decisions. • Empirical methods to measure change in pretest diagnostic probabilities assumed by clinicians are probably best for determining the absence of diagnostic thinking efficacy, rather than estimating the magnitude of change in diagnostic thinking due to imaging information. • Imaging examination result may influence clinician's diagnostic thinking, but has no impact on patient treatment.
4. Therapeutic efficacy	• Number (%) of times images judged helpful in planning management of the patient in a case series • % of times medical procedure avoided due to image information • % of times therapy planned pretest changed after imaging information was obtained (retrospectively inferred from patient records) • % of times clinicians prospectively stated therapeutic choices changed after test information	• In situations where RCTs of decision making with and without the imaging information cannot be performed ethically or because of the momentum for using a particular procedure, asking Level 4 questions may be only efficacy study possible. • Integrating negative information about a test from Level 3 and 4 studies may help to direct clinical use away from imaging tests that are not useful or have been supplanted by other tests.
5. Patient outcome efficacy	• % of patients improved with test compared with no test • Morbidity (or procedures) avoided with test • Change in quality-adjusted life expectancy • Expected value of test information in QALYS • Cost per QALY saved with imaging information	• Definitive answer re efficacy with respect to patient outcome requires RCT (involving withholding test from some patients). • RCTs may be associated with formidable statistical, empirical, and ethical problems and are justified only in carefully selected situations. • Weaker evidence may be derived from case control studies or case series. • Independent contribution of imaging to patient outcome may be small, requiring very large sample sizes. • Decision analytic approach can be alternative to RCT, but the analyses may suffer from the same biases as their secondary data sources. • Decision analyses can highlight critical pieces of information and guide future studies.
6. Societal efficacy	• Cost-benefit analysis from societal viewpoint • Cost-effectiveness analysis from societal viewpoint • Cost-utility analysis from societal viewpoint	• Economic evaluations of evolving technologies do not provide definitive answers, since values and judgments play a significant role in interpretation of results. • Cost utility analyses imply at least Level 5 efficacy data or models.

Adapted from Fryback and Thornbury, 1991

Abbreviations: RCT, randomized clinical trial
ROC, receiver operating characteristics
QALY, quality adjusted life year

2. Assess the quality of individual studies of diagnostic tests

Criteria for assessing the quality of a diagnostic test evaluation have been defined for use in evidence-based medicine (Haynes and Sackett, 1995). These criteria, listed below, will be applied to individual studies in the report. If the criteria are not met, the study will generally be considered insufficiently rigorous to provide the basis for patient care decisions. However, such studies often provide useful information on the technical characteristics of a diagnostic test, or may provide information necessary to subsequent diagnostic accuracy studies.

Evidence-based medicine criteria for evaluating studies of diagnosis

- Clearly identified comparison groups, of which ≥ 1is free of the target disorder.
- Either an objective diagnostic standard (e.g., a machine-produced laboratory result) or a contemporary clinical diagnostic standard (e.g., a venogram for deep venous thrombosis) with demonstrably reproducible criteria for any subjectively interpreted component (e.g:, report of better-than-chance agreement among interpreters).
- Interpretation of the test without knowledge of the diagnostic standard result (no test review bias).
- Interpretation of the diagnostic standard without knowledge of the test result (no diagnostic review bias).

Haynes and Sackett, 1995

Documentation of test accuracy does not translate into documentation that the test is clinically useful. Sensitivity and specificity, while not as dependent on prevalence of disease as predictive values, can be biased by differences in patient mix in the study population and the patients on whom the test will be used in clinical practice (Sackett et al. 1991). A published study that does not supply valid information needed to calculate posttest probability of disease (i.e., predictive values or likelihood ratios) would not assist clinicians in interpreting its results, or taking action based on those results.

Evidence-based criteria provide a broad quality screen for clinicians who are contemplating using a test in their own patients. A somewhat more detailed set of quality criteria, that expand on those of evidence-based medicine, have been used by the American College of Physicians in evaluations of the literature on magnetic resonance imaging (Kent et al., 1994; Kent and Larson, 1992; Kent and Larson, 1988). These criteria were applied to studies of **diagnostic accuracy and diagnostic thinking efficacy**.

Methodologic quality of diagnostic accuracy studies

Grade	*Criteria*
A	**Studies with broad generalizability to a variety of patients and no significant flaws in research methods** • ≥ 35 patients with disease and ≥ 35 patients without disease (since such numbers yield 95% CIs whose lower bound excludes 0.90 if Se = 1) • patients drawn from a clinically relevant sample (not filtered to include only severe disease) whose clinical symptoms are completely described • diagnoses defined by an appropriate reference standard • PET studies technically of high quality and evaluated independently of the reference diagnosis
B	**Studies with a narrower spectrum of generalizability, and with only a few flaws that are well described (and impact on conclusions can be assessed)** • ≥ 35 cases with and without disease • more limited spectrum of patients, typically reflecting referral bias of university centers (more severe illness) • free of other methods flaws that promote interaction between test result and disease determination • prospective study still required
C	**Studies with several flaws in methods** • small sample sizes • incomplete reporting • retrospective studies of diagnostic accuracy
D	**Studies with multiple flaws in methods** • no credible reference standard for diagnosis • test result and determination of final diagnosis not independent (diagnostic review and/or test review bias) • source of patient cohort could not be determined or was obviously influenced by the test result (work-up bias) • opinions without substantiating data

Studies that assess the efficacy of diagnostic tests, particularly estimates of sensitivity and specificity, are susceptible to a variety of biases (Begg, 1987). Thornbury et al. (1991) described five aspects of research methodology that may influence accuracy estimates. ***Insufficient sample size*** may result in failure to detect differences between imaging modalities, if in fact they do exist, and may provide imprecise estimates of imaging accuracy.

Differences among patient populations in the spectrum of disease presentation (case mix) and severity result in ***referral bias***. The spectrum of patients needed to assess a diagnostic test will depend on the clinical situation. For example, at initial presentation of abnormality the spectrum should also include patients with no abnormality as well as patients with abnormalities that may be confused with malignancy. For diagnosing recurrent disease the spectrum should include patients with recurrence, patients with no recurrence, and patients with treatment changes that may be confused with malignancy on testing. A wider spectrum of patients would be needed to assess a test when there is a high prevalence of benign conditions (eg. SPN), whereas a test could be assessed in a narrower spectrum of patients with higher prevalence cancers.

Biases related to the appropriate use of a diagnostic reference standard are ***work up bias, test review bias,*** and ***diagnostic review bias***. Presence of referral bias and reference standard methodologic biases result in overestimation of true positive rates and underestimation of false positive and negative rates.

Considerable activity in the diagnostic testing literature is focusing on developing study designs and analytic techniques to correct for, or minimize the effect of, these biases. Some of the more common methods for limiting their influence on diagnostic accuracy estimates are presented below:

Biases in Studies of Diagnostic Imaging Tests

Type of bias	***Techniques to minimize bias***	***Comments***
Referral/spectrum *the influence of spectrum and severity of disease (case mix) on test characteristics*	• referral sources from a variety of medical practice settings in which potential patient subjects are first encountered • clearly defined referral • define patient groups based on physician's pre-test probability estimate of disease • adequate subgroup sizes	⇒ gives sufficient number and mix of patients needed to define predictive values ⇒ can determine generalizability of study results to own population ⇒ allows subgroup analysis of diagnostic accuracy estimates
Work-up/verification • *results from imaging test determine the choice of patient verified by the gold standard, or* • *study is restricted to biopsy verified cases*	• all patients have all competing tests • prospective study in which all patients receive definitive verification of disease status • sufficient follow-up time • retrospective adjustments • algebraic correction involving regression of empirical disease frequencies against the probability of disease as determined in a predictive model	⇒ magnitude of the bias is related to association between selection for verification and test result ⇒ maximizes diagnostic certainty ⇒ require test results and covariate data from the source population and verified sample
Test review *imaging test interpretation is not independent of final diagnosis, clinical information or results of comparison test*	• randomized, blinded, independent interpretation of imaging test • readings with and without clinical information • allow sufficient time between readings • standardize diagnostic terms and degrees of abnormality • document impact of uninterpretable results • use multiple readers and determine interobserver variability and methods for resolving differences	⇒ can determine effect of clinical information on diagnostic probability estimates ⇒ frequency of uninterpretability is an important consideration in the cost-effectiveness of a test
Diagnostic review/incorporation *gold standard diagnosis is not independent of imaging test results*	• extensive nodal sampling regardless of imaging results • expert interdisciplinary panel to review patient information and revise diagnostic and probability estimates incrementally	⇒ blinding practitioner to imaging may be impractical , but effect of bias can be minimized ⇒ panel process optimizes the final diagnosis in cases in which biopsy result is and is not available

Adapted from Begg (1987), Thornbury et al. (1991), and Webb et al. (1991)

3. Evaluate the strength of the evidence supporting a causal link between the use of the technology and improved outcomes of care

The third analytic framework for the literature review will rank the available evidence for the degree to which it supports a causal link between the use of the technology and improved outcomes. Recommendations about the use of a technology should be linked to the quality of the available evidence, with the strength of the evidence dependent on the quality of the available evidence.

Several models for this framework exist that are based on well-established scientific principles of study design. Flynn (1996) used the model below by Cook (1992) to summarize the relative strengths associated with various study designs and to rank the

persuasiveness of their findings between the use of the technology and the outcome of interest:

Classifications of study designs and levels of evidence
(when high quality meta analyses/overviews are not available)

Level	*Description*
I	**Randomized trials with low false-positive (alpha) and low false-negative (beta) errors (high power)** • positive trial with statistically significant treatment effect (low alpha error) • negative trial that was large enough to exclude the possibility of a clinically important benefit (low beta error/high power; i.e. had a narrow confidence interval around the treatment effect, the lower end of which was greater than the minimum clinically important benefit) • meta analysis can be used to generate a pooled estimate of treatment efficacy across all high quality, relevant studies and can reveal any inconsistencies in results
II	**Randomized trials with high false-positive (alpha) and/or high false negative (beta) errors (low power)** • trial with interesting positive trend that is not statistically significant (high alpha error) • negative trial but possibility of a clinically important benefit (high beta error/low power; i.e. very wide confidence intervals around the treatment effect) • small positive trials with wide confidence intervals around the treatment effect, making it difficult to judge the magnitude of the effect • when Level II studies are pooled (through quantitative meta analysis), the aggregate effects may provide Level I evidence
III	**Nonrandomized concurrent cohort comparisons between contemporaneous patients who did and did not (through refusal, noncompliance, contraindication, local practice, oversight, etc.) receive treatment** • results subject to biases • Level III data can be subjected to meta analysis, but the result would not shift these data to another Level, and is not usually recommended
IV	**Nonrandomized historical cohort comparison between current patients who did receive treatment (as a result of local policy) and former patients (from the same institution or from the literature) who did not (since at another time or in another institution different treatment policies prevailed)** • *results subject to biases, including those that result from inappropriate comparisons over time and space*
V	**Case series without control subjects** • *may contain useful information about clinical course and prognosis but can only hint at efficacy*

Source: Cook et al. (1992)

Ibrahim (1987) presented a similar framework to display the continuum of study designs and their causal implications.

Continuum of study designs and their causal implications

*Level**	*Study design*	*Inference/strength of evidence*
I	Randomized controlled trials (RCT) Community randomized trials Systematic reviews of RCTs	Firm
II	Prospective cohort	Moderately firm
III	Before-after with controls Historical cohort	Highly suggestive
IV	Case-control	Moderately suggestive
V	Time series Ecologic correlations Cross-sectional	Suggestive
VI	Anecdote Clinical hunches Case history	Speculative

Adapted from Ibrahim, (1985).
*For simplicity, the numerical order was reversed for this review to align with the levels found in the previous table.

Levels IV, V, and VI are observational (nonexperimental) studies. Observational studies are subject to many forms of bias, which can diminish the accuracy of their findings. They do not provide strong evidence linking interventions with the observed outcomes; however, they can be useful for generating hypotheses for future research. Levels II and III are considered quasi-experimental designs. They are commonly used in health care and provide stronger evidence than can be obtained from observational studies. Level I studies are true experimental studies and provide the most persuasive evidence for linking interventions with the observed outcomes.

Both frameworks will be used to appraise the strength of the evidence that links use of PET with desired outcomes, particularly to effect change in diagnosis and treatment management.

XV. APPENDIX 2

Models of High Quality Efficacy Studies of Diagnostic Imaging Technologies

Study	Highlights of study design
Mushlin (1993) MRI vs. CT in patients with suspected multiple sclerosis	▪ multi-site study with well-defined referral sources and filters, included patients with an uncertain diagnosis, representing those in whom the tests might be used ▪ sufficient sample size ▪ all patients receive all tests under evaluation ▪ independent, blinded image interpretation ▪ varying degrees of abnormality on the images were noted to permit calculation of receiver-operating characteristics (ROC) analysis and likelihood ratios for summary comparisons ▪ sufficient follow-up to permit reasonable diagnostic certainty ▪ use of technology that is representative of what is available and widely used in most medical communities
Stark (1987) MRI vs. CT in patients diagnosed with liver metastases	▪ included patients with and without disease, and patients with benign disease commonly confused with metastases ▪ independent, blinded interpretation of each test and gold standard diagnosis ▪ used ROC analysis to permit comparison of tests over a range of confidence levels and diagnostic thresholds
Webb (1991) MRI vs. CT to determine extent of disease in patients with non-small cell bronchogenic carcinoma	▪ multi-site study with a detailed description of the filter through which patients entering into the study were passed (to reduce referral bias) ▪ data dichotomized to analyze lower and advanced stage disease ▪ blinded, independent interpretation of test results and interobserver variability calculated ▪ independent pathologic data available for all patients analyzed ▪ use of standardized forms for data analysis ▪ extensive nodal sampling not limited to abnormal results on imaging ▪ assessed influence of sampling procedure on results
Rifkin (1990) MRI vs. transrectal ultrasonography to determine extent of disease in surgical candidates with probable localized prostate cancer	▪ large consecutive case series and a multi-site study ▪ used standardized forms for data analysis ▪ blinded, independent interpretation of test results using a five-point grading scale appropriate for ROC analysis ▪ lesions identified on diagnostic imaging were matched with pathological findings using a computer algorithm
Thornbury (1993) MRI vs. plain CT vs. CT myelography in patients with acute low-back pain and radicular pain	▪ patients with a range of probability of disease were included, based on initial clinical diagnosis before imaging ▪ sample size sufficient to provide reasonable statistical power ▪ MRI and one of the two CT tests were performed in all patients ▪ follow-up time sufficient to permit reasonable diagnostic certainty ▪ randomized, unpaired blinded interpretation of all tests ▪ use of an expert interdisciplinary panel to determine true diagnosis ▪ data collection provided information for use in a cost-effectiveness analysis
Zerhouni (1996) CT vs. MRI in staging colorectal carcinoma	▪ multi-institutional study with well defined and described study population and referral filter ▪ all subjects received either histopathologic, follow-up verification, or corrected for work up bias using technique of Gray et al (1984) ▪ well-defined positivity criteria ▪ blind, independent interpretation of each test compared to joint interpretation ▪ standardized surgical form for data collection of extent of disease for gold standard determination ▪ extensive quality control procedures to monitor data collection and compliance ▪ data analysis stratified based on pre-test knowledge of disease

XVI. APPENDIX 3

Active Funded Research at VHA PET Facilities as of October 1, 1998

Site	*Study Title/Number*	*Funding/Sponsor*	*Start/Completion Dates*
St. Louis	18F-Fluorodeoxyglucose (FDG) Positron Emission Tomography (PET) Imaging in the Management of Patients with Solitary Pulmonary Nodules (CSP 27)	$2,306,632 – funded by VHA ORD Cooperative Studies Program	1998/5 year project
West Haven	Neurobehavioral Correlates of Mental Stress Ischemia (R01 HL59619-01A1)	$1,300,000 - NIH National Heart, Lung and Blood Institute	1998-2001
	Psychological, CNS and Myocardial Mechanisms in Mental Stress Ischemia	$374,000 - Merit Review Award	1998-2000
	CNS Correlates of Mental Stress Induced Myocardial Ischemia in Women	$100,000 - Charles A. Dana Foundation, Neuroscience Research Program on Brain-Body Interaction	Starts 1998, duration 3 years
	Study to Determine the Effect of Atorvastatin on the Progression of Atherosclerosis	$210,000 - Parke-Davis Pharmaceutical Research	1998-1999 (6-month project)
	Impact of PET on Patient Care Algorithm	$50,000 - funded by VHA Office of Patient Care Services	1998-1999
	PET Measurement of Cerebral Blood Flow Correlates of Memory in Posttraumatic Stress Disorder	$421,094 - Career Development Award	10/1/97-9/30/00
	PET Measurement of Hippocampal Function (Memory) in Depression	$56,500 - National Alliance for Research in Schizophrenia and Depression, Young Investigator Award	7/1/97-6/30/99
	Cerebral Metabolic Correlates of AMPT-induced Depressive Relapse	$306,000	7/1/96-6/30/99
	PET Measurement of Cerebral Blood Flow Correlates of Traumatic Memory in PTSD	$850,000 per year	Continuing Renewal
	Hippocampal Function in Gulf War Combat-related PTSD	$299,400	7/1/98-6/30/02
	Hippocampus in Women with Abuse-related PTSD	$967,000 - NIMH	1/1/99-12/30/02
	PET Measurement of Benzodiazepine Receptor in Anxiety	$850,000 per year - National Center for Posttraumatic Stress Disorder Grant	Continuing Renewal
	PET Measurement of Cerebral Blood Flow Correlates of Conditioned Fear	$850,000 - National Center for Posttraumatic Stress Disorder Grant	Continuing Renewal
	Transmyocardial Laser Revascularization in Chronic Canine Model of Ischemia	$80,000 - United States Surgical Corp.	10/96-12/98
	Dynamic SPECT BMIPP Imaging comparison with Perfusion and FDG Accumulation	$149,400 - Nihon Mediphysics	3/96-6/99
	PET Neuroreceptor Imaging (Serotonin-2A and Serotonin-1A)	• $100,000 - National Institute of Mental Health Clinical Research Center • $55,000 - VA Schizophrenia Research Center	• 10/1/96-9/30/01 • 10/1/94-12/31/99
Minneapolis	Quantitative Assessment of Functional Connectivity in the Hereditary Ataxias (PO1 NS33718)	$87,720 - Sponsored by NIH/NINDS	1/1/95-12/31/99
	Spatial and Temporal Patterns in Functional Neuroimaging (P20 MH57180)	$1,113,418 - Sponsored by NIH	$9/30/96-9/29/01
	Correlation of Cholinergic Reserve and Cognitive Function with Positron Emission Tomography (LOI-96-001)	$106,446 - With the Alzheimer's Association	10/15/96-10/14/98
	Motor Cortex and the Control of Dynamic Force	$75,500 - Merit Review Award by VA	11/1/96-10/30/01
	Functional MRI of Human Motor Cortex (5RO1 NS32437-02)	$150,178 - Sponsored by NIH.NINDS	4/1/95-3/30/98

Site	*Study Title/Number*	*Funding/Sponsor*	*Start/Completion Dates*
	Functional reorganization with cortical motor areas	$33,000 - Funded by Charles A. Dana Foundation	1/1/95-12/31/98
	Neural mechanisms of drawing movements under different load conditions	$73,300 - Funded by the National Science Foundation	4/1/97-3/31/00
	Optimizing 3D Iterative Reconstructions for PET (R29 NS33721)	$71,369 - Sponsored by NINDS	12/1/94-11/30/99
	Regional FDG Uptake in Stunned vs Hibernating Myocardium (R29 HL52157)	$78,012 - Sponsored by NIH/NHLBI	2/1/96-1/31/01
	Quantitative Magnetic Resonance Assessment of Microvascular Dysfunction (R01 HL58876)	$194,475 - Sponsored by NIH	9/1/97-8/31/00
	Functional Anatomy of Human Cognition	$99,000 - VA Merit Review Award	10/1/95-9/30/99
	PET studies of Lexical Processing in Schizophrenia	$30,000 - Young Investigator Award from NARSAD	7/1/96-6/30/98
	Lexical Processing in the Differential Diagnosis of Mania from Depression	$12,151 - Funded by Minnesota Medical Foundation	4/1/98-3/31/99
	PET Imaging of Hunger and Satiety	$38,704 - Minnesota Obesity Center	8/1/96-7/31/97
	Hippocampal and Memory Dysfunction in Normal Aging	$29,700 - Alzheimer's Disease Association	7/1/96-12/31/97
Buffalo	Positron Emission Tomographic Study of Tinnitus and Auditory Plasticity	$46,125 - American Tinnitus Association	6/1/96-10/30/97
	Positron Emission Tomographic Studies of the Auditory System	Jane H. Cummings Foundation	6/1/97
	A Comparison of Cerebral Blood Flow in Migraineurs During Headache, Headache Free, and Treatment Periods	$114,300 - Department of Defense	Start 7/1/95 duration of two years
	PET Studies of Temporal Mandibular Joint Pain	$20,000 - State University of New York	Start 6./1/97 duration of one year
	Glucose Transport in Stunned and Hibernating Myocardium	$105,000 - New York State Affiliate, American Heart Association	7/1/97-6/30/00
	Chronic Alterations in Glucose Transport in Hibernating and Stunned Myocardium	$277,800 - American Heart Association	7/1/96-6/30/01
	Chronic Adaptations to Myocardial Ischemia	$1,120,447 - NIH and National Heart Blood and Lung Institute	
	PET Studies of Tinnitus and Hearing Loss	$1,272,652 - NIH and National Institute on Deafness and Communicative Disorders	Starts 1/98 duration of 5 years
	PET Imaging subproject	$48,240 - NIH and National Institute of Aging	
San Antonio	Fluoxetine Effects on Mood, Cognition & Metabolism	$507,446 - National Institute of Mental Health	Ends 8/31/98
	Anterior Cingulate Metabolism in Depression	$99,992 - NARSAD	Ends 9/14/98
	Multimethodological Studies in Cognitive Neuroscience	$85,440 - Blue List Neurobiology	Ends 12/31/98
	The Role of PET in Conjunction with Maximal Exercise Stress in Assessment of Chronic Stable Coronary Artery Disease	$25,000 - Dupont Pharmaceuticals, Inc.	Ends 01/01/99
	The Effects of Prozac Treatment on Mood, Cognition and Brain Glucose Metabolism in Patients with Primary Unipolar Depression	$49,940 - Eli Lilly and Co.	Ends 01/01/99
	PET/TMS Mapping of the Neural Circuitry of Developmental Stuttering	$100,000 - Dan Foundation	Ends 12/31/99
	Interactive Effects of Mood and Cognition Challenges on Anterior Cingulate Function in Remitted Depression	$60,000 - NARSAD Young Investigator Award	Ends 06/30/00
	Hunger for Air Study	$140,000 - Mathers Foundation	Ends 06/30/00
	Investigating the Neural Bases of Chronic Stuttering	$435,231 - NIH	Ends 11/30/01
Indianapolis	Role of Hemodynamics in In-Vivo Insulin Resistance (R01 DK 42469)	$207,453 - sponsored by NIH	7/1/95-6/30/00

Site	*Study Title/Number*	*Funding/Sponsor*	*Start/Completion Dates*
	SCOR in Sudden Cardiac Death (P50 DK 52323)	$258,274 - sponsored by NIH	1/1/95-12/31/99
	PET Imaging in the Surgical Management of Melanoma	$127,918 - Sponsored by NIH	4/1/97-3/31/01
Pittsburgh	Effect of NIDDM on Glucose Transport into Skeletal Muscle	Not available	1998
	The Effect of Troglitazone, Metformin, and Sulfonylurea on Insulin-stimulated Glucose Transport and Phosphorylation, Oxidative Enzyme Capacity and Muscle Composition in NIDDM	Not available	Ongoing
	Echocardiographic Assessment of Myocardial Viability in patients with Impaired Left Ventricular Function	Not available	1998
	The Role of PET Scanning in Staging the Patient with Intrathoracic Malignancies: Non-small Cell Lung Cancer	Not available	1998
West Los Angeles	Pre-frontal Dysfunction in Frontal Lobe Epilepsy	VA Merit Review	
	Psychiatric and Behavioral Disturbances in Alzheimer's Disease	NIMH	
	The Study of Cognitive Processes in Normal Individuals: Activation Studies of the Normal Human Frontal Lobe	Mathers Charitable Foundation	
	Effect of Smoking on Coronary Blood Flow Reserve and Attenuation Effect on Coronary Vasodilator Response of Nitroglycerine	California Tobacco Institute	
	Perception and Modulation of Visceral Sensations	NIH and Astra Pharmaceuticals	
	Central Nervous System Processing of Sensory Information in Irritable Bowel Syndrome (IBS) and Fibromyalgia	CAP	
	Functional Electrical Stimulation on Spinal Cord Injured Patients	VA PM&R R&D	
	Evaluation of Limb Blood Flow with ^{15}O-H_20 PET	VA PM&R R&D	
	^{15}O-H_20 Scanning in Schizophrenia; Assessing Training-Related Improvement	Stanley Foundation and/or NARSAD Young Investigator Award	
	Brain Metabolic Changes with Cigarette Craving	California Tobacco institute	
	PET-FDG Imaging of Opioid Dependent Subjects	NIDA	
	Pathogenesis of Symptomatic vs. Silent Myocardial Ischemia	Not Available	
	Assessment of Myocardial Viability Using PET to Determine Benefit for Revascularization	Not Available	
Ann Arbor	Michigan Alzheimer's Disease Research Center	NIA	
	PET study of Biochemistry and Metabolism of CNS	NIND&S	
	Forebrain Mechanisms of Pain and Analgesia	$300,000 - VA Merit Award	
	Forebrain Responses to Chronic Pain and Its Treatment	NICH&HD	
	Concomitant Chemotherapy and Radiation for Organ Preservation in Patients with Advanced (Stage III, IV) Laryngeal Cancer	University of Mich./VA	
	Combined Hormone Replacement Therapy and Myocardial Blood Flow	VA	

Site	*Study Title/Number*	*Funding/Sponsor*	*Start/Completion Dates*
	Effect of Conjugated Equine Estrogen and Micronized Progesterone on Coronary Artery Endothelial Function as Assessed by Positron Emission Tomography	VA	
	Limbic Blood Flow & Opiate Receptor PET in Posttraumatic Stress Disorder	$288,500 - VA Merit Award	
	Paroxysmal Dystonia-Choreoathetosis	NIND&S	
	PET Studies of Dopaminergic Neurons in Chronic Severe Alcoholism	NIAA&A	
	Metabolic Imaging of Renal Masses with Positron Emission Tomography	VA	
	Metabolic Imaging of Pancreatic Disease with Positron Emission Tomography	University of Mich./VA	
	Imaging of Intermediary Metabolism in Neoplasia using C-11 Acetate PET	VA	

XVII. APPENDIX 4: Data Abstraction Tables of Included Diagnostic Efficacy Studies of FDG-PET in Cancer

Diagnostic Efficacy of FDG PET in Head and Neck Cancer

Study	Patients/Methods	Results/Comments
Lowe et al. (1997) *(St. Louis Health Sciences Center, Missouri)*	**Purpose** To evaluate (?prospectively) chemotherapy response using PET in patients with advanced head and neck cancer **Cases** 28 consecutive patients with Stage III/IV head and neck cancer who were participating in a neoadjuvant organ-preservation protocol using taxol and carboplatin **Methods** • PET scans and tissue biopsy performed on all patients before and after (1-2 weeks) chemotherapy • Tissue obtained after 2 courses if clinical response by clinical exam and CT was < 50% as determined by change in primary tumor size; and after 3 courses if response was > 50% • Blinded visual consensus analysis of PET by two readers using a before and after comparison format on one page per patient using a 4-point scale • ROIs measured and SURs calculated corrected for body weight while blinded to pathology results • Post therapy biopsies obtained after PET blinded to PET data • Patients classified as pathologic complete response (PCR) or residual disease (RD) based on biopsy results **Limitations of study design** • Small sample size • Short follow up time • No comparison data presented	**Detecting disease recurrence (with 95% CI) (21 positive cases, 6 negative cases)** PET: Se=90% (77-100%); Sp=83% (53-100%); PPV=95%; NPV=71%; Accuracy=89%; LR+=5.43; LR-=0.11 **Authors' comments** • PET may be used in situations when sampling bias is more likely eg. difficult access, questionable post-therapy biopsy results, or a normal, reepithelialized appearance of tumor site post-therapy • PET scan should be obtained before biopsy to avoid possible confusing effects of post biopsy inflammation or wait 5-7 days after needle biopsy or 6 weeks after surgical resection • Positive PET scan may be indicative of residual tumor and warrant repeat tissue sampling or resection • Negative PET scan may also necessitate tissue confirmation to rule out false negative results

Wong et al. (1997) *(Clinical PET Centre of Guy's and St. Thomas' Hospitals, London, UK)*	**Purposes** To prospectively evaluate primary disease, early nodal metastases, and recurrent disease in patients head and neck cancer **Cases** 54 consecutive patients who presented to head and neck clinics at both hospitals for assessment of squamous cell carcinoma (31 with primary disease, 23 with suspected recurrence or residual disease) • of which 16 had neck dissections (N0=8; N1=4; N2a=2; N2b=2) **Methods** • All patients examined under anesthesia and clinically staged before imaging • All suspicious areas of aerodigestive tract were biopsied • All patients had PET and anatomic imaging: CT (n=37); MRI (n=13); CT+MRI (n=4) • Clinical assessment, CT, MRI, PET evaluation, and histology each performed independently • Standard size and morphological criteria used to assess nodal disease on CT/MRI **Limitations of study design** • Small sample sizes in subgroup analyses • High suspicion of malignancy (referral bias) • Only biopsy verified cases analyzed (work up bias) • Criteria for positive PET not defined (potential test review bias) • Blinding unclear (potential test review bias) • PET data analysis not fully described • Order of biopsy and imaging not described.	**Detecting known primary disease (31 positive cases, 0 negative cases)** All 31 primary malignant tumors were detected by PET as hypermetabolic areas **Detecting nodal metastases (12 positive cases, 4 negative cases)** PET: Se=67% CR/MRI: Se=67% Clinical exam: Se=58% • 7 of 12 patients with involved nodes had clinically obvious nodal metastases; all were identified by PET • PET identified 1 of 5 patients with occult nodal disease **Detecting local recurrence (10 patients with recurrences, 2 non recurrences)** PET correctly identified presence or absence of disease in all 12 patients Se=100% **Detecting nodal recurrence (8 positive patients, 5 negative patients)** PET: Se=100%; Sp=100% CR/MRI: Se=75%; Sp=80% Clinical exam: Se=100%; Sp=60% **Other findings** • Abnormal uptake unrelated to malignancy was caused by osteomyelitis of the mandible following dental extraction and surgery performed within 2 months of PET imaging. Findings did not affect clinical management • Effects of post-biopsy inflammation on imaging unclear.

Study	Patients/Methods	Results/Comments
Myers et al. (1998) *(SUNY, VA Medical Center, Buffalo, NY)*	**Purpose** To assess retrospectively the clinical effectiveness of PET in the evaluation of N0 staged neck patients with squamous cell cancer (SCC) of the upper aerodigestive tract **Cases** 14 patients with N0 disease (24 total neck dissections) on clinical exam: • Stage I=1; Stage II=8; Stage III=2; Stage IV=3 from a larger study of 116 consecutive patients diagnosed with head and neck cancer, of which: • 72 had biopsy-proven SCC • 26 underwent neck dissections **Methods** • All patients had complete exam consisting of PET and panendoscopy • Nine patients had preoperative CT • All patients had modified radical neck dissections with removal of levels 1 to V • Pathologic specimens examined for number of nodes, presence of malignancy, and extracapsular spread. • PET scans correlated with pathologic results, site of primary tumor, and CT **Limitations of study design** • Small number of cases • Imaging tests influenced selection of patients for surgery and nodal sampling not described (work up bias) • Thresholds for characterizing disease on imaging not reported (potential test review bias) • Independent blind evaluation of tests and gold standard not reported (potential test review bias and diagnostic review bias)	**Detecting nodal metastases (9 positive dissections, 15 negative dissections)** PET: Se=78%; Sp=100%; PPV=100%; NPV=88%; accuracy=92% CT: Se=57%; Sp=90%; PPV=80%; NPV=75%; accuracy=76% *trend in increased accuracy of PET over CT (P = 0.11) PET + CT: Se=86%; Sp=100%; PPV=100%; NPV=91%; accuracy=95% **Other findings** • PET accurately detected presence or absence of cervical metastases in all 8 patients with SCC of the oral cavity • In 5 patients with either carcinoma of the oropharynx or hyopharynx PET correctly identified cervical metastases in two of four patients with neck metastases

Diagnostic Accuracy Efficacy Studies of FDG PET in Breast Cancer

Study	Patients/Methods	Results/Comments
Utech et al. (1996) *(Univ. of Illinois College of Medicine and Downstate Clinical PET Center, Peoria, Illinois)*	**Purpose** To study PET for staging axillary lymph node metastases in breast cancer **Cases** 124 patients with newly diagnosed and histologically proven breast cancer who were studied with PET prior to therapy: • Stage I=65; Stage IIa=30; Stage IIb=23; Stage IIIa=6 • Tumor size: < 1cm=16; >1cm=49; >2cm=30; >3cm=29 • Mixed types- 82% invasive ductal carcinoma • Hyperglycemic patients excluded • Axillary status: 10 patients with positive lymph nodes on clinical exam, 4 patients on mammography **Methods** • All patients had ER and PR assays, DNA flow cytometry, SPF, PET • DUR calculated for primary and axillary node uptake: DUR 1-3 for metastases and ≥3 for primary tumor • Qualitative PET read as positive if discrete focal uptake > background • Images read by experienced radiologists, final read by nuclear med physician from hard copy and video monitor blinded to lymph node status; reader aware of primary carcinoma • All patients had level I dissection, some had Level II, none had level III • Average # dissected nodes=20 (range 7-39) for true positives; 16 (range 7-36) for true negatives; 20 (range 9-46) for false positives • Qualitative PET compared to pathology • 20 patients with false positive results followed for 1-2 years for recurrence • DUR correlated with tumor size, grade and histopathology, SPF, DNA ploidy, and hormone receptors **Limitations of study design** • Source population unclear; ?consecutive series (potential referral bias) • Order of testing unclear • Influence of PET results on biopsy procedure unclear (potential diagnostic review bias but minimized by extensive nodal sampling)	**Detecting axillary lymph node involvement (44 positive cases, 80 negative cases)** PET: Se=100%; Sp=75%%; PPV=69%; NPV=100% • Reasons for false positive results undetermined **Other findings** • Weak correlation between DUR of metastatic axillary lymph nodes and tumor size and between SPF • No correlation between DUR and tumor grade or histopathology • Correlation between DUR and hormone receptors was undetermined **Authors' comments** • False positive findings may benefit from complementary use of lymphoscintigraphy • Confirmation is needed to determine if PET should be considered the initial test in axillary lymph nodes

Study	Patients/Methods	Results/Comments
Adler et al. (1997) *(University Hospitals and Case Western Reserve University, Cleveland, Ohio)*	**Purpose** to prospectively evaluate FDG-PET as a screening test for axillary lymph node metastases in breast cancer **Cases** 50 patients with 52 axillary dissections (2 patients with bilateral disease) who met the following inclusion criteria: • Age ≥ 30 years • Operable breast cancer • At least level 2 axillary lymph node dissection to be performed within 3 months of PET • Minimum of 10 lymph nodes dissected • Ability to fast for at least 4 hours Exclusion criteria were: • History of ipsilateral axillary lymph node dissection • Preoperative systemic therapy • Primary tumor < 5mm • Uninterpretable PET scan (2) Final primary tumor staging: T0=1; T1=31; T2=17; T3=3 **Methods** • Transmission and emission PET scans obtained on two scanners (Scanditronix SP3000 and CTI ECAT EXACT) • PET scans reviewed by two independent readers blinded to all information other than axilla side • PET scans graded on a 5-point Likert scale for presence of increased FDG uptake and scan quality; discrepancies resolved by consensus; scores ≥ 3 are positive • Histopathology obtained; all lymph nodes dissected (average #/patient=17) • Operating characteristics of both scanners compared **Limitations of Study Design** • Patient source unclear; ?consecutive series (potential referral bias) • Association between test result and gold standard determination unclear (potential diagnostic review bias minimized by extensive nodal sampling) • No follow-up data presented • Authors used higher dose of FDG and longer scanning times than in other studies	**Detecting axillary lymph node involvement (20 positive axilla, 32 negative axilla)** ***overall*** PET: Se=95%; Sp=66%; PPV=63%; NPV=95% • false positives caused by extensive sinus histiocytosis and fat replacement, mild plasmacytosis, and hemosiderin-laden macrophages • false negative likely due to patient's large size **Other findings** • Authors found no differences in results between two scanners **Authors' comments** • PET is acceptable for use as a screening test • axillary dissection should be performed in patients with positive PET scans to confirm metastases, to determine the number of lymph nodes involved, and for local control • estimates of $120,000 in cost savings and decreased morbidity would have occurred in the 22 patients with negative PET scans, but no follow up data presented to confirm • inclusion of additional interpretive criteria such as the presence of bilateral axillary activity, intensity and extent of activity may help improve specificity of axillary PET in the future *Study includes 20 patients reviewed in 1996 MDRC PET report*

Study	Patients/Methods	Results/Comments
Crippa et al. (1998) *(National Cancer Institute, Milan, Italy)*	**Purpose** to evaluate prospectively noninvasive staging of axillary nodes using PET for metastases **Cases** 68 consecutive patients with palpable breast nodules (unilateral disease=64; bilateral=4) who were scheduled for breast surgery w/ or w/o ALND based on clinical and instrumental (mammography and/or ultrasonography) results • 61 with ALND with 72 total axilla sampled: N0=36; N1a=21; N1b=13; N2=2 • 11 with no ALND were classified as negative • total # breast nodules=81 (73 malignancies, 8 benign): T1=45; T2=30; T3=2; T4=4 - 63% of malignancies were infiltrating ductal carcinomas - benign conditions were proliferative dysplasia without atypica or focal inflammation - average size = 20mm (4mm-67mm) **Methods** • PET emission and transmission scans performed 1 to 7 days before surgery • PET visual interpretation blinded to histopathology; localized uptake > surrounding tissue classified as positive • Mean SUVs of breast carcinoma were calculated • ROC analysis performed using SUVs between two groups to assess SUV as a prognostic indicator • Average # dissected nodes/axilla=21 (range 12-38) • No treatment decisions made on the basis of PET **Limitations of study design** • High index of suspicion of malignancy (referral bias, minimized by using consecutive series) • Blinding of readers to other clinical information not reported (potential test review bias) • Diagnostic review bias minimized by extensive nodal sampling • PET used in test sequence, incremental value not assessed	**Detecting axillary node involvement** *N0 disease (10 positive axilla, 26 negative axilla)* PET: Se=70%; Sp=92%; accuracy=86% *N1a disease (8 positive axilla, 13 negative axilla)* PET: Se=87.5%; Sp=100%; accuracy=95% *N1b-2 (9 positive axilla, 6 negative axilla)* PET: Se=100%; Sp=67%; accuracy=87% *overall (27 positive axilla, 45 negative axilla)* PET: Se=85%; Sp=91%; accuracy=89%; PPV=85%; NPV=91% • false positives due to vascular uptake and undetermined causes • false negatives due to microscopic and unexplained macroscopic involvement, mean size=6mm (5-8mm) **Other findings** • Median SUV in carcinomas with axillary metastases (4.6) was higher than that in carcinomas without axillary metastases (2.9), but there was a significant overlap between the two groups (interquartile ranges= 2.7-7.2 and 1.9-4.5, respectively) • ROC analysis showed best cutoff value of SUV (2.9) was associated with a Se=74% and Sp=58% • Any change in reading of ROC curve was not useful in patient management referred for ALND • Authors propose using PET in patients with very low probability of axillary metastases (t1a), in whom axillary surgery may be avoided, to monitor relapses • SUV value of primary not a prognostic indicator of axillary spread

Study	Patients/Methods	Results/Comments
Palmedo et al. (1997) *(University of Bonn, Germany)*	**Purpose** to compare prospectively the diagnostic accuracy of FDG PET vs. scintimammography (SMM) (planar and single-photon emission tomography) using ^{99m}Tc MIBI **Cases** 20 patients with 22 suspicious primary lesions detected by PE or mammography scheduled for excisional biopsy • 14 patients with 15 malignant primary lesions (including 2 local recurrences); mean size 29mm (range 8-53mm); 3 patients with tumors ≤ 8mm • 5 patients with 30 axillary node metastases (all diameters ≥ 12mm) **Methods** • SMM followed by PET scans of breasts and axillary regions were performed > 24 hrs apart in all patients during the week prior to surgery • ROIs analyzed, MIBI TNR on planar images and SUV FDG uptake calculated • Mammograms used for scintigraphic localization • Focal uptake classified as normal or abnormal • Independent, blinded interpretation of SMM and PET by two nuclear medicine physicians • PET and SMM results each compared with histopathology **Limitations of study design** • Small size • High index of suspicion for malignancy (potential referral bias) • Extent of blinding to other clinical information not clear (potential test review bias) • Association between test result and gold standard determination unclear (potential diagnostic review bias) • Incremental value of test used in work up not assessed	**Detecting unknown primary (13 malignant tumors, 7 benign tumors)** PET: Se=92%; Sp=86% SMM: Se=92%; Sp=86% • calculations excluded 2 recurrences • false positives due to fibroadenoma • false negatives due to local recurrences with diameters < 9mm **Axillary node involvement (5 positive patients, 7 negative patients)** • PET correctly detected axillary involvement in all 12 patients • (30 positive nodes) • PET detected 9 of 30 nodes • SMM detected 8 of 30 nodes **Quantitative analysis** PET: mean SUV=2.57 (0.3-6.2 with median SUV=1.6) SMM: mean TNR=1.97 (1.42-3.1 with median TNR=1.8) **Authors' comments** • Menstrual cycle may alter MIBI uptake in normal tissue • Diffuse FDG uptake in normal tissue declines with age • Thresholds and variations in SUV calculations can impact test characteristics • Both tests could detect axillary node involvement but not extent of disease • Larger cohort needed to confirm results

Study	Patients/Methods	Results/Comments
Bender et al. (1997) *(University of Bonn, Germany)*	**Purpose** To assess the feasibility of PET in staging recurrent breast carcinoma **Cases** 75 patients with suspected recurrent or with metastatic disease in undecided or equivocal cases • primary tumor histology: well-differentiated ductal carcinoma (n=46); infiltrating lobular carcinoma (n=10) • all patients had PET; 63 patients had PET and CT/MRI of which: - 14 pts with confirmed local recurrence; 17 pts w/ confirmed lymph node mets • no recurrence (N=15); local recurrence (N=20); lymph node involvement (N=28); distant metastases (N=22) **Methods** • All patients had routine work-up consisting of physical exam, axillary lymph node US, optional thorax/abdomen CT and/or MRI, bone scintigraphy and serum tumor markers • All had histologic confirmation (surgery or biopsy), except 4 by follow-up • PET evaluated qualitatively using a 4-point scale (intense, moderate, low, none) • Two readers of PET not blinded to available data • PET compared independently to each of the standard imaging modalities • All patients followed for at least 6 months (1-2 visits) **Limitations of study design** • Small study size in subgroup analyses • High index of suspicion for malignancy (potential referral bias) • Patient inclusion based in part on results of CT/MRI (work up bias but minimized by all patients receiving definitive disease verification) • Interpretation of PET results not blinded to other clinical and imaging data; positivity criteria not defined (test review bias) • Association between test result and gold standard determination unclear (potential diagnostic review bias)	**Direct visual comparison (63 patients available for direct comparison)** **Local recurrence (15 positive lesions, 48 negative lesions)** PET: Se=73%; Sp=96%; PPV=85%; NPV=92%; Acc=90% CT/MRI: Se=91%; Sp=98%; PPV=91%; NPV=98%; Acc=97% **Lymph nodes (22 positive lesions, 41 negative lesions)** PET: Se=95%; Sp=93%; PPV=88%; NPV=97%; Acc=94% CT/MRI: Se=74%; Sp=95%; PPV=89%; NPV=87%; Acc=88% (discrepancies in total numbers of PET lesions vs. total number of CT/MRI lesions) **Bone (13 positive lesions, 50 negative lesions)** PET: Se=100%; Sp=96%; PPV=87%; NPV=100%; Acc=97% CT/MRI: Se=46%; Sp=98%; PPV=88%; NPV=86%; Acc=87% **Lung (6 positive sites, 57 negative sites; in 5 patients)** • PET had 2 false positive* and 1 false negative results • CT/MRI had 2 false positive and 1 false negative results • Reasons for CT/MRI false positives not reported **Liver (2 positive sites, 73 negative sites; in 2 patients)** • PET had one false positive* result and no false negative results • CT/MRI had 1 false positive and 1 false negative result • *due to artifact wrongly interpreted during the learning phase of the facility **Authors' comments** • PET and CT/MRI had similar results re lung and liver metastases • CT/MRI identified more local recurrences correctly • PET identified 1/3 more patients with lymph node metastases, suggesting the use of PET early in restaging of breast cancer. • Semiquantitative analysis and tumor appearance info may decrease number of false positive results • Results suggest the importance of PET as a complement to morphologic tests in the staging of recurrence • PET may also play a role in whole body staging of high risk patients • Further studies are needed to assess the clinical impact of PET in the management of recurrent breast cancer and its consequence on overall survival

Study	Patients/Methods	Results/Comments
Moon et al. (1998) *(UCLA, Los Angeles, California and University of Ulsan, Seoul, Korea)*	**Purpose** To retrospectively evaluate the diagnostic accuracy of PET in patients with suspected recurrent or metastatic breast cancer **Cases** 57 female patients with 83 reference sites (29 with disease, 28 no recurrence or metastases) • who underwent primary surgery with or without adjuvant chemo- or radiation therapy and • who were referred to the UCLA PET center from October 1990 to October 1995 (mean time interval between diagnosis and PET scan=4 yrs. range 1 mo. To 17 yr 9 mo) • who had a clinical suspicion of disease recurrence not resolved by conventional imaging excluded cases: • patients who underwent chemo-or radiation therapy within 3 mo before PET • lesions that were biopsied • lesions diagnosed with known disease **Methods** • All patients underwent history and PE, multiple labs and imaging tests • Diagnostic confirmation based on biopsy, lesion morphology for tumor on 2 or more imaging studies, and for at least 6 months clinical and radiographic follow up • PET abnormalities that resolved without treatment were considered to be false-positive results • ECAT 931 and ECAT 961 were used • Independent visual inspection of PET by 3 readers informed clinical suspicion of metastases, but blinded to gold standard; discrepancies resolved by 4th reader aware of a discrepancy but not of specifics • PET images scored from 1 (definitely negative) to 5 (definitely positive) • Lesion site defined as any abnormality suggesting the possibly of breast recurrence or metastases either clinically or on imaging, therefore, analysis was biased toward positive lesions • Analysis by patient and by lesion **Limitations of study design** • Only more difficult cases included and suspicious sites assessed (referral bias) • Not all regions were prospectively examined with other conventional imaging tests • Partial blinding (potential test review bias) • Blinding of gold standard diagnosis (confirmation) to PET results unclear (potential diagnostic review bias)	**Overall diagnostic accuracy (29 positive cases, 28 negative cases)** ***Scores ³ 4 defined as positive*** PET: Se=93%; Sp=79%; PPV=82%; NPV=92% ***Scores ³ 3 defined as positive*** PET: Se=93%; Sp=61%; PPV=82%; NPV=92% **Overall diagnostic accuracy (41 positive lesions, 39 negative lesions)** ***Scores ³ 4 defined as positive*** PET: Se=85%; Sp=79%; PPV=81%; NPV=84% ***Scores ³ 3 defined as positive*** PET: Se=90%; Sp=54%; PPV=67%; NPV=84% **ROC analysis** *Az*=0.91 for patient detection; *Az*=0.88 for lesion detection **Interobserver variability** • In 48% of patients, scores of all 3 observers were the same • In 38% of patients, a score from one reader deviated one score grade from the other 2 readers • In 14% of patients, a score from 1 reader deviated more than 1 score grade from the other readers **Other findings** • Bone metastases had a larger proportion of false-negative lesions than other malignant sites when scores of 4 or 5 were regarded as positive; Bone Se=69% (11 of 16) vs. Non-bone Se=96% (24 of 25) ($p<0.05$) • False negative lesions (scored ≥ 4) included 5 bone metastases and one small breast site, of which 3 bone and one breast lesion showed mild uptake (scored 2-3), one lesion was confirmed positive on follow up PET scan • When scores of ≥ 3 were regarded as positive, lymph node sites had more false positives than other sites; Lymph Sp=13% vs. Other Sp=79% ($p<0.05$); with scores ≥ 4, Lymph Sp=60% vs. Other Sp=92% ($p<0.001$) • PPV=62% for lesions with a score of 4 and PPV=90% for lesions with a score of 5 • False positive lesions (scored ≥ 3) were attributed to muscle uptake, inflammation, and physiological and artifactual FDG uptake, and unknown causes • Characteristics of 5 patients with fasting blood glucose levels > 110 mg/dl: 4 diabetics, 2 received insulin, 3 were true negative, one true positive. **Authors' comments** • More strict attention to patient preparation, recognition of artifactual uptakes, and information on clinical history (re inflammatory disease) will improve specificity of PET • Limitations of study: study biased toward positive lesions and more difficult cases, not all regions were prospectively examined with other conventional imaging studies • Attenuation correction would provide a more accurate representation of tracer distribution • A prospective study is needed to further assess the role of PET in post-surgical breast cancer management

Diagnostic Accuracy Efficacy of FDG PET in Lung Cancer

Study	Patients/Methods	Results/Comments
Bury et al. (1997) *(CHU Liège, Belgium)*	**Purpose** to prospectively compare the accuracy of FDG-PET and conventional imaging (CI) for staging NSCLC **Cases** 141 consecutive patients between 9/94-10/96 with newly diagnosed NSCLC based on sputum cytology, needle biopsy, or flexible bronchoscopy: • exclusion criteria included poor physiological status (n=21) and inappropriate follow up (n=11) • of which 109 were enrolled in the study: • 77 men, 32 women; mean age=64 (44-83) • squamous cell=50; adenosquamous cell=8; adenocarcinoma=46; undifferentiated large cell=5 • stage I=32; stage II=8; stage IIIA=22; stage IIIB=8; stage IV=39; N0=32; N1=20; N2=10; N3=4; T4=4 • benign conditions = nonspecific inflammation, pneumonia, multinodular goiter, localized FDG uptake in hepatic-splenic angle of colon • 66 cases with suspected mediastinal involvement on CT or PET had biopsy confirmation **Methods** • all patients had CI before PET; CI= chest and abdominal CT scanning and bone scintigraphy • suspicious lesions on bone scintigraphy confirmed by bone radiography • contrast CT positive criteria > 10mm on short axis • PET data analyzed by visual interpretation; positive results=moderate (about twice the activity in contralateral or reference region) and intense (markedly higher than reference activity) FDG uptake • PET and CI interpreted separately by 2 nuclear medicine readers and 2 radiology readers blinded to N and M histology but not blinded to histology of primary tumor • confirmation of suspected mediastinal involvement or distant metastases on CI or PET done within 21 days of imaging • N staging confirmed by extensive nodal sampling • M staging confirmed by biopsy (21)or clinical and/or radiologic follow-up (88); absence of demonstrated metastases 6 months after negative imaging considered negative for metastases • statistical analysis by patient **Limitations of study design** • choice of N stage cohort influenced by imaging tests; only biopsy verified cases included in N cohort (work up bias minimized by all patients having both tests) • readers not blinded to primary tumor histology (partial test review bias) • strong association between test results and determination of gold standard (diagnostic review bias minimized by extensive nodal sampling in N staging • incomplete reporting of methods for evaluating changes in treatment strategy	**Defining known primary disease (109 cases)** • all primary tumors showed increased focal FDG uptake; intense in 101 cases, moderate in 8 cases • no correlation between histopathology and FDG uptake **Mediastinal involvement (34 positive cases, 32 negative cases)** PET: Se=89% (95% CI: 72-96%); Sp=87% (95% CI: 71-97%); PPV=89% (95% CI:72-96%); NPV=87% (95% CI: 71-96%); accuracy=88% CT: Se=79%; Sp=71%; PPV=75%; NPV=76%; accuracy=75% (no 95% CI reported) • disagreement between PET and CT in 29 cases (44%); correct changes by PET=22 cases (33%); correct changes by CT=7 (11%) **Distant metastases (39 positive cases, 70 negative cases)** PET: Se=100% (95% CI: 91-100%); Sp=94% (95% CI: 86-98%); PPV=90% (95% CI: 78-97%); NPV=100% (95% CI: 95-100%); accuracy=96% (95% CI: 90-98%) CI: Se=82%; Sp=89%; PPV=80%; NPV=89%; accuracy=86% (95% CI not reported) • moderate FDG uptake in 7 of 8 cases were < 2 cm. • PET false positives caused by nonspecific inflammation in axillary lymph node, pneumonia sequelae, benign multinodular goiter, anatomical misidentification • PET had no false positive FDG uptake in adrenal glands • PET correctly changed M stage, as determined by CI, in 15 cases (14%) **Changes in therapeutic strategy** • PET modified therapy in 27 patients (20%) (10 to curative surgery, 8 to a more curative approach with chemo- and/or radiation therapy, 9 to more palliative approach) • no patient follow up data reported **Other findings** • authors found PET more useful than CI in evaluating adrenal masses • false positive PET in axillary site probably caused by extravasation of antecubital vein during FDG injection • lack of anatomical markers limited precise localization of some PET findings • CI + PET could increase the accuracy of detection • prospective comparison needed to compare PET with bone scanning

Study	Patients/Methods	Results/Comments
Erasmus et al. (1997) *(Duke Univ. Medical Center, Durham, NC)*	**Purpose** To assess PET in differentiating benign from metastatic adrenal masses in patients with bronchogenic carcinoma **Cases** 27 consecutive cases with 33 total lesions (23 malignant, 10 benign) presenting to thoracic surgery, oncology, or pulmonary between January 1993 and January 1996 with bronchogenic carcinoma and an adrenal mass detected by CT Characteristics: • 19 men, 8 women; mean age 57 yrs (range 39-76) • 24 with NSCLC, 3 small cell; bilateral masses in 6 patients • mean diameter of adrenal masses=3 cm (range 1-9cm) **Methods** • FDG PET performed after CT • Independent interpretation of adrenal activity by 3 readers blinded to clinical and pathologic findings and other imaging test • Positive activity= activity > background; negative activity= activity ≤ background • ROI and SUR determined blinded to biopsy results • Confirmation of adrenal masses by: - percutaneous needle biopsy (n=11) within a mean of 5 days before PET (n=9) and after (n=2), - growth characteristics on follow up (mean=4 months) CT (n=16), and - CT Hounsfield unit measurement < 10H diagnostic of a benign lesion (n=6) **Limitations of study design** • High probability of malignancy and benign conditions not depicted (potential referral bias) • Association between PET results and choice of confirmation method unclear (potential diagnostic review bias) • Incremental value of PET in test sequence not determined	**Defining adrenal disease (23 malignant lesions, 10 benign lesions)** PET visual analysis: Se=100%; Sp=80% SUR analysis: malignant lesions mean=6.28 ± 2.5 vs. benign lesions mean=1.77 ± 0.89 ($p < 0.0001$) **Other findings** • Characteristics of malignant masses: - mean diameter = 4 cm - n=6 new on follow up CT (mean time=4 months), n=10 growth changes on CT; n=7 by biopsy - changes in adrenal masses on CT consistent with changes in thorax • Characteristics of benign masses: - n=4 by biopsy; n=5 with features < 10 H on CT; n=1 with benign features on CT

Study	Patients/Methods	Results/Comments
Guhlman et al. (1997) *(University of Ulm, Germany)*	**Purpose** to evaluate retrospectively the accuracy of FDG PET in thoracic lymph node staging in patients with NSCLC **Cases** 46 consecutive patients (32 cases, 14 benign processes) who underwent thoracotomy for lung tumors from 1994 to 8/95: • 41 men, 5 women; mean age=56.7 yrs (24-78) • squamous cell (n=19); adenocarcinoma (n=7); large cell (n=6) • T1=3; T2=12; T3=10; T4=7; N0=12; N1=5; N2=11; N3=4 • benign conditions: pneumonia (n=4); tuberculosis (n=3); one each of florid abscess, aspergilloma, hamartoma, aneurysm of subclavian artery, lung fibrosis, inflammatory pseudotumor of which 32 malignant cases underwent further mediastinal evaluation **Methods** • all patients underwent contrast CT of chest prior to PET 3 weeks before surgery • positive node on CT defined as > 10mm in short axis diameter • blind, independent interpretation of CT by two experienced radiologists to clinical and PET findings • blind, independent visual interpretation of PET by 2 experienced nuclear medicine physicians • histopathology and TN classification confirmed surgically on patients with primary lung cancer • surgeon conducted thorough dissection of mediastinal nodes, data on extent not reported • PET and CT results mapped and compared to histologic findings • statistical analysis reported by patient **Limitations of study design** • retrospective study of surgical series—high probability of malignancy (potential referral bias) • small sample size limits subgroup analyses • only biopsy verified cases analyzed (work-up bias) • association between test results and biopsy confirmation unclear (potential diagnostic review bias, minimized by extensive nodal sampling)	**Defining unknown primary tumor (32 malignant cases, 14 benign cases)** PET: Se=94%; Sp=86%; accuracy=91% CT: data not reported • PET false positives caused by an aspergilloma with active inflammation and a florid abscess • PET false negatives caused by a 1 cm intrapulmonary adenocarcinoma metastasis and bronchioalveolar carcinoma **Mediastinal/Hilar involvement (with 95% CI) (20 positive cases, 12 negative cases)** PET: Se=80% (56%-94%); Sp=100% (73%-100%); accuracy=87% (71%-96%) • PET accuracy (# patients): N0=12/12; N1=3/5; N2=9/11; N3=4/4 CT: Se=50% (27%-73%); Sp=75% (43%-95%); accuracy=59% (41%-76%) (p<.02 for overall accuracy) • CT accuracy (# patients): N0=9/12; N1=2/5; N2=6/11; N3=2/4 • CT false positive nodes caused by nonspecific inflammation including sarcoidosis, inflammatory pseudotumor, and pneumonia • PET false positive hilar nodes caused by nonspecific inflammation • PET differentiated N1/N2 disease from N3 disease in 4 patients, but only 2 of 4 with CT **Other findings** • curative resection would have been avoided in 2 patients with N3 disease on PET but not with CT • tumor involvement of peribronchial hilar nodes and mediastinal nodes adjacent to the bronchus may be enhanced with anatometabolic imaging • PET + CT may decrease the need for invasive diagnostic procedures such as mediastinoscopy

Study	Patients/Methods	Results/Comments
Hagberg et al. (1997) *(VA Palo Alto Health Care System and Stanford University School of Medicine)*	**Purpose** • to evaluate retrospectively PET in characterizing pulmonary nodules and staging bronchogenic carcinoma • to compare CT with PET for diagnosing N2 disease **Cases** 49 consecutive patients presenting between 9/94 and 3/96 (31 malignant cases, 18 benign cases) with 54 pulmonary nodules (44 positive nodules, 10 negative nodules): • 45 men, 4 women; mean age=63 (37-85) • squamous=15 nodules; adenocarcinoma=16; large cell=3;adenosquamous=3;bronchoalveolar=2; atypical carcinoid=1; small cell=1; renal cell=2;malignant melanoma=1 • benign conditions: granuloma=4; hamartoma=3; necrotic tissue=2; fungal ball=1 • exclusion criteria: indeterminate PET scan (2) and inadequate histopathologic information of mediastinum (11) • 18 of 31 malignant cases had complete PET, CT, and histopathology information and were included in analysis of mediastinal nodes (N2 only) **Methods** • CT performed and interpreted prior to PET • CT interpreted by one investigator; positive mediastinal lymph nodes > 1 cm in short axis diameter • initial PET scans visually interpreted by one investigator not blinded to CXR or CT findings • blinded mediastinal PET images reread by two investigators • PET FDG uptake classified as positive, negative or indeterminate—no difference between initial read and reread, but methods for comparison not described • all pulmonary nodules confirmed by histo- or cytopathology; extent of nodal sampling not reported • data analyzed by node **Limitations of study design** • small number of subjects • retrospective design--patient source and filters unclear (potential referral bias) • influence of imaging tests on selection of surgical candidates unclear; N stage cohort restricted to biopsy verified cases (work up bias) • association of test results and determination of gold standard unclear (potential diagnostic review bias) • interpretation of primary tumor not blinded (test review bias)	**Defining unknown primary tumor (44 positive nodes, 10 negative nodes)** PET: Se=93%; Sp=70% CT: no data reported • all false positives caused by granulomas • 3 false negatives caused by poor quality PET scans (hyperglycemia at the time of scan, no attenuation correction, r outdated scanner); one false negative due to renal cell carcinoma unexplained **Mediastinal involvement—N2 disease only (9 positive nodes, 9 negative nodes)** PET: Se=67%; Sp= 100% CT: Se=56%; Sp=100% • small numbers, selection and verification bias may contribute to lack of significance between PET and CT **Other findings/Comments** • quantitative analysis may allow more accurate differentiation of disease • balanced discussion of the influence of study size, selection bias and verification biases on results • further study needed to clarify role of PET in staging • cost-effectiveness studies needed prior to advocating the routine use of PET in management of NSCLS

Study	Patients/Methods	Results/Comments
Steinert et al. (1997) *(University Hospital, Zurich, Switzerland)*	**Purpose** to compare prospectively (??) the accuracy of FDG PET with CT in staging NSCLC **Cases** 62 surgical candidates with suspected or proven NSCLC who had PET between 2/94 and 3/96 and who had no prior neoadjuvant therapy or diabetes • exclusion criteria: inadequate CT=2; distant metastases=8; inadequate nodal sampling=5 • therefore, 47 patients with suspected or confirmed NSCLC of mixed types remained in the study • squamous=24; adenocarcinoma=17; large cell=6 • 29 (62%) with nodal metastases; N0-N1=34; N2=7; N3=6 **Methods** • CT, emission and transmission PET obtained on all patients • blind, independent interpretation of PET and CT scans before surgical staging • CT positive criteria=nodes > 10mm in short axis diameter, except upper paratracheal stations > 7mm or infracarinal station > 11mm • presence and site of mediastinal and tracheobronchial nodes recorded according to ATS lymph node station mapping system; extent of lymph node metastases classified according to AJC lung cancer staging system • all patients underwent extended surgical lymph node staging regardless of PET or CT findings or nodal size • PET and CT available to surgeon during surgery • patients undergoing left thoracotomy had limited sampling of mediastinal lymph nodes; care was taken to resect all preoperatively staged positive lymph nodes • PET and CT correlated with histopathologic results **Limitations of study design** • Real-time prospective design unclear • Patient source unclear and high probability of malignancy in surgical series (potential referral bias) • Only biopsy verified cases analyzed (work up bias) • Strong correlation between imaging results and biopsy confirmation (diagnostic review bias) minimized by nodal sampling	**Detecting mediastinal lymph node metastases (28 positive nodes, 84 negative nodes)** PET: Se=89%*; Sp=99%; PPV= 96%; NPV=97%; accuracy=97% CT: Se=57%*; Sp=94%; PPV=76%; NPV=87%; accuracy=85% * (*P*=.0066) (n=47 patients) • PET: 45/47 (96%) correctly staged vs. CT: 37/47 (79%) (*P*=.01341) **Other findings** • no statistically significant difference between staging with conventional CT and with spiral CT • correlation of nodal station on imaging with surgical results matched except for one case

Study	Patients/Methods	Results/Comments
Vansteenkiste et al. (1997) *(UHG, Leuven, Belgium)*	**Purpose** to compare prospectively?? CT, PET, and PET + CT in staging mediastinal lymph nodes in patients with NSCLS **Cases** Unknown # of patients with suspected or biopsy proven NSCLC who were potentially operable after standard staging for distant metastases • exclusion criteria: diabetes; treatment with oral corticosteroids; ischemic cardiomyopathy; mediastinal invasion of primary tumor; obvious bulky metastases • records of 50 patients treated between 9/95 and 4/96 were analyzed - squamous=32; adenocarcinoma=10; large cell=8 - T1==3;T2=32; T3=15; N0=35; N2=15 **Methods** • interpretation of contrast CT by two interpreters blinded to bronchoscopic or pathologic findings; positive node ≥15 mm long axis diameter • SUV PET images interpreted blinded to clinical, CT, and pathologic data by two independent readers; five point semiquantitative scale used 1-5; positive node= 4 or 5 • CT + PET visually interpreted by two readers blinded to pathologic data • surgical staging done by mediastinoscopy and intraoperative staging in case of resection • CT, PET, and surgical staging carried out within one month • MLN map used for imaging and surgical staging • data analyzed by patient **Limitations of study design** • Small sample size • Limited patient data; high probability of malignancy (potential referral bias) • prospective design unclear • only biopsy verified cases analyzed (work up bias) • association between test results and determination of gold standard unclear (diagnostic review bias) • extent of nodal sampling not reported • blinding to clinical data on visual interpretation not reported (potential test review bias) • methods for assessing changes in treatment not reported	**Detecting mediastinal involvement (15 positive cases, 35 negative cases)** PET: Se=67%; Sp=97%; accuracy=88%; PPV=91%; NPV=87% CT: Se=67%; Sp=63%; accuracy=64%; PPV=43%; NPV=81% PET + CT: Se=93%; Sp=97%; accuracy=96%; PPV=93%; NPV=97% • of 18 discordant results, CT was correct in 1/18, and PET was correct in 17/18 (p=0.004) **Authors' comments** • mediastinoscopy could be omitted in patients with normal CT and PET or in patients with abnormal CT but normal PET • all patients with abnormal mediastinal PET should still proceed to invasive mediastinal staging, to be sure that no patient with N0 or N1 disease is denied the chance of cure by direct surgical resection • in this study the need for invasive mediastinal staging could be reduced to 13 or 50 patients, resulting in important savings in operation time • if accuracy of mediastinal PET will be confirmed in future data, it is likely that PET will substantially change current clinical practice of staging of NSCLC

Study	Patients/Methods	Results/Comments
Sasaki et al. (1996) *(Kyushu University, Fukuoka, Japan)*	**Purpose** to compare ?prospectively FDG PET with CT in the detection of mediastinal lymph node metastases **Cases** 29 newly diagnosed patients with NSCLC of mixed types who had undergone surgery and who had pathologic confirmation of disease • adenocarcinoma=18; squamous=9; adenosquamous=1; large cell=1 • N0-N1=17; N2=11; N3=1 • 132 out of 261 mediastinal lymph nodes were surgically resected and histopathology confirmed of - 71 regions had CT, PET, and histopathologic information and were included in the study **Methods** • mediastinal lymph nodes classified into nine regions based on mapping proposed by the Japan Lung Cancer Society • FDG uptake measured by TMR of the primary tumor • visual interpretation of PET images performed by 3 nuclear medicine readers • PET positive criteria were FDG uptake in nodes> that in other mediastinal structures • contrast CT interpreted by 2 radiologists • CT positive nodes ≥10 mm on short axis diameter • gold standard biopsy obtained surgically • data analyzed by nodal region **Limitations of study design** • not real-time prospective design (referral bias) • regions included in analysis represent an unknown number of patients • only biopsy verified cases analyzed (work up bias) • blinding of readers not reported (test review bias) • association between test results and determination of gold standard unclear, and extent of nodal sampling not described (potential diagnostic review bias)	**Detection of known primary (29 lesions)** PET FDG uptake: 9.1±4.6 (TMR±SD) CT: 43.3±18.5mm (mean ± SD) **Mediastinal lymph node metastases (17 positive regions, 54 negative regions)** PET: Se=76%; Sp=98%*; Accuracy=93%**; PPV=93%; NPV=93% CT: Se=65%; Sp=87%*; Accuracy=82%**; PPV=61%; NPV=89% *,** (P<0.05) • All PET false negatives < 7mm in short axis diameter due to partial volume effect • smallest true positive on PET was 7mm in short axis diameter • CT false positives caused by non-specific inflammatory changes and an enlarged tracheobronchial lymph node of unreported cause • PET false positive caused by an enlarged tracheobronchial lymph node of unreported cause **Authors' Comments** • voluntary and involuntary movement can contribute to underestimation of FDG uptake • mediastinal evaluation may have been limited by PET field of view • use of quantitative analysis using PET is limited in lymph node evaluation due its complexity, the influence of the partial volume effect, and limited visual identification of lymph nodes • authors suggest PET as a complementary diagnostic method with CT; improvements in technical and quantitative methods should improve the diagnostic ability of PET

Diagnostic Accuracy and Diagnostic Thinking Efficacy Studies of FDG PET in Solitary Pulmonary Nodules

Study	Patients/Methods	Results/Comments
Dewan et al. (1997) *(Creighton University and VAMC Omaha, NE)*	**Purpose** to compare the probability of cancer (pCA) in a SPN using standard criteria with Bayesian Analysis and PET (retrospective analysis) **Cases** 52 consecutive patients (37 malignant cases, 15 benign cases) who met the following selection criteria: • underwent PET imaging between April 1990 and February 1994 • noncalcified, noncavitary SPN based on CXR and CT classified as indeterminate • age > 30 years • nodule size ≤ 3 cm • group included 3 patients with extrathoracic malignancy and one patient with stable nodule for > 2 yrs Reported patient characteristics: • 43 men, 9 women; mean age=63.6 yrs ±11.3yrs; 79% current smokers of which 52% smoked ≥ 20 cigs/day • Edge characteristics (%malignant cases vs. % benign cases): Sharp, smooth=14% vs.20%; Lobulated=30% vs. 40%; Slightly irregular w/ few spiculations=38% vs. 33%; Grossly irregular and spiculated=19% vs. 7% **Methods** • PET performed within 2-4 weeks after CT; CT densitometry not performed • histologic diagnosis obtained by thoracotomy, mediastinoscopy, bronchoscopy, or needle lung biopsy • qualitative PET scans read by one reader blinded to histology; clinical and radiologic data available to the reader varied, but size and location was known in all patients • benign PET=no focal FDG uptake; malignant PET=focal FDG accumulation greater than surrounding tissue but more than mild • nodule edge on CT interpreted independently by 2 pulmonologists using 4 type classification system blinded to clinical diagnosis; discrepant interpretations reached by consensus • odds-likelihood of malignancy estimated using Bayes Theorem • standard criteria for probability of cancer (pCA) based on patient's age, smoking history, history of prior malignancy, nodule size and edge, and presence of calcification • pCA of standard criteria compared to standard criteria + PET and PET alone **Limitations of Study Design** • retrospective design • all patients has invasive biopsy determination, implying a high index of suspicion for malignancy (referral bias) • source of patient cohort influenced by test results (work up bias) • association between test results and determination of gold standard unclear (potential diagnostic review bias) • blinding of clinical and radiologic information varied (test review bias) • PET and other tests not independent, a requirement of Bayes' Theorem • pre-PET probability of cancer in patients unknown	**Diagnostic Accuracy (37 malignancies, 15 benign)** *Overall* PET+CT: Se=95%; Sp=87%; accuracy=92% *nodules £1.5 cm* PET+CT: Se=83%; Sp=100% *nodules >1.5 cm* PET+CT: Se=100%; Sp=67% • 2 false positives due to histoplasma granuloma with active inflammation • 2 false negatives were 1 cm scar adenocarcinoma and adenocarcinoma **Diagnostic Thinking Efficacy** **Bayes' Theorem** • LR for malignant SPN with abnormal PET=7.11 (95% CI, 6.36 to 7.96) • LR for malignant SPN with normal PET=0.06 (95% CI, 0.05 to 0.07) **ROC curve analysis** PET alone was the best predictor of malignancy at different levels of pCA, the standard criteria the worst, and standard criteria + PET was intermediate

Study	Patients/Methods	Results/Comments
Lowe et al. (1998) *(multi-site study from 9 U.S. sites)*	**Purpose** To prospectively evaluate the diagnostic accuracy of FDG-PET imaging in evaluating SPNs **Cases** 89 of 105 consecutive patients who met the following inclusion criteria: • Imaging performed between October 1993 and August 1994 • SPNs 0.7cm-4.0cm in size visualized on CT • Considered indeterminate for malignancy by CXR and CT and clinical data • Definitive pathologic confirmation by TTNA (n=29) or surgery (n=60) Excluded patients: 8 no definitive pathology; 4 definitely benign SPN on CXR/CT; 2 w/o CT Reported patient characteristics: • 61 men, 28 women; mean age= 63 ±9.5 yrs **Methods** • Imaging performed prior to treatment of SPN • AP and lateral CXR and CT of at least chest and adrenals obtained; thin-section transaxial images and IV contrast used in some studies • Independent qualitative interpretation of CXR and CT by readers at 2 participating sites other than where the studies were performed, blinded to clinical, PET, and gold standard results • Semiquantitative analysis (SUV) performed; SUV> 2.5 = malignant • Independent visual analysis of PET by 2 readers (of 3 available readers) blinded to clinical, CXR/CT, and gold standard results in each case; focal uptake > mediastinal blood pool structures= malignant • PET compared to histology; Se, Sp, accuracy, and LRs calculated • K value assessed for interobserver variability **Limitations of study design** • small sample size in subgroup analyses • ? real-time prospective data collection or retrospective data collection from surgical series; conditional independence among tests unclear • high prevalence of malignancy in study population (referral bias) • inclusion criterion of biopsy verification biased toward patients with high probability of malignancy (work-up bias) • blinding of surgeons to PET results not always conducted for ethical reasons (diagnostic review bias) • pre-PET probability of cancer in patients unknown *Note: Some sites (eg. Creighton University Medical Center, Omaha, Nebraska) may have included data from previously published patient series, which were reviewed in the 1996 MDRC PET assessment.*	**Defining SPN (60 malignant cases, 29 benign cases) reported with 95% CI** *Overall (£ 4cm)* SUV: Se=92% (82-100%); Sp=90% (79-100%); Acc=91%; LR+=9.0; LR-=0.09 Visual: Se=98% (95-100%); Sp=69% (57-81%); Acc=89%; LR+=3.0; LR-=0.02 • Malignant mean SUV ±1 SD=6.9 ± 3.9 vs. benign =1.7 ± 1.0 ($P < 0.001$) *£ 1.5cm (15 positives, 19 negatives)* SUV: Se=80% (60-100%); Sp=95% (85-100%); Acc=88%; LR+=15.0; LR-=0.2 Visual: Se=100% (100-100%); Sp=74% (55-93%); Acc=85%; LR+=4.0; LR-=0.0 *>1.5cm (45 positives, 10 negatives)* SUV: Se=96% (90-100%); Sp=80% (55-100%); Acc=93%; LR+=5.0; LR-=0.06 Visual: Se=98% (94-100%); Sp=60% (46-74%); Acc=91%; LR+=2.0; LR-=0.04 *£ 3.0cm (51 positives, 26 negatives)* SUV: Se=90% (82-98%); Sp=92% (85-99%); Acc=91%; LR+=12.0; LR-=0.1 Visual: Se=98% (94-100%); Sp=69% (56-82%); Acc=88%; LR+=3.0; LR-=0.03 **Other findings** • K=0.95; 2/89 discrepant cases caused by granuloma and acute inflammation • benign conditions: granuloma (7), coccidiomycosis (4), benign cellular debris (4), nonspecific inflammation (3), necrotizing granuloma (3), fibrosis (1), hemangioma (1), aspergillosis (1), metaplasia (1) • malignancies: NSCLC (50), melanoma (5), Hodgkin's lymphoma (1), small-cell (1), malignant neural tumor (1), malignant carcinoid (1), colon cancer (1) • SUV false negatives (5): 2.0cm bronchioalveolar cancer, 1.5cm squamous cell cancer, 1.0cm melanoma mets, 2.5cm squamous cell cancer in a patient with blood glucose=341; all false negatives in R upper lobe • Visual false negatives (1): 2.0cm bronchioalveolar cancer • SUV and visual false positives (3): granuloma, necrotizing granuloma, necrotizing granuloma with histoplasmosis; 2 in R upper lobe, 1 in l upper lobe • Serum glucose values on 61 patients, mean ±SD=99 ± 56 mg/dL; 2/7 diabetics had elevated glucose levels with 1 false positive and 2 false negative results • In 4 cases CXR did not identify the SPN **Authors' comments** • Results from visual analysis might be more helpful than quantitative analysis for small nodules (≤ 1.5cm) and in cases in which elevated glucose levels are unavoidable to reduce the number of false negative cases • Given the multisite nature of the study and multiple radiologists used to interpret the films across sites, study population is representative of the proportion of indeterminate SPNs from the sites included

Diagnostic Accuracy and Therapeutic Efficacy Studies of FDG PET in Colorectal Cancer

Study	Patients/Methods	Results/Comments
Delbeke et al. (1997) *(Vanderbilt University Medical Center, Nashville, Tennessee)*	**Purposes** Prospective assessment: • To assess the accuracy of FDG PET vs. CT vs. CT arterial portography (CTAP) in detecting liver metastases • To assess the accuracy of FDG PET vs. CT in detecting extrahepatic metastases • To evaluate the impact on management of patients with recurrent colorectal carcinoma (retrospective) **Cases** 52 consecutive patients presented on 61 occasions for evaluation of suspected recurrent carcinoma based elevated CEA levels or abnormal findings on CT (includes 9 repeat patients) • 45 had liver metastases, including 16 with concomitant extrahepatic disease, 10 had extrahepatic disease only • Total liver lesions: 104 malignant, 23 benign (0.3 cm-6 cm in size) • Total extrahepatic lesions: 34 malignant, 5 benign • Benign conditions: Normal liver (7), Post-surgical site (8), Local fibrosis (2), Resolving abscess (1), hepatic cyst (1), hematoma (1) • 31 men, 21 women; Mean age 63 ± 11 yrs **Methods** • 40 patients underwent abdominal CT; CT portography=40; or both=29 • PET, CT, and CT portography (both with contrast) performed within 2 months of each other • Patients with abnormal PET scans in extra-abdominal areas had additional CT scan of that region • PET visually interpreted, and analyzed semiquantitatively using SUR corrected for body weight, by two nuclear medicine physicians; SUR calculations excluded lesions < 1 cm in diameter • CT and CT portography interpreted independently by two experienced radiologists • All readers blinded to other imaging results • Disease confirmed with clinical or radiologic follow up (n=17) or histopathology obtained surgically (n=44) , except for two lesions that were examined after percutaneous fine needle aspiration • Surgical exam and intraoperative ultrasound used to confirm nonresected liver lesions • Recurrence defined pathologically or by suspected recurrence on imaging • Changes in patient management retrospectively reviewed with the surgeons **Limitations of Study Design** • High prevalence of malignancy and unclear patient source (potential referral bias) • Strong association between imaging results and choice of patient cohort; CT of extrahepatic areas dependent on PET results (work up bias, minimized by follow up of all patients) • Criteria for positive test and cut-off for semiquantitative analysis not reported (potential test review bias) • Blinding to other clinical information not reported (potential test review bias) • Association between imaging test results and gold standard determination unclear (potential diagnostic review bias) • Details of methods for evaluating therapeutic efficacy not reported	**Detecting recurrences overall** (55 patients with recurrences, 6 with scar)** PET: Se=98%; Sp=83% CT: insufficient data to calculate results CTAP: insufficient data to calculate results **Detecting liver lesions** (104 malignant lesions, 23 benign lesions)** PET: Se=91%; Sp=96%; accuracy=92% CT: Se=81%; Sp=60%; accuracy=78% CTAP: Se=97%; Sp=5%; accuracy=80% *Excluding lesions < 1 cm (18 malignant, 5 benign)* PET: Se=99%; accuracy=98% CT: Se=87%; accuracy=83% CTAP: Se=97%; accuracy=80% **Other findings** • If only histologically proven lesions were included, test characteristics remained within 1% of above values, but no data available to replicate calculations • Accurate differentiation of postsurgical changes from malignant recurrence: PET= 12/14 sites; CT=7/11 sites; CTAP=5/11 sites **Detecting extrahepatic lesions**(34 malignant lesions, 5 benign lesions)** PET: Se=100% CT: Se=74% **Quantitative analysis of hepatic lesions**** • SUR malignant = 8.1 ± 4.1 vs. SUR benign = 2.0 ± 1.0 (p< 0.0001) • For extrahepatic lesions the SUR was less helpful than CT in differentiating bowel uptake from metastases **Therapeutic efficacy**** • PET helped to plan surgery by identifying site of recurrence in 10% of patients (n=6) • PET helped to avoid unnecessary surgery in 18% of patients (n=11) • Impact of false positive and false negative PET scans on patient management was not reported ***Note: PET utility was evaluated complementary to diagnostic tests done earlier in the work up.*

Study	Patients/Methods	Results/Comments
Ogunbiyi et al. (1997) *(Washington Univ. School of Medicine, St. Louis, Missouri)*	**Purpose** Retrospective assessment: • To evaluate PET versus CT for staging recurrent and metastatic colorectal cancers • To assess the impact of PET on clinical management of patients with colorectal cancer **Cases** 58 patients had PET between 1/91 and 1/95 for evaluation of suspected recurrent (n=47) or advanced primary (n=11) disease: • based on high clinical suspicion and equivocal or positive CT findings (n=39) or clinical suspicion alone, including raised CEA levels with normal CT (n=19) • 33 men, 25 women; mean age 60 yrs. (23-81 yrs) • benign conditions not reported in reproducible detail **Methods** • All patients underwent colonoscopy and contrast CT of chest, abdomen and pelvis within 4 wks prior to PET • CT interpreted for extent of local pelvic recurrence and presence of metastases • Qualitative PET interpreted by two readers with access to CT results • Malignancy=FDG uptake moderately or markedly intense; benign=no or mild uptake, or if abnormality identified on other imaging for which no corresponding abnormality was present on PET • Gold standard= surgery, histology, or both (n=40); clinical and radiologic follow up (n=16); autopsy reports(n=2), and treatment outcomes • All patients followed for at least 12 months after PET or until death • Impact of PET on patient management was assessed; positive impact=alteration in clinical decisions with PET results **Limitations of study design** • ? Consecutive series • Retrospective analysis • High prevalence of malignancy (potential referral bias) • Strong association between test results and choice of patient cohort (work-up bias, minimized by follow up of all subjects for at least 12 months after pet or until death) • Criteria for positive result on imaging not reported • Blinding not reported; ct results available to pet readers (test review bias) • Incremental value of pet not assessed • Association between test results and gold standard determination unclear (potential diagnostic review bias) • Methods for assessing therapeutic efficacy not reported	**Defining local pelvic recurrence (21 disease, 26 no disease)** PET+CT: Se=90%*; Sp=100%; PPV=100%; NPV=93%; Acc=96% CT: Se=57%*; Sp=81%; PPV=71%; NPV=70%; Acc=70% *(P = 0.008) • PET correctly identified presence of disease in all patients with true positive CT findings • PET was useful in differentiating postoperative fibrosis from recurrence in 6 patients with positive CT scans • PET confirmed disease in 4 patients with equivocal CT findings • 2 false negatives on both PET and CT were diffuse mesorectal and anastomotic histologies proven by transrectal US-guided biopsies. **Defining hepatic metastases (23 disease, 35 no disease)** PET+CT: Se=96%*; Sp=100%; PPV=100%; NPV=97%; Acc=98% CT: Se=74%*; Sp=86%; PPV=77%; NPV=83%; Acc=81% *(P = 0.02) • PET identified all 5 patients with solitary metastases, CT identified 2 patients with solitary mets • PET identified 17/18 patients and CT identified 10/18 patients with multiple lesions • One false negative on both CT and PET found to be multiple superficial hepatic lesions up to 3 cm in diameter **Defining extrahepatic metastases (20 disease, 38 no disease)** PET identified extra-hepatic metastases in 21 sites in 20 patients, of which 9 lesions were missed on CT or CXR _____ **Therapeutic efficacy** PET influenced clinical management in 47% (10/21) patients with local recurrent disease, 43% (10/23) with hepatic metastases, and 38% (8/20) with extrahepatic metastases

Study	Patients/Methods	Results/Comments
Flanagan et al. (1998) *(Washington Univ. School of Medicine, St. Louis, Missouri)*	**Purpose** To retrospectively assess PET in patients with unexplained rising carcinoembryonic antigen (CEA) levels after treatment of colorectal cancer **Cases** 22 of 128 patients with a history of colorectal carcinoma who underwent PET from 6/93 to 6/96, were enrolled and were potential candidates for exploratory laparotomy: • all had plasma CEA level > 5.0 ng/ml (mean 25 ng/ml), normal imaging studies, endoscopy, and physical exam on routine follow-up • all patients had normal CEA levels after resection of their primary tumors • 17 men, 5 women; ages 17-84 • Primary site: colon (9), rectum (10), rectosigmoid (2), appendix (1) • Stage B (10); Stages C (5), C1 (2), C2 (3); Stage D (2) **Methods** • Patients with history of rectal or rectosigmoid carcinoma had contrast CT of chest, abdomen, and pelvis • Patients with history of colon cancer had contrast CT of abdomen and pelvis • CT scans performed ≤ 4 weeks before PET • CT interpreted in "routine clinical fashion" • PET interpreted qualitatively in "routine clinical fashion", including correlating with CT, and by consensus of at least two readers • PET used in treatment management at the discretion of the referring surgeon • PET correlated with histology, long term radiologic and clinical follow-up ≥ 6 months • PET true positive=confirmation by biopsy or obvious disease site on follow up imaging directed by PET and within 6 months of PET • PET true negative= confirmation by biopsy or no abnormality verified by other imaging or clinical follow-up within 6 months of PET **Limitations of study design** • Retrospective analysis • High probability of malignancy (potential referral bias) • Methods for image interpretation unclear and readers not blinded to CT(test review bias) • Blinding of PET results and reference standard not reported; strong correlation between test results and gold standard determination (diagnostic review bias) • Methods for systematic assessment of therapeutic efficacy not reported	**Detecting recurrent disease (15 recurrence, 7 no recurrence)** PET: Se=100%; Sp=71%; PPV=89%; NPV=100% • 2 false positives due to asymmetric activity in bowel and bladder diverticulum, and increased uptake in dome of liver in a patient in whom a poor quality PET scan was produced due to large patient size --- **Therapeutic efficacy** • Guided by the PET results, curative surgery was attempted in only 4 or 15 patients with disease • Neither false positives on PET resulted in mismanagement; both patients had equivocal findings, and referring physicians opted for additional radiologic and follow up studies • All 5 patients with negative PET scans were alive and disease free 9-24 months after PET; 2 patients had negative biopsy of anastomotic site, other 3 patients had no disease progression on follow up **Note: overlapping patient populations with previous study*

XVIII. APPENDIX 5: Technology Assessments of PET Produced by Other Organizations

Date	Issuing Agency	Topic/Title	Methods	Findings/comments
1997	Blue Cross and Blue Shield Association	PET for Managing Medically Refractory Partial Seizures	Systematic review and Medical Advisory Panel	proprietary
1997	Blue Cross and Blue Shield Association	PET in the Diagnosis and Management of Brain Tumors	Systematic review and Medical Advisory Panel	proprietary
1997	Blue Cross and Blue Shield Association	PET for the Assessment of Cerebrovascular Disease	Systematic review and Medical Advisory Panel	proprietary
1997	Agencia de Evaluación de Tecnologías Sanitarias (AETS) Spain	PET for non-CNS tumors	Systematic Review ***(uses VHA methods and frameworks, for studies through 1996)***	Spanish text with English abstract available. • Evidence is of poor methodological quality, small and uncontrolled studies • Potentially a good alternative for lung cancer staging and solitary pulmonary nodules • Relative contribution of PET in the management of patients with cancer is inconclusive • PET is considered a technology under investigation.
1998	Blue Cross and Blue Shield Association	PET with FDG for non-CNS cancer	Systematic reviews and Medical Advisory Panel	proprietary, but abstract available on-line • PET with FDG for staging lung cancer and imaging patients with a solitary pulmonary nodule that cannot be determined malignant by X-ray or CT (provided the results of the test could change the patient's medical management) meet the BC/BS Association's TEC criteria. • FDG-PET of other non-CNS cancers studies do not meet TEC criteria. (Specifically, treatment monitoring for lung cancer, detection, staging or monitoring breast cancer, pancreatic cancer, colorectal cancer, head and neck cancer, lymphoma, melanoma, musculoskeletal cancers, thyroid cancer, ovarian cancer, hepatocellular carcinoma, parathyroid cancer, thymoma, prostate cancer, germ-cell cancer, or esophageal cancer.)
1998	Hayes, Inc.	PET for diagnosing and staging lung cancer, for cardiac applications, neurologic applications, CNS tumors, non-CNS head and neck tumors, other malignancies	Systematic reviews	proprietary

Date	Issuing Agency	Topic/Title	Methods	Findings/comments
1998	ECRI	Cost-effectiveness analysis of PET in lung cancer diagnosis and staging	Meta-analysis and decision analysis	Proprietary, abstract available on-line • PET added to the diagnostic algorithm is cost-effective for patients with proven lung cancer to confirm resectability, but is not cost-effective when used earlier in the diagnostic algorithm • diagnosing SPN decreased life expectancy and increased costs, compared to the reference strategy • SPN strategy using CT for initial diagnosis, needle biopsy to confirmation positive results, and PET to confirm negatives attained greatest life expectancy
1998	Basque Office for Health Technology Assessment (OSTEBA) Spain	The clinical utility of PET	Narrative review	English text not available
1998	Medical Technology Section Swiss Federal Office of Social Security (SFOSS), Switzerland	PET use at two Swiss hospitals	Evaluation registry	English text not available
1998	Alberta Heritage Foundation for Medical Research (AHFMR), Canada	Functional diagnostic imaging in epilepsy	Systematic review	• PET has advantages over existing functional imaging methods in terms of accuracy of localization of lesions in patients with MRE. However, it has not yet been able to replace other technologies, and is not helpful for many patients with non-temporal lobe epilepsy. • Of the functional diagnostic imaging methods considered, only PET has a potential place in routine management of some epilepsy patients. Further work would be needed to define its role and economic costs and benefits.
1998	Center for Practice and Technology Assessment, Agency for Health Care Policy and Research (AHCPR), USA	FDG-PET scans for the localization of epileptogenic foci	Systematic review	• PET, SPECT and invasive EEG have been used at various epilepsy centers to identify additional candidates who might benefit from curative epilepsy surgery • FDG-PET scans show hypometabolic areas concordant with epileptogenic foci indicated by other diagnostic tests such as EEG and MRI. PET also showed discordant results in many patients with EEG-indicated epileptogenic foci. • Available data were insufficient to determine whether PET scans might reliably substitute for EEG, or to determine the contribution of confirmatory PET scans to the management of patients with complex partial seizures
1998	Committee for Evaluation and Diffusion of Innovative Technologies (CEDIT), France	FDG-PET and Cdet (coincidence detection emission tomography) imaging in Assistance Publique-Hôpitaux de Paris (AP-HP)	Expert panel	French text with English abstract: • Assessment addressing technical aspects, clinical uses, economics, regulatory issues, and recommendations from perspective of AP-HP system: • Literature supports positron imaging in prostatic cancer and has potential value in at least four areas: bronchopulmonary cancer, colorectal cancer, lymphoma, and breast cancer • CEDIT recommends establishing a PET center for AP-HP cancer patients and making Cdet available for routine oncological use. And funding an evaluation comparing the effectiveness and diagnostic contribution of PET and Cdet in pre-operative staging patients with lung cancer

Date	Issuing Agency	Topic/Title	Methods	Findings/comments
1999	NHS Health Technology Assessment Programme for the Medical Research Council, NHS UK	Potential role of PET in the NHS and establishing key health tech. assessment questions relating to PET in the UK	Systematic review and Delphi survey ***(Updates and expands VHA review using VHA methods and frameworks, for studies through January 1998)***	Under Council review • Includes myocardial and neuropsychiatric applications and all positron imaging modalities • evidence relating to diagnostic accuracy limited by bias and often relate only to small patient numbers • evidence is needed on the cost-effectiveness of positron imaging modalities in all of the advocated clinical indications Research priorities identified in descending order: • relative cost-effectiveness of full ring PET and gamma camera PET to pre-operatively stage patients with lung cancer • Compare partial ring to full ring PET in oncology • Relative cost-effectiveness of full ring PET and gamma camera PET to stage and monitor treatment response in patients with breast cancer • relative cost-effectiveness of gamma PET to collimated 511 keV positron imaging for selecting patients for myocardial revascularization surgery
1999	AETS, Spain	PET in neurology	Systematic review	In preparation

References

Asua J, Hurtado de Saracho I. *Tomografía por emisión de positrones: su utilidad clínica.* 1998, Osasunerako Teknologien Ebaluaketa (Osteba): Vitoria-Gasteiz, Spain.

Charpentier E, Fay A-F. *Utilisation des cameras TEDC pour la scintigraphie conventionelle.* 1998, Comite d'Evaluation et de Diffusion des Innovations Technologiques (CEDIT): Paris, France.

Comite d'Evaluation et de diffusion des Innovations Technologiques (CEDIT). *La tomographie par emission de positrons: rapport du groupe de travail.* 1998, Comite d'Evaluation et de Diffusion des Innovations Technologiques (CEDIT): Paris, France.

Corabian P, Hailey D. *Functional diagnostic imaging in epilepsy (Draft).* 1998, Alberta Heritage Foundation for Medical Research (AHFMR): Edmonton, Alberta, Canada.

Ministerio de Sanidad y Consumo. Instituo de Salud "Carlos III". Agencia de Evaluación de Tecnologías Sanitarias (AETS). *Tomografía por emisión de positrones en cardiología (PET).* 1995, Ministerio de Sanidad y Consumo. Instituo de Salud "Carlos III". Agencia de Evaluación de Tecnologías Sanitarias (AETS): Madrid, Spain.

Ministerio de Sanidad y Consumo. Instituo de Salud "Carlos III". Agencia de Evaluación de Tecnologías Sanitarias (AETS). *Tomografía por emisión de positrones (PET) en oncología clínica no neurológica.* 1997, Ministerio de Sanidad y Consumo. Instituo de Salud "Carlos III". Agencia de Evaluación de Tecnologías Sanitarias (AETS): Madrid, Spain.

Mitchell MD, Turkelson C, Doggett D. *Cost-effectiveness analysis of PET in lung cancer diagnosis and staging*. Annual Meeting of the International Society for Technology Assessment in Health Care, *Abstracts.* Vol. 14; 33, 1998.

Robert G, Milne R. *Positron emission tomography: establishing priorities for health technology assessment (under council review).* 1999, University of Southampton. National Co-ordinating Centre for Health Technology Assessment: Bassett Crescent East, Southampton.

U.S. Department of Health and Human Services. Health Care Financing Administration. Technology Advisory Committee. (September 29, 1997). *Technology Advisory Committee Minutes, August 5 and 6, 1997.* [Web site]. Available: **http://www.hcfa.gov/events/0897tmin.htm,** [May 3, 1999].