



**T e c h n o l o g y
A s s e s s m e n t
P r o g r a m**

Office of Patient Care Services

UPDATED INFORMATION FOR VA TECHNOLOGY ASSESSMENT PROGRAM (VATAP) REPORTS

In June 2000, VATAP was relocated within the Veterans Health Administration from the Office of Research & Development to the Office of Patient Care Services. The following report was produced prior to the relocation of VATAP.

Current VATAP contact information is as follows:

VA Technology Assessment Program (11T)

VA Boston Healthcare System

150 South Huntington Avenue

Boston, MA 02130

Tel: 617.278.4469 Fax: 617.264.6587

vatap@med.va.gov

<http://www.va.gov/vatap> <http://vawww.va.gov/vatap>

Positron Emission Tomography:

- Descriptive Analysis of Experience with PET in VA
- Systematic Reviews: FDG-PET as a Diagnostic Test for Cancer and Alzheimer's Disease

October, 1996

Primary Authors:

Karen Flynn, D.D.S., M.S., *Manager*

Elizabeth Adams, R.R.T., M.P.H., *Management & Program Analyst*

Contributing Author:

Diana Anderson, R.N., M.P.H., *Research Analyst*

Technology Assessment Program

Management Decision and Research Center
Health Services Research and Development Service
Office of Research and Development
Department of Veterans Affairs

Karen Flynn, D.D.S., M.S., Manager
Technology Assessment Program
Management Decision and Research Center
VA Medical Center (152M)
150 South Huntington Avenue
Boston, MA 02130

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

TABLE OF CONTENTS

Executive Summary

I.	BACKGROUND	2
II.	METHODS	3
	A. Overview of the assessment methodology	3
	B. Systematic reviews in technology assessment	4
	C. Search strategies	5
	D. Systematic review approach and protocol	6
	E. Meta analysis considered for studies of PET diagnostic accuracy	9
	F. Selected alternatives to PET	9
	G. Review of the assessment report	9
III.	RESULTS: Site visits and surveys	10
	A. Characteristics of sites	10
	B. Activity at each PET site	10
	C. Barriers and incentives to the use of PET	11
IV.	RESULTS: Systematic reviews	13
V.	RISKS ASSOCIATED WITH PET RADIOPHARMACEUTICALS	16
VI.	FDA STATUS OF PET RADIOPHARMACEUTICALS	16
VII.	CONCLUSIONS	17
VIII.	REFERENCES	27
<i>Table 1</i>	<i>Summary of the literature: Diagnostic accuracy efficacy of PET and alternatives in head and neck cancer</i>	19
<i>Table 2</i>	<i>Summary of the literature: Diagnostic accuracy efficacy of PET and alternatives in colorectal cancer</i>	20
<i>Table 3</i>	<i>Summary of the literature: Diagnostic accuracy efficacy of PET and alternatives in breast cancer</i>	22
<i>Table 4</i>	<i>Summary of the literature: Diagnostic accuracy efficacy of PET and alternatives in lung cancer</i>	23
<i>Table 5</i>	<i>Summary of the literature: Diagnostic accuracy efficacy of PET and alternatives in solitary pulmonary nodules</i>	25
<i>Table 6</i>	<i>Summary of the literature: Diagnostic accuracy and diagnostic thinking efficacy of PET and its neuroimaging alternatives</i>	26

Appendix 1 Advisory Committee to the PET Assessment

Appendix 2 Assessing Diagnostic Technologies

I.	BACKGROUND	A2 - 2
II.	CONDUCTING STUDIES TO EVALUATE DIAGNOSTIC TEST ACCURACY	A2 - 2
III.	MEASURES OF THE ACCURACY OF DIAGNOSTIC TESTS	A2 - 4
	A. Is disease present or absent?	A2 - 4
	B. What is the severity of disease?	A2 - 6
IV.	INTERPRETING RESULTS AFTER AN ACCURATE TEST HAS BEEN SELECTED AND PERFORMED	A2 - 6
V.	ANALYTIC FRAMEWORK FOR MDRC TECHNOLOGY ASSESSMENT PROGRAM SYSTEMATIC REVIEWS OF DIAGNOSTIC TEST LITERATURE ...	A2 - 7
	A. What is the quality of the individual studies that were intended to measure the technology's characteristics as a diagnostic test?	A2 - 8
	B. Where does an individual study fall in the hierarchy of diagnostic efficacy?	A2 - 9
	C. How strong is the evidence supporting a causal link between the use of the technology and improved outcomes of care?	A2 - 10
VI.	SYSTEMATIC REVIEW PROTOCOL	A2 - 10
VII.	REFERENCES	A2 - 16
VIII.	GLOSSARY	A2 - 18
	<i>Table 1</i>	A2 - 13
	<i>Table 2</i>	A2 - 15

Appendix 3 Systematic review: PET as a Diagnostic Test in Head and Neck Cancer

I.	BACKGROUND	A3 - 3
	A. General sources	A3 - 3
	B. Description	A3 - 3
	C. Epidemiology	A3 - 3
	D. Diagnosis	A3 - 3
	E. Staging, treatment, and survival	A3 - 4
	F. Potential roles for PET	A3 - 6
II.	RESULTS	A3 - 7
	A. Detecting unknown primaries in patients who present with metastatic cervical nodes	A3 - 8
	B. Detecting primary disease	A3 - 8
	C. Detecting cervical metastases	A3 - 8
	D. Detecting recurrent disease	A3 - 9
III.	SUMMARY	A3 - 9
IV.	DISCUSSION	A3 - 9
V.	SUGGESTIONS FOR FURTHER RESEARCH	A3 - 10
VI.	REFERENCES: BACKGROUND AND DIAGNOSTIC ACCURACY STUDIES	A3 - 16

VII.	REFERENCES: TECHNICAL EFFICACY STUDIES	A3 - 17
VIII.	REFERENCES: EXCLUDED STUDIES	A3 - 17
Table 1	<i>Tumor, node, metastases staging system for head and neck cancer</i>	A3 - 4
Table 2	<i>Head and neck cancer treatment and survival by stage</i>	A3 - 5
Table 3	<i>Summary of the literature: PET diagnostic accuracy studies in head and neck cancer</i>	A3 - 11
Table 4	<i>Data abstraction table: Diagnostic accuracy efficacy studies</i>	A3 - 12
Table 5	<i>Data abstraction table: Hypothetical therapeutic efficacy study</i>	A3 - 15

Appendix 4 Systematic review: PET as a Diagnostic Test in Colorectal Cancer

I.	BACKGROUND	A4 - 3
	A. General sources	A4 - 3
	B. Epidemiology	A4 - 3
	C. General description	A4 - 3
	D. Staging, treatment, and survival	A4 - 3
	E. Follow-up after primary treatment	A4 - 4
	F. Potential roles for PET	A4 - 6
II.	RESULTS	A4 - 6
III.	ALTERNATIVES TO PET AND DISCUSSION	A4 - 7
IV.	SUMMARY	A4 - 8
V.	SUGGESTIONS FOR FUTURE RESEARCH	A4 - 11
VI.	REFERENCES: General background and diagnostic accuracy efficacy studies	A4 - 21
VII.	REFERENCES: Technical efficacy studies	A4 - 23
VIII.	REFERENCES: Studies reviewed but not included in evidence tables	A4 - 23
Table 1	<i>Modified Dukes classification of colorectal cancer, standard treatment options, and survival</i>	A4 - 5
Table 2	<i>Summary of the literature: Diagnostic accuracy efficacy of PET and alternatives in colorectal cancer</i>	A4 - 9
Table 3	<i>Data abstraction table: PET diagnostic accuracy efficacy studies</i>	A4 - 12
Table 4	<i>Data abstraction table: PET therapeutic accuracy efficacy studies</i>	A4 - 15
Table 5	<i>Data abstraction table: Diagnostic accuracy efficacy of alternative technologies to PET</i>	A4 - 17

Appendix 5 Systematic review: PET as a Diagnostic Test in Breast Cancer

I.	BACKGROUND	A5 - 3
	A. General sources	A5 - 3
	B. Description	A5 - 3
	C. Epidemiology	A5 - 3
	D. Diagnosis	A5 - 4
	E. Staging, treatment, and survival	A5 - 4
	F. Potential roles for PET	A5 - 5

II.	RESULTS	A5 - 5
	A. Defining primary breast disease	A5 - 6
	B. Defining axillary lymph node involvement	A5 - 7
	C. Detecting distant metastases	A5 - 7
III.	SUMMARY	A5 - 7
IV.	DISCUSSION	A5 - 7
	A. Alternatives to PET in some of its potential breast cancer applications	A5 - 8
	B. A breast cancer research agenda	A5 - 9
V.	SUGGESTIONS FOR FURTHER RESEARCH	A5 - 10
VI.	REFERENCES: Background and studies meeting evidence-based medicine criteria for evaluations of diagnostic tests	A5 - 18
VII.	REFERENCES: Technical efficacy studies	A5 - 19
VIII.	REFERENCES: Excluded studies	A5 - 20
<i>Table 1</i>	<i>Breast cancer staging, treatment and survival by stage</i>	A5 - 11
<i>Table 2</i>	<i>Summary of the literature: Diagnostic accuracy studies of PET and alternatives in breast cancer</i>	A5 - 13
<i>Table 3</i>	<i>Data abstraction table: Diagnostic accuracy efficacy studies</i>	A5 - 14

Appendix 6 Systematic review: PET as a Diagnostic Test in Lung Cancer

I.	BACKGROUND	A6 - 3
	A. General sources	A6 - 3
	B. Description	A6 - 3
	C. Epidemiology	A6 - 3
	D. Diagnosis	A6 - 4
	E. Staging, treatment, and survival	A6 - 4
	F. Potential roles for PET	A6 - 6
II.	RESULTS	A6 - 8
	A. Detecting unknown primary disease	A6 - 9
	B. Detecting hilar and mediastinal metastases	A6 - 9
	C. Detecting recurrent disease	A6 - 10
III.	SUMMARY	A6 - 10
IV.	DISCUSSION	A6 - 10
V.	SUGGESTIONS FOR FURTHER PET RESEARCH	A6 - 12
VI.	REFERENCES: Background and studies meeting evidence-based medicine criteria for evaluations of diagnostic tests	A6 - 23
VII.	REFERENCES: Technical efficacy studies	A6 - 26
VIII.	REFERENCES: Excluded studies	A6 - 27
<i>Table 1</i>	<i>Lung cancer TNM staging system</i>	A6 - 4
<i>Table 2</i>	<i>Lung cancer staging, treatment and survival</i>	A6 - 5
<i>Table 3</i>	<i>Summary of the literature: PET diagnostic accuracy studies in lung cancer</i>	A6 - 13

Table 4 Data abstraction table: Diagnostic accuracy efficacy of PET in lung cancer A6 - 15

Table 5 Data abstraction table: Hypothetical diagnostic thinking efficacy of PET in lung cancer A6 - 22

Appendix 7 Systematic review: PET as a Diagnostic Test in Solitary Pulmonary Nodules

I. BACKGROUND A7 - 3

 A. General sources A7 - 3

 B. Description A7 - 3

 C. Epidemiology A7 - 3

 D. Diagnosis A7 - 3

 E. Staging, treatment, and survival A7 - 5

 F. Potential roles for PET A7 - 6

II. RESULTS A7 - 7

 A. Characterizing indeterminate solitary pulmonary nodules A7 - 7

III. SUMMARY A7 - 9

IV. DISCUSSION A7 - 9

V. SUGGESTIONS FOR PET FURTHER RESEARCH A7 - 10

VI. REFERENCES: Background and studies meeting evidence-based medicine criteria for evaluations of diagnostic tests A7 - 17

VII. REFERENCES: Technical efficacy studies A7 - 18

VIII. REFERENCES: Excluded studies A7 - 19

Figure 1 Algorithm for the management of solitary pulmonary nodules A7 - 4

Table 1 Lung cancer tumor, node, metastases staging system A7 - 6

Table 2 Summary of the literature: Diagnostic accuracy of PET in solitary pulmonary nodules A7 - 12

Table 3 Data abstraction table: Diagnostic accuracy efficacy studies A7 - 13

Table 4 Data abstraction table: Diagnostic thinking efficacy studies A7 - 15

Appendix 8 Systematic review: PET as a Diagnostic Test in Alzheimer’s Disease

I. BACKGROUND A8 - 2

 A. Description A8 - 2

 B. Epidemiology A8 - 2

 C. Diagnosis A8 - 3

 D. Treatment A8 - 5

 E. Rationale for PET in AD A8 - 6

 F. Special considerations in evaluating a diagnostic test for Alzheimer’s disease A8 - 6

 G. Alternative neuroimaging technologies for diagnosing AD A8 - 7

 H. Ethical considerations in testing for AD A8 - 9

II. RESULTS A8 - 9

III. SUMMARY A8 - 11

IV. DISCUSSION A8 - 12

V.	CONCLUSIONS: Clinical use of PET in Alzheimer's disease	A8 - 13
VI.	REFERENCES: Background and diagnostic accuracy/diagnostic thinking efficacy studies	A8 - 22
VII.	REFERENCES: Technical efficacy studies	A8 - 25
VIII.	REFERENCES: Excluded studies	A8 - 26
Table 1	<i>NINCDS-ADRDA criteria for clinical diagnosis of Alzheimer's disease</i>	A8 - 4
Table 2	<i>Tests for dementia and Alzheimer's disease</i>	A8 - 5
Table 3	<i>Summary of diagnostic accuracy and diagnostic thinking efficacy studies for PET and neuroimaging alternatives</i>	A8 - 11
Table 4	<i>Data abstraction table: PET diagnostic accuracy efficacy studies</i>	A8 - 14
Table 5	<i>Data abstraction table: PET and alternatives diagnostic accuracy efficacy studies</i>	A8 - 18
Table 6	<i>Data abstraction table: Study design models for diagnostic thinking efficacy</i>	A8 - 20

Appendix 9 Experience With PET in VHA

I.	BACKGROUND	A9 - 2
II.	METHODS	A9 - 2
III.	RESULTS	A9 - 2
	A. Characteristics of interview subjects	A9 - 3
	B. Characteristics of PET centers	A9 - 3
	C. Types and volumes of PET studies	A9 - 12
	D. Costs	A9 - 19
	E. Barriers and incentives to PET use	A9 - 19
IV.	SUMMARY	A9 - 40
Table 1	<i>Site Visit Interview Subjects According to Specialty, Job Role, and Referral Status</i>	A9 - 4
Table 2	<i>Summary of Site Visit Interview Subjects According to Specialty</i>	A9 - 8
Table 3	<i>Summary of Site Visit Interview Subjects According to Referral Patterns</i>	A9 - 8
Table 4	<i>Summary of Site Visit Interview Subjects According to Job Role</i>	A9 - 8
Table 5	<i>A Comparison of Ancillary Services Offered at Each VHA PET Site</i>	A9 - 9
Table 6	<i>General Information of VHA PET Sites as of Fiscal Year 1994</i>	A9 - 10
Table 7	<i>Summary of the General Characteristics of the VHA PET Sites</i>	A9 - 11
Table 8	<i>Patient Volume at VHA PET Sites for Fiscal Year 1994</i>	A9 - 13
Table 9	<i>A Comparison of VA to non-VA Patient Volume Within Each Clinical and Research Application Across All VHA PET Sites for Fiscal Year 1994</i>	A9 - 14
Table 10	<i>Patient Volume at VHA PET Sites for Fiscal Year 1993</i>	A9 - 15
Table 11	<i>A Comparison of VA to non-VA Patient Volume Within Each Clinical and Research Application Across All VHA PET Sites for Fiscal Year 1993</i>	A9 - 16
Table 12	<i>Follow-up Survey of Activity at VHA PET Sites for Fiscal Year 1995</i>	A9 - 17
Table 13	<i>Results of Site Visit Interviews Reflecting Major Barriers and Incentives to the Use of PET Within Each Site</i>	A9 - 24

Table 14 Recommendations Volunteered During VHA PET Site Visit Interviews A9 - 30

Table 15 Best Practices Identified at VHA PET Sites A9 - 33

Table 16 Research Activity at VHA PET Sites as of October 1994 A9 - 34

Figure 1 Locations of VHA PET Centers A9 - 41

Survey Instruments A9 - 42

Appendix 10 Assessments, Guidelines, and Policy Statements Produced by Other Agencies

Acknowledgements

The MDRC Technology Assessment Program wishes to thank the members of the Advisory Committee to the PET Assessment, who contributed to the design of the assessment, confirmed the completeness of literature retrieval, reviewed the report, and reached consensus on its recommendations.

Chair: *Marguerite Hays, M.D.*, ACOS Research Service, Palo Alto VA Medical Center

Members: *Martin Charns, D.B.A.*, Director, Management Decision and Research Center

Daniel Deykin, M.D., Director, Health Services Research and Development Service

James Fletcher, M.D., Chief, Nuclear Medicine Service, St. Louis VA Medical Center

Clifford Goodman, Ph.D., Consultant

Milton Gross, M.D., Program Director, VA Nuclear Medicine Service

Thomas Holohan, M.D., Director, Office of Health Technology Assessment, Agency for Health Care Policy and Research

Steven Hotta, M.D., Ph.D., Project Manager for PET, Office of Health Technology Assessment, Agency for Health Care Policy and Research

Steven Larson, M.D., Chief, Nuclear Medicine Service, Memorial Sloane Kettering Hospital

H. William Strauss, M.D., Division of Nuclear Medicine, Stanford University School of Medicine

The MDRC Technology Assessment Program thanks the physicians who suggested improvements to the disease-specific systematic reviews and other sections of the report, and acknowledges their contributions:

Alan Garber, M.D., Ph.D., Associate Director of Health Economics and Technology Assessment, Center for Health Care Evaluation, VA Palo Alto Health Care System

Methods

Thomas Holohan, M.D., Chief Patient Care Services Officer, Department of Veterans Affairs

Head and neck, colorectal, and breast cancer

L. Jack Faling, M.D., Associate Chief, Medical Service, VA Medical Center, Boston

Lung cancer
Solitary pulmonary nodules

Charles Powell, M.D., Staff Physician, Boston University Pulmonary Center and VA Medical Center, Boston

John Booss, M.D., Director of Neurology, Department of Veterans Affairs

Alzheimer's disease

Thomas Bird, M.D., Chief, Neurology Service, Seattle VA Medical Center, and Professor of Neurology, University of Washington

Jeffrey Cummings, M.D., Professor of Neurology and Psychiatry, University of California at Los Angeles, and Neurology Service, West Los Angeles VA Medical Center

Judith Salerno, M.D., Chief Consultant for Geriatrics and Extended Care, Department of Veterans Affairs

The MDRC Technology Assessment Program also wishes to thank Lois Camberg, Ph.D., Mark Prashker, M.D., M.P.H., Elizabeth Kramer, and the many VA and university affiliate physicians and other staff who participated in the interviews, site visits to PET centers, and surveys.

Technology Assessment:
Positron Emission Tomography

Executive Summary

The Veterans Health Administration (VHA) shares, with some of its academic affiliates, the ownership and operation of 10 positron emission tomography (PET) imaging facilities. Significant resource commitments are associated with the acquisition, maintenance, and ongoing operation of these facilities. In late 1993, the Acting Under Secretary for Health, Department of Veterans Affairs (VA), requested that the Management Decision and Research Center (within the Health Services Research and Development Service) conduct an assessment of PET. The assessment would supply the Under Secretary with information that would assist in setting future VHA policy regarding PET. The Acting Under Secretary asked two questions:

What is known about the utilization of PET, and other experience with the technology, in VHA today?

Should VHA establish additional PET centers?

This document reports the results of the assessment. The overall approach and findings of the assessment are presented in this summary section. The appendices detail the individual components of the assessment and provide background to the development of the assessment methodology.

The Technology Assessment Program of the Management Decision and Research Center (MDRC) focuses on evaluating the clinical applications (rather than the technical performance or technical specifications) of health care technologies, using systematic reviews of published evidence supplemented by primary data collection. The Program uses the broad definition of health care technology developed in 1978 by the Office of Technology Assessment:

“... the drugs, devices, and medical and surgical procedures used in health care, and the organizational and supportive systems within which such care is delivered.”

and the Institute of Medicine’s 1985 definition of technology assessment:

“...any process of examining and reporting properties of a medical technology used in health care, such as safety, efficacy, feasibility, and indications for use, cost, and cost-effectiveness, as well as social, economic, and ethical consequences, whether intended or unintended.”

The purpose of technology assessment is to inform technology-related policy making in health care.

I. BACKGROUND

Positron emission tomography (PET) is a nuclear medicine technology that allows the visualization and measurement of biochemical processes within tissues. PET's particular functional imaging capacity is related to the physics of the positron emission detection method and to the variety of radiolabelled compounds that can be used.

Nuclear medicine imaging techniques rely on the detection of photons produced from the decay of radioactive isotopes attached to tracers that target physiologic processes (Gritters and Wahl, 1993). PET, like other nuclear medicine techniques, makes it possible to measure local tissue and organ function, re-defining disease in terms of quantifiably abnormal regional chemistry. PET and other nuclear medicine imaging therefore may complement the information obtained from other imaging methods, such as radiography, computed tomography (CT), or magnetic resonance imaging (MRI), which rely on predominantly anatomic definitions of disease (Maisey and Jeffery, 1991).

Traditional nuclear imaging is based on photon detection using a stationary single or double-headed gamma camera that produces two-dimensional images. Tomographic techniques (single photon emission tomography, SPECT) may use mechanically rotating camera heads to acquire many pictures in a 360° circle around the patient. The imaging data are then reconstructed to produce multiple cross-sectional images.

Most radioactive isotopes with potential uses in medical imaging decay by releasing energy as single gamma rays (photons) whose energies fall within a range from 80 to 400 KeV. The relatively low energy of the photon released during SPECT imaging means that attenuation and scatter by tissues can degrade the image.

The radioactive isotopes used in PET decay by other means: they emit a positively charged electron (positron) from the nucleus. The positron usually travels only a very short distance (1 to 2 millimeters) before colliding with a local electron. The collision results in the annihilation of the mass of the two particles, and the emission of two gamma rays (photons) of high energy (511 KeV), which travel out at approximately 180° from each other. Radiation detectors in a PET camera, which are arranged in a ring around the patient, detect the two gamma rays from each such collision simultaneously. The exact site of origin of each signal is recorded, and a cross-sectional image is displayed.

The high energy of the photon released during PET imaging means that very little of that energy is attenuated or scattered by tissue. Other sources of scatter are minimized by coincidence counting (the recording of only those photons which have been emitted at 180° from each other and hit opposing crystals in the camera simultaneously).

All medical imaging involves comparisons: of an image with the interpreter's mental pictures of the patterns representative of "normal" and of different disease states; or of changes in sequential images from the same patient (Links and Devous, 1994). PET and other nuclear medicine image patterns represent spatial and temporal arrangements and rearrangements of the physiological or biochemical process under investigation. A variety of ways to detect and compare these patterns are illustrated by the literature that will be summarized in this document and reviewed in detail in the Appendices. Pattern detection approaches include: visual analysis of patterns of metabolism; region of interest (ROI) analysis where the regions are hand-drawn or placed (sometimes with co-registration with anatomic images); and neural networks.

Kippenhan, et al. (1992), report that much of PET research involves improving the performance of particular links in the chain of highly complex data transformation that results in regional metabolic representations. Approaches to PET data management may include: normalization to a reference value (e.g., in brain studies to global brain metabolic rate or to an anatomic reference area that is relatively unaffected by the disease process) to generate metabolic ratios; or the use of absolute

metabolic values. Links and Devous (1994) suggest that (for brain studies) the effect of normalization on diagnostic results may be dependent on region of interest (ROI) size, tomographic resolution, and biological and technical variation in the data and the type of normalization.

PET has been recognized as a valuable basic research tool during its approximately 20 years of development. Clinical diagnostic applications for PET are now emerging, particularly in the areas of neurology, cardiology, and oncology. Constructing and equipping, maintaining, and supporting PET facilities are resource-intensive activities requiring high levels of medical, technical, and managerial expertise. In the context of the widely recognized need to use available health care resources to maximize quality of care and achieve optimal patient outcomes, there is a compelling rationale for evaluating the clinical applications of PET as they emerge, and for applying evaluation results to policies regarding PET (Chalmers, 1988; Cooper, et al., 1988; Powers, et al., 1991; Hoffman, et al., 1992).

II. METHODS

A. Overview of the assessment methodology

The MDRC Technology Assessment Program convened a PET Advisory Committee, whose members are listed in *Appendix 1*, to focus the assessment. The Acting Under Secretary's question on experience with PET within VA was addressed by conducting surveys and site visits of VA PET centers to collect information on PET imaging utilization, center operations, and research activities. The results of this component of the assessment are outlined in *Appendix 9*.

PET has potential clinical applications in six conditions identified by the Advisory Committee as being of particular importance to the veteran population. These conditions are: solitary pulmonary nodules; lung cancer; head and neck cancers; breast cancer; colorectal cancer; and Alzheimer's disease. In the clinical management of these conditions, PET is applied as a diagnostic test.

One rationale for VA to invest in additional PET centers would be to make clinically useful PET studies more widely accessible to veterans. To respond to the Acting Under Secretary's question regarding whether to acquire additional PET capacity, a systematic review of research articles published in peer reviewed medical journals was used to evaluate what is known about the usefulness of PET in diagnosing diseases of importance to the veteran population. The systematic reviews (*Appendices 4 through 8*) addressed two additional queries:

Is PET an accurate diagnostic test when applied to patients with head and neck cancer, colorectal cancer, breast cancer, lung cancer/solitary pulmonary nodules, and Alzheimer's disease?

Does PET affect patient management decisions, outcomes of care, costs of care, or cost-effectiveness of care in head and neck cancer, colorectal cancer, breast cancer, lung cancer/solitary pulmonary nodules, or Alzheimer's disease?

The final literature database searches were performed on September 10, 1996; the assessment represents peer-reviewed literature published and indexed as of that date.

B. Systematic reviews in technology assessment

Expanding on the Banta and Luce (1993) systematic process for technology assessment, Goodman, et al. (1996) defined the following steps for conducting an assessment:

- identify assessment topics;
- specify the assessment problem;
- identify the locus for the assessment;
- retrieve evidence;
- collect new primary data (as necessary and appropriate);
- interpret the evidence;
- synthesize/consolidate the evidence;
- formulate findings and recommendations;
- disseminate findings and recommendations;
- monitor impact.

Banta and Luce (1993) note that synthesis is a critical part of the process. Synthesis involves a critical analysis of research results and other information, and often takes the form of judgments or recommendations. Synthesis is necessary to provide a responsible basis for decisions regarding the technology. Since policy makers are not generally trained in research study design and interpretation, raw data or unsynthesized results may be of little use to them. The purpose of synthesis is to make knowledge relevant to policy.

Synthesis provides focused, user-oriented information at a relatively low cost. If carefully performed, with attention to limitations of knowledge, synthesis can both guide technology-related decision making and help to define new research to answer important questions (Banta and Luce, 1993). However, the traditional narrative literature review has several shortcomings (Light and Pillemer, 1984; Mulrow, 1987 and 1994): the lack of formal rules for its conduct leads to subjectivity and bias; frequently used methods for synthesizing the results of multiple studies are inconsistent with good statistical practice; and it is an inefficient way to extract useful information.

Mulrow (1994) and other authors (e.g. Light and Pillemer, 1984; Slavin, 1986 and 1995) note that systematic reviews use a rigorous scientific approach and provide an alternative to traditional reviews. A systematic review frequently leads to different conclusions than does a traditional review of the same topic (Mulrow, 1994).

A systematic review or overview of the literature (Guyatt, et al., 1995):

- addresses a focused clinical question;
- uses appropriate criteria to select studies for inclusion;
- conducts a comprehensive search;
- appraises the validity of the individual studies in a reproducible fashion.

The purpose of a systematic review is to reduce unmanageable amounts of information to a form that is usable by decision makers, enabling health care decisions to be based on the best available evidence.

Systematic reviews can include both qualitative overviews of study findings and quantitative meta analyses of results. As currently understood, such reviews answer Slavin's call (1986; 1995) to use "best evidence synthesis" to avoid the shortcomings associated with both qualitative reviews and indiscriminately applied meta analyses. A "best evidence" systematic review combines the quantification of effect sizes and systematic study selection procedures of quantitative syntheses with the attention to individual studies and methodological and substantive issues typical of the best narrative reviews. These

reviews focus on the studies highest in internal and external validity, using well-specified and defended *a priori* inclusion criteria, and use effect size data as an adjunct to a full discussion of the literature being reviewed.

C. Search strategies

For each of the disease-specific systematic reviews conducted as part of this assessment of PET, literature was identified using formal search strategies. Comprehensive, multi-step search protocols were designed to ensure the broadest possible retrieval in each of the six disease areas: breast cancer, lung cancer, solitary pulmonary nodules, colorectal cancer, head and neck cancer, and Alzheimer's disease. Three searches for each disease were run on current files of the National Library of Medicine's MEDLINE®, and HEALTH® Planning databases for reviews of the literature, articles dealing with diagnosis, and for articles reporting the use of PET. Both free text words and MeSH subject headings were used to describe the concepts of interest and to ensure identification of the most comprehensive range of articles in the databases.

To ensure complete retrieval for the current period, when citations would not yet have appeared in the MEDLINE and HEALTH databases, searches were also performed on the ©Institute for Scientific Information's Current Contents® databases. Free text searches, using multiple synonyms, were employed for Current Contents searches.

Additional searches were performed for all of the cancers on the PDQ® Physicians' Data Query database (National Cancer Institute and National Library of Medicine). These non-bibliographic searches yielded information on diagnosis and staging of disease and on currently available treatment options; this information is incorporated into the background sections of the individual systematic reviews. All of the searches were refined according to the following rationale:

1. Early PET research in Alzheimer's disease used "first generation" scanners. These machines had limited spatial resolution, which contributed to potentially biased estimates of glucose metabolism due to partial volume effects (inclusion of cerebrospinal and subarachnoid spaces in the areas being analyzed for glucose metabolism) and the restriction of metabolic data to large neocortical areas. In addition, transmission scans were not used to correct for attenuation, venous blood was "arterialized" to estimate plasma radioactivity and glucose concentrations for metabolic rate calculations, and a highly subjective trace method was employed to determine regions of interest for analysis. Later generation scanners have improved resolution; the protocols used with these scanners correct for attenuation, collect arterial blood for metabolic calculations, and use devices to minimize patient movement during the relatively lengthy scanning procedures (Kumar, et al., 1991).

The rapid evolution of the technology also affected the use of PET in oncology, and supported the restriction of the searches to the years 1991 to 1995. Significant articles appearing before that period were identified by selected searches of the years 1986 to 1991, and from the reference lists of the articles retrieved.

2. Publication in a peer-reviewed journal was required. This decision was based on preliminary review of the quality of studies in peer-reviewed journals, many of which failed to meet criteria for avoiding bias in diagnostic test evaluations. It was felt that abstracts that had not been subjected to the peer review process necessary for publication would have a high probability of representing studies of equivalent or lesser quality, and would generally also fail to meet criteria for diagnostic test evaluations.

3. Only studies using 2-[F-18]-2-deoxy-D-glucose (FDG) PET were included, as the majority of studies in oncology and Alzheimer's disease use this radiopharmaceutical.

D. Systematic review approach and protocol

The first question addressed by the diagnosis-specific systematic reviews of the literature concerned the accuracy of PET as a diagnostic test. The validity of the estimates of accuracy supplied by published studies was evaluated by applying a set of study design and reporting criteria from the methodologic literature, codified in the review protocol.

Accurate estimation of the characteristics of a diagnostic test is one of the early steps in the assessment of that test. However, accuracy does not extrapolate automatically to clinical utility, and a complete assessment requires further research. The second question (regarding outcomes of care) addressed by the reviews was focused by assigning published studies to levels in a hierarchy of "diagnostic efficacy", and by applying quality criteria (based on accepted principles of research design) appropriate to each level.

The diagnostic efficacy hierarchy explicitly acknowledges the goal of diagnostic testing to be improving processes of care, outcomes of care, and efficiency of resource use. It outlines the progression of research into a new diagnostic technology from the initial studies documenting technical performance of the imaging device, through accuracy studies, to studies documenting changes in treatment decisions based on diagnostic information, changes in outcome, and finally to studies of societal efficacy (i.e., cost-effectiveness, cost-benefit, or cost-utility studies from a societal perspective).

Each of the disease-specific systematic reviews was conducted using the following protocol, which codifies the analytic frameworks presented in *Appendix 2: Assessing Diagnostic Technologies*.

Systematic Review Protocol

- 1) Conduct MEDLINE and other database searches; retrieve full text articles that meet screening criteria:
 - English language articles reporting primary data and published in a peer reviewed journal (not abstracts)
 - studies \geq 12 human subjects (not animal studies) with the disease of interest (sample sized defined by PET Advisory Committee)
 - studies using the radiopharmaceutical 2-[¹⁸F]fluoro-2-D-glucose (FDG)
- 2) Apply screening criteria to bibliographies of retrieved articles as above, and retrieve additional articles.
- 3) Review full text articles and assign to level of Fryback and Thornbury (1991) diagnostic efficacy hierarchy.
- 4) Assign to **technical efficacy** level of Fryback and Thornbury diagnostic efficacy hierarchy:
 - uncontrolled studies
 - feasibility studies
 - correlation studies of glucose metabolic rate changes with treatment

Systematic review protocol, continued

Studies whose stated purpose is to define diagnostic accuracy but which report results in a way that measures of diagnostic accuracy cannot be duplicated or interpreted, or in which some patients entered are not accounted for, will also be assigned to the technical efficacy level.

5) Assign to **diagnostic accuracy efficacy** level:

- stated purpose is to define diagnostic accuracy, and clinically useful measures (Se/Sp) provided or can be calculated
- meets full or modified (case series with internal controls; blinding if image analysis qualitative) evidence-based medicine criteria
- determines optimal cutpoint from ROC analysis or applies previously determined optimal cutpoint

Caveats will be attached to reports of sensitivity and specificity reported for case series with internal controls if prevalence of severe disease is high.

6) Assign to **diagnostic thinking efficacy** level if meets evidence-based medicine criteria (in box below) for evaluations of diagnostic tests and:

- numbers of subjects without target disorder \geq numbers of cases with disorder (i.e., pretest probability of disease \approx 50%)
- information useful in interpreting test results (i.e. converting pre- test probability of disease to post-test probability using predictive values or likelihood ratios) is provided or can be calculated from information in article.

Evidence-based medicine criteria for studies of diagnostic tests*

- Clearly identified comparison groups, \geq 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.
- Interpretation of the diagnostic standard without knowledge of the test result.

* Purpose and Procedure, *Evidence-Based Medicine*, November/December 1995

7) To further refine judgments of methodologic quality, grade **diagnostic accuracy or thinking efficacy** studies according to criteria in the box on the next page.

Systematic review protocol, continued

Methodologic quality of diagnostic accuracy and diagnostic thinking efficacy studies*

Grade	Criteria
A	<p>Studies with broad generalizability to a variety of patients and no significant flaws in research methods</p> <ul style="list-style-type: none"> • 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 completely described • diagnoses defined by an appropriate reference standard • PET studies technically of high quality and evaluated independently of the reference diagnosis
B	<p>Studies with a narrower spectrum of generalizability, and with only a few flaws that are well described (and impact on conclusions can be assessed)</p> <ul style="list-style-type: none"> • 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	<p>Studies with several methods flaws</p> <ul style="list-style-type: none"> • small sample sizes • incomplete reporting • retrospective studies of diagnostic accuracy
D	<p>Studies with multiple flaws in methods</p> <ul style="list-style-type: none"> • no credible reference standard for diagnosis • test result and determination of final diagnosis not independent • source of patient cohort could not be determined or was obviously influenced by the test result (work up bias) • opinions without substantiating data

* Adapted from: Kent DL, Larson EB. Disease, level of impact, and quality of research methods: three dimensions of clinical efficacy assessment applied to magnetic resonance imaging. *Investigative Radiology* 1992; 27:245-54.

Kent DL, Haynor DR, Longstreth WT, Larson EB. The clinical efficacy of magnetic resonance imaging in neuroimaging. *Annals of Internal Medicine* 1994; 120:856-71.

- 8) Assign to **therapeutic efficacy level** if meets evidence-based criteria for evaluations of diagnostic tests and/or:
- authors discuss how test results did change, or could have changed, treatment for the patients enrolled in the study
 - % of times subsequent procedure avoided due to test results, % of times prospectively stated therapeutic plans changed post-test documented.
- 9) Assign to **patient outcome efficacy** level if patient outcomes with PET are compared to those without PET in a case-control study, cohort study, or randomized controlled trial and/or:
- change in quality adjusted survival or cost/quality adjusted life year gained documented.
- 10) Assign to **societal efficacy** level if both costs (from a societal perspective) and consequences (efficacy, effectiveness, or utility) determined for both PET and an alternative.
- 11) Evaluate quality of studies at each efficacy level; conduct meta analyses if appropriate.
- 12) Articles are excluded from the review if they:
- are duplicated or superseded by subsequent study (at the same level of the hierarchy and with the same purpose) from the same institution
 - contain insufficient 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.

E. Meta analysis was considered for studies of PET diagnostic accuracy

Quantitative (statistical) pooling of the results of diagnostic accuracy studies to arrive at summary measures of sensitivity and specificity (defined in *Appendix 4*) or summary effect measures (Hasselblad and Hedges, 1995) for each disease was considered. The disease-specific systematic reviews described in *Appendices 3 through 8* resulted in decisions that meta analysis would not contribute to the assessment results; reasons for that decision are detailed in each appendix. In general, significant methodologic limitations that would tend to overestimate accuracy were present in all of the literature reviewed. These limitations argued against the validity and usefulness of pooling study results (Eysenck, 1994).

F. Selected alternatives to PET were addressed in the review

Brief discussions of alternate diagnostic technologies are included in the reviews. Alternate technologies were identified according to the following criteria:

- technologies that have been directly compared to PET;
- technologies that have been more rigorously assessed than PET (i.e. using stronger study designs and/or at a higher level in the diagnostic efficacy hierarchy) for a particular application, resulting in documentation that the alternate technologies are equally or more accurate and/or have a better defined role in patient management.

G. Review of the assessment report

The final draft of the assessment report was reviewed and approved by all members of the Advisory Committee, by one of the co-directors of the San Antonio Cochrane Center (Gilbert Ramirez, Ph.D.), and by the Under Secretary for Health, Kenneth W. Kizer, M.D., M.P.H. One of the committee members, an oncologist (Dr. Holohan), also reviewed all of the original literature cited in the oncology sections of the report to confirm the MDRC Technology Assessment Program's evaluation of that literature. Other reviewers included: Alan Garber, M.D., Ph D. (methods), L. Jack Faling, M.D. and Charles Powell, M.D. (lung cancer, solitary pulmonary nodules), John Booss, M.D., Thomas Bird, M.D., Jeffrey Cummings, M.D., and Judith Salerno, M.D. (Alzheimer's disease).

Summary of the Full Assessment Methodology

I. Systematic review of the literature

- Systematic review protocol applied.
- External reviewer independently judged studies for quality and position in hierarchy.
- Meta analysis was contemplated for diagnostic accuracy studies, but was not performed due to methodologic limitation in available studies.

II. Survey of VA PET facilities/site visits

- Data collected by written surveys covering fiscal years 1993, 1994, and 1995.
- Site visits conducted in August and September 1994
- Descriptive analysis and tabulation.

III. RESULTS: Site visits and surveys

PET is a relatively new addition to the repertoire of clinical diagnostic tests available both within and outside VA. Many of VA's PET facilities have become operational since 1990, and the information collected through the site visits and surveys represents preliminary data on VA experience with the technology.

The MDRC Technology Assessment Program obtained information on experience at eleven VA PET centers. A written survey and subsequent site visits were carried out from August through October, 1994. Another brief follow-up survey was distributed in December, 1995. Of the twelve initially approved PET sites, eleven were fully operational at the time of the assessment; support for the twelfth had been withdrawn. After completion of the site visits, support for another PET center was discontinued by the local VA medical center administration. At the time of release of this report, ten VA PET centers were in operation.

Interview subjects were selected by the VA PET centers, and included nuclear medicine physicians, PET center staff, referring physician specialists (in cardiology, neurology, oncology, and psychiatry, representing the clinical and research areas where PET is most commonly used), and hospital administrators. Most of the interview subjects had multiple job roles (administrative, clinical, and/or research), reflecting the academic environment for VA PET activity.

Full details of the site visit and survey findings are presented in *Appendix 9*.

A. Characteristics of sites

The pre-site visit surveys indicated that an equivalent array of ancillary services was offered at each location; substantial differences in PET utilization would not be likely to be attributable to differences in the types of patients (as represented by the range of services in place) treated at the VA medical centers (VAMCs). Most of the PET sites became operational in 1992 and 1993. The location of the PET camera was equally distributed between VAMCs and university affiliates, although the sharing partner/university affiliate tended to be the main source for radiopharmaceuticals. At most sites responsibility for PET center personnel was distributed equally between VA and university sharing partner.

Fluorodeoxyglucose (FDG) was the only radiopharmaceutical common to all sites. An important distinction among sites was the main mission or focus, which ranged from primarily research to primarily clinical work. The centers also differed in the models of PET scanners used.

B. Activity at each PET site

Activity at PET sites was compared using total number of patients studied rather than total number of scans. Expressing utilization according to the number of patients studied was felt to reflect most accurately the existing referral base for each site. While the total number of patients scanned is a relatively crude measure, adjusted measures reflecting variations among scanning protocols with respect to scan time and resources used, and PET technology across sites, would require a standardized workload unit and prospective data collection, neither of which has been systematically implemented across VA centers.

A wide range of types and volumes of PET studies was performed across VA in 1993 and 1994. During that time, more subjects appeared to have been scanned for clinical purposes than for research purposes. These trends may be attributed, in part, to differences in centers' definitions of studies classified as "clinical" versus those classified as "research."

- In 1993, most research activity system-wide was in neurology and psychiatry, followed by oncology; clinical activity consisted mainly of neurology and cardiology studies, followed by oncology.
- In 1994, most research activity was in neurology and psychiatry, followed by a growing interest in oncology; research activity in cardiology appeared to decrease. Neurology applications comprised the majority of clinical studies, followed by a growing interest in oncology applications. Some clinical cardiology studies were performed, but cardiology did not contribute substantially to overall clinical activity.
- In the brief follow-up survey conducted in 1995, six sites reported an increasing interest in clinical PET studies, which was attributed largely to increasing demand for clinical oncology studies. Interest in clinical cardiology applications continued to decrease. The increased interest in oncology studies may be attributed, in part, to the results of educational and marketing efforts made by PET center staff in recent years, and to the growing body of PET literature reporting clinical oncology applications. Two sites reported an increased use of PET in psychiatric and neurologic research.

C. Barriers and incentives to the use of PET

VA made a significant contribution to overall PET activity by committing substantial resources to the start-up of twelve PET centers. In return, PET has contributed significantly to overall research activity within VA. PET is regarded by many researchers in neurology and psychiatry as an essential tool for research into mental disorders, an area which is important to the veteran population. Additionally, many investigators view PET as a critical tool for basic physiologic research.

Foci of strong academic and clinical interests in functional imaging were important to obtaining initial support for PET at individual VA medical centers. Variations in current research activity across sites reflect the degree to which the initial interests extended into other research areas. The depth and breadth of the clinical and research bases at each site influenced the types of applications studied, the kinds of patients included in these studies, and the relative proportions of clinical and research studies conducted. At all sites, the reputation and expertise of the PET director and core PET center staff contributed to the willingness of medical staff and researchers to use PET as a clinical and research tool. The site visit interviews indicated that there are important organizational, professional, scientific, and reimbursement factors contributing to the relatively slow diffusion of PET into clinical practice. Interview subjects felt that limited FDA approved clinical PET applications and lack of demonstrated clinical utility perpetuated the perception of the general medical community and regulators that PET is primarily a research tool. Subjects felt that these factors also contributed to third party payers' inconsistent reimbursement policies.

PET is a very costly technology that requires a significant investment to cover start up costs and annual operating expenses. The major costs at each PET site were: equipment amortization; maintenance contracts for the scanner and cyclotron; scanner-related supplies; cyclotron supplies including target materials; and personnel, particularly highly skilled radiochemists, clinical and research specialists, analysts and programmers. Other significant costs included installation and maintenance of pneumatic tube systems used to transport radioactive isotopes between facilities.

PET directors and medical center directors have attempted to recover and reduce some of these costs. Those sites able to obtain reimbursement for clinical studies generally

developed *a priori* consensus-building efforts among payers and providers within their communities in exchange for data collection. Multiple studies were often coordinated with production and use of radiotracers in an effort to minimize waste. Some sites generated revenue by selling cyclotron products, while others extended their catchment areas to include a broader patient base. One site made a decision to maintain low operating costs by purchasing cyclotron products from a private source, rather than producing its own. Two recommendations to offset the high and often unexpected maintenance costs of the scanner and cyclotron were made during the site visits: 1) establish an escrow account from equal contributions made by the sharing partners, and 2) support a “roving” maintenance team within VA to service all VA PET centers.

Inadequate staffing (particularly radiochemists) was cited as impeding the conduct of certain studies. Four PET centers cited the need for a qualified radiochemist as a major influence on the volume and variety of studies; competition for these specialists is intense. In VA hospitals, PET centers’ hours of operation were frequently curtailed by inflexible tours of duty, restrictions in overtime salary, and restrictions and/or cutbacks in the number of Full Time Equivalent Employees. Reimbursement of patient transport costs for non-VA patients and the inability to transport less medically stable patients were barriers to access for some patients.

Competition among clinical specialties for access to PET, between PET and other technologies, and among PET centers in the same city may also affect access to PET for some patients. One center that developed a process to facilitate research protocol approval based on a NIH model; this assured equal representation of the sharing partners and the medical specialists interested in PET. Competition with other technologies and other local PET centers may dilute support for VA’s PET facilities.

Several issues were related specifically to VA and to VA patients. Some interview subjects reported poor patient compliance in keeping scheduled appointments was noted. Others noted private sector patients’ concerns about VA quality of care or perceptions that the services provided by PET centers at VA hospitals were restricted to VA patients only. Many VA PET center directors expressed frustration at not having the authority or resources to properly market their services to the private sector. The inability to attract VA patients for PET scans was attributed to either a lower burden of particular diseases among veterans compared to the general population, or to the failure of many veteran patients to meet protocol inclusion criteria.

Interview subjects saw centralized strategic planning around distribution, construction and maintenance as necessary to the overall investment in costly technologies such as PET. Nevertheless, subjects described these processes as frustrating, inefficient, and protracted. Local VA administrators perceived a lack of vision and commitment to PET by Headquarters; many felt that they were expected to support new, costly programs and services within existing funding levels.

Variations in VA’s financial commitment among the centers appeared to be related to the degree to which local medical center directors sustained the support, often through the sharing agreements with academic affiliates. The agreement negotiating team typically included representatives from Fiscal Service and the Director’s Office. The degree to which the Director’s Office participated in these negotiations varied across sites; the most active participation tended to produce some of the most functional arrangements. To comply with VA policy, PET center directors with dual appointments were excluded from negotiations. Consequently, interview subjects felt that the negotiations could not benefit from the insight of the individual who was most familiar with the needs of the center.

In these agreements, PET center cost sharing varied; under some agreements, costs were evenly distributed between partners, while other agreements stipulated alternate means of distributing costs. VA's contribution ranged from covering partial costs of the scanner to covering partial costs of both the scanner and overhead. Unrealistically high volume projections and unreasonably low overhead costs formed the basis on which some of the original sharing agreements were negotiated. Negotiations in recent years have used more realistic volume projections or a patient charge based on the national average. One site developed a workload unit to better reflect true utilization of resources.

Interview subjects felt that the sharing arrangement most favorable for VAMCs with PET centers located at the academic affiliate was one that required full payment up front by the affiliate for its portion of the scanner. If contributing to overhead costs, the VAMC was subsequently billed on a fee-for-service basis at a charge approximately equal to the national average. Another arrangement favorable to the VAMC was one in which a fixed number of "free" scans for VA patients was determined up front, in exchange for partial use of the scanner by other sharing partners. These arrangements insure that each VAMC recovers its portion of the investment up front, without risk of financial loss, should volume projections be unfulfilled or overhead costs be excessive.

IV. RESULTS: Systematic reviews

The full background, results, and discussion texts, data abstraction tables for diagnostic accuracy and therapeutic efficacy studies, and comparisons of PET to alternate technologies are presented in the appendices (*Appendix 3: Head and Neck Cancer; Appendix 4: Colorectal Cancer; Appendix 5: Breast Cancer; Appendix 6: Lung Cancer; Appendix 7: Solitary Pulmonary Nodules; and Appendix 8: Alzheimer's Disease*). The overall results of the systematic reviews are summarized here in Tables 1 through 6 (pages 20 through 28).

The systematic reviews indicate that research into the clinical utility of PET in selected conditions relevant to the veteran population is in its preliminary stages. The available studies have focused on the feasibility of using PET in these conditions, and on defining its accuracy as a diagnostic test. A few studies have addressed changes in treatment decisions based on PET findings. However, the MDRC was unable to locate any studies documenting changes in outcomes of care or costs of care associated with incorporating PET into diagnostic strategies for the conditions addressed in this assessment. Critical research into defining the clinical consequences of using PET for diagnosis has yet to be performed or reported.

Since most of the PET studies analyzed for the systematic reviews address diagnostic accuracy, revisiting criteria for a valid evaluation of diagnostic test accuracy is advisable here. The McMaster University Department of Clinical Epidemiology and Biostatistics provided a seminal (1981) and concise list of the questions to ask regarding published clinical evaluations of diagnostic tests. These are:

- Was there an independent, "blind" comparison with a "gold standard" of diagnosis?
- Did the patient sample include an appropriate spectrum of mild and severe, treated and untreated disease, plus individuals with different but commonly confused disorders?
- Was the setting for the study, as well as the filter through which study patients passed, adequately described?

- Was the reproducibility of the test result (precision) and its interpretation (observer variation) determined?
- Was the term “normal” defined sensibly?
- If the test is advocated as part of a cluster or sequence of tests, was its contribution to the overall validity of the cluster or sequence determined?
- Were the tactics for carrying out the test described in sufficient detail to permit their exact replication?
- Was the “utility “ of the test determined?

According to these and analogous criteria incorporated into this assessment’s systematic review protocol, the published studies using PET to diagnose Alzheimer’s disease have been relatively well constructed and present a coherent set of observations on PET’s good level of agreement with widely used clinical criteria for dementia of the Alzheimer’s type. As most results fell within a relatively narrow range of estimates of accuracy, meta analyses of the diagnostic accuracy results were not conducted.

There are, however, barriers to moving PET into routine use in diagnosing Alzheimer’s disease and to affecting outcomes of care by means of PET diagnosis:

- relatively few of the published studies prospectively evaluated large numbers of patients with causes of dementia that can be confused, or present concurrently, with Alzheimer’s disease, making a valid estimate of the positive predictive value of PET difficult to determine;
- histologically verified Alzheimer’s disease represents a subset of patients with clinically diagnosed dementia of the Alzheimer’s type, and the results of ongoing studies defining the agreement of PET with the gold standard of autopsy diagnosis in Alzheimer’s disease are not yet available;
- effective treatments for Alzheimer’s disease are not available. An accurate diagnostic test (relative to the gold standard of autopsy result) is needed for research into treatments for Alzheimer’s disease; both PET and other tests that have been shown to have equivalent accuracy may be useful in this role.

The published evidence for the accuracy of PET in diagnosing cancer is less convincing than that for its accuracy in diagnosing dementia of the Alzheimer’s type. While the available studies report good face accuracy for PET (particularly in clinical settings where PET was used to differentiate recurrent cancer from treatment artifacts such as scars), many of the studies did not adhere to the principles of study design outlined above.

Almost all PET cancer studies are retrospectively analyzed case series. They enrolled relatively few patients (too few to allow one to comfortably draw conclusions from the data), did not include control groups (i.e., did not adequately account for biologic variation in test results or differential diagnosis with other conditions), and, when PET images were visually interpreted, often did not blind image interpreters or address issues of interobserver variation. Many published studies can be assumed to be subject to context bias (Eggin and Feinstein, 1996). The studies that compared PET to other diagnostic technologies did not randomize the order of test administration, and in some cases were subject to work up bias (where the results of one test led to the decision to perform another, or to confirm diagnosis by biopsy); these biases will have affected accuracy estimates of both PET and the alternative test or tests.

A critical shortcoming in the diagnostic accuracy PET oncology literature for organizations, like VA, that are seeking to rationalize the provision of services on a regional or system-wide basis, is the lack of epidemiologic information in the published studies. The filters through which patients passed to be included in the published case series are often inadequately described, making extrapolation of the results to defined populations, and subsequent planning for these populations, difficult. To summarize, like the early studies into other diagnostic technologies, the methodologic weaknesses of the available PET oncology studies will have tended to overestimate accuracy and clinical value. Accordingly, meta analyses of these studies were not performed.

A widely credited role for PET in the nuclear medicine and surgery literature is that of increasing diagnostic certainty regarding the need for invasive procedures (e.g., neck dissection in patients with head and neck cancer, resection of metastases from colorectal cancer that are potentially curable if isolated, axillary dissection in breast cancer, thoracotomy for solitary pulmonary nodules). The authors of a few oncology studies discussed the potential or actual changes in treatment that resulted from incorporating PET into diagnostic strategies at critical decision points in cancer treatment processes. These studies were retrospective case series that had not been specifically designed to document changes in treatment; methods for recording changes in treatment plans were not specified, and results data tended not to be systematically analyzed or presented. The studies generally enrolled highly selected patients whose previous work-up was not clearly specified, nor was the size or composition of the referral base from which the patient sample was drawn. Information from PET studies resulted in more appropriate treatment for some patients. However, the published studies tended to give inadequate details about what happened to patients whose PET studies did not accurately reflect their disease status.

PET is generally presented as complementary to anatomic imaging studies such as CT or MRI. Accordingly, further work on PET's treatment impact and role in a multi-test diagnostic strategy is needed before the population impact of PET on a health care system such as VHA can be estimated. It should be noted that the management of patients with cancer is a highly complex area, with many uncertainties beyond those related to the impact of PET; specifically, treatment for many cancers (particularly the solid tumors addressed in this assessment) is less than optimally effective. Before population outcomes (e.g., mortality rates from specific cancers) can be improved, a wide range of interventions for prevention, screening, diagnosis, treatment, and follow-up also need to be improved.

The disease-specific systematic reviews in this assessment included, for comparative purposes, information on some of the diagnostic technologies that may be alternatives to PET. While information on these alternatives was not identified and retrieved with the same thoroughness as the information on PET, a sample of articles from the recent peer reviewed literature indicates that research into alternate tests has resulted in substantial improvements in accuracy for many of the conditions discussed in this assessment. Some of the research that has been conducted for alternative diagnostic tests surpasses the PET literature in its methodologic rigor.

Systematic review summary Tables 1 through 6 (pages 20 through 28) present the results of diagnostic accuracy efficacy studies in each of the diseases considered for this assessment. Studies were included if they met all or some of the evidence-based medicine criteria for diagnostic test evaluations. Methodologic quality grades, which further refine the quality judgments implicit in the evidence-based medicine criteria, are also noted. Many of the included studies do not meet high methodologic standards; in the absence of more rigorous studies, they are presented here in an effort to make the review methods and conclusions as transparent as possible.

V. RISKS ASSOCIATED WITH PET RADIOPHARMACEUTICALS

Research articles reviewed for this assessment either:

- provided no comments on any risks associated with PET radiopharmaceuticals, or
- included a statement indicating that no patients in the study experienced adverse events after radiopharmaceutical administration.

VI. FDA STATUS OF PET RADIOPHARMACEUTICALS

The Food and Drug Administration has approved two PET radiopharmaceuticals. The quotations below are from the package inserts.

- CardioGen (Rubidium Rb 82 Generator) is indicated for use as “a myocardial imaging agent that is useful in distinguishing normal from abnormal myocardium in patients with suspected myocardial infarction.”
- Fludeoxyglucose F 18 Injection [(¹⁸F-FDG) The Methodist Medical Center of Illinois] is indicated for “the identification of regions of abnormal glucose metabolism associated with foci of epileptic seizure. FDG is not indicated for distinguishing epileptogenic foci from brain tumors or other brain lesions which may cause seizures.”

The many PET imaging facilities that “compound” radiopharmaceuticals on-site do so under the aegis of state practice of medicine and pharmacy laws. The FDA has determined that PET centers are manufacturing a new drug and that they are subject to existing new drug regulations. Facilities that manufacture radiopharmaceuticals for clinical use must file a New Drug Application (NDA) or an Abbreviated New Drug Applications (ANDA) and conform with Current Good Manufacturing Practice (CGMP) standards. Clinical investigators who wish to conduct clinical trials with an unapproved PET radiopharmaceutical must file an Investigational New Drug (IND) application for each drug.

The FDA has offered to work with the PET community to help sponsors and investigators interpret and utilize the appropriate regulations in offer to comply with the existing new drug regulations.

VII. CONCLUSIONS

The site visits and surveys confirm that VA has made a substantial resource commitment to its PET imaging facilities. This commitment has the potential to contribute to fulfilling two parts of VA's mission: research and clinical care. PET is widely credited as an important basic research tool in the literature; VA PET researchers who were interviewed for this assessment share this belief. The efficiency of basic research activities would be enhanced by implementation of suggestions for improving operations that were made by VA PET center personnel during the site visits.

The site visits and surveys outlined a wide range of research and clinical activities in VA PET centers. There are many site-specific protocols and areas of research interest. Coordination of these activities has not been systematically addressed.

The presence of PET on the lists of many health care assessment agencies (*Appendix 10*), nationally and internationally, and many discussions in the medical literature attest to concerns that PET will follow a familiar diffusion trajectory into clinical care before its usefulness and contribution to improved outcomes have been adequately evaluated. The trends seen in PET utilization during the assessment period indicate that oncology is an increasing focus for clinical activity in VA. This trend should be of interest to policy makers in the context of the findings of the systematic reviews reported here.

This assessment's systematic reviews of the literature indicate, to the extent that the published literature represents the existing data, that the knowledge base supporting clinical diagnostic applications of PET has significant deficiencies. Methodologic weaknesses in published studies seriously limit the validity and generalizability of the available evidence on the accuracy of PET as a diagnostic test, and PET's contribution to improving outcomes has not been systematically addressed. Accordingly, the assessment team believes that the literature as of September, 1996 does not support widespread incorporation of PET studies into routine diagnostic strategies for the applications addressed in this assessment.

The Advisory Committee to the PET assessment believed that the assessment results supported a conclusion that VA should maximize the value derived from its existing resource commitment, rather than invest in additional PET centers at the present time. Maximizing the value of the existing commitment could include:

- Building organizational structures to coordinate its PET activities across the VA system.
- Implementing a VA PET registry. Systematic, standardized data (including those related to work load, resources used, and operations) specific to PET would facilitate future assessment efforts. A registry for tracking diagnosis-specific utilization, the marginal contribution of PET to a diagnostic strategy involving other tests, impact of PET on treatment decision making, and treatment outcomes would also facilitate future assessments.
- Organizing a cooperative group consisting of VA PET centers and their academic affiliates. Such a group could facilitate efforts to comply with FDA regulations. Efficacy research in oncology is frequently conducted by cooperative groups, and could supply a model for PET oncology research. A VA cooperative PET group could also attempt to define clinical research areas of interest to the entire VA system, and to design multi-center studies of high methodologic quality.
- Supporting rigorous, prospectively designed clinical research that corrects the methodologic limitations outlined in the diagnosis-specific systematic reviews. Once the diagnostic accuracy of PET has been adequately defined, attention should be directed to defining the changes in patient management decisions, outcomes of care, patient outcomes, cost-

effectiveness of care, and cost-utility of care that are associated with incorporating PET into diagnostic strategies.

- Submitting currently unpublished data from studies of high methodologic quality for peer review. Advocates of PET both within and outside VA feel strongly that the clinical utility of PET is increasingly evident; these opinions may be supported by currently unpublished data.

Table 1 Summary of the literature: Diagnostic accuracy efficacy of PET and alternatives in head and neck cancer
(from studies comparing PET directly to other diagnostic tests)

Role	Study	N	Operating characteristics*				Evidence-based medicine criteria**			Methodologic quality grade***
			PET	CT	MRI	Other	controls****	standard	blinding	
Unknown primary	Rege, et al., 1994	4 cases 0 controls	Se = 50%		Se = 0%		-	+	-	D
Known primary site	Rege, et al., 1994	30 cases 0 controls	Se = 97%		Se = 77%		-	+	-	D
	Laubenbacher, et al., 1995	17 cases 0 controls	Se = 100%		Se = 100%	endoscopy, Se = 100%	-	+	-	D
Primary tumor staging (size, extent)	Laubenbacher, et al., 1995	17 cases 0 controls	Se = 41%		Se = 41%	endoscopy, Se = 59%	-	+	-	D
Cervical node involvement	Rege, et al., 1994	16 pos 18 neg	Se = 88% Sp = 89%		Se = 81% Sp = 89%		+	+	-	D
	McGurt, et al., 1995	14 pos 31 neg	accuracy = 82%	accuracy = 82%		clinical exam accuracy = 71%	+	+	-	D
	Laubenbacher, et al., 1995	83 pos nodes 438 neg nodes	Se = 90% Sp = 96%		Se = 78% Sp = 71%		+	+	-	D
		18 pos neck sides 16 neg neck sides	Se = 89% Sp = 100%		Se = 72% Sp = 56%		+	+	-	D
	Braams, et al., 1995	22 pos nodes 177 neg nodes	Se = 91% Sp = 88%		Se = 36% Sp = 94%		+	+	-	D
	Benchaou, et al., 1996	54 pos node groups 414 neg node groups	Se = 72% Sp = 99% PPV = 89% NPV = 99%	Se = 67% Sp = 97% PPV = 74% NPV = 95%		clinical exam Se = 61% Sp = 97% PPV = 72% NPV = 95%	+	+	+	B
Suspected recurrent disease	Rege, et al., 1994	10 pos 7 neg	Se = 90% Sp = 100%		Se = 67% Sp = 57%		+	+	-	D
	Lapela, et al., 1995	16 pos 17 neg	Se = 88 -94% Sp = 43 -86% depending on criteria for pos	Se = 92% Sp = 50%			+	+	+	C

Abbreviations: Ct, computed tomography
MRI, magnetic resonance imaging
neg, negative for disease
pos, positive for disease
Se, sensitivity
Sp, specificity

PPV, positive predictive value
NPV, negative predictive value
US/FNA, ultrasound/fine needle aspiration

* operating characteristics defined in Appendix 2: Assessing Diagnostic Technologies, pages 5-7
** Appendix 2, page 8
*** Appendix 2, page 9
****"controls" were case series patients with benign conditions

Table 2 Summary of the literature: Diagnostic accuracy efficacy of PET and alternatives in colorectal cancer

Notes The PET studies in this table were retrospectively analyzed case series; internal controls (cases with benign, rather than malignant, conditions) allowed the calculation of specificity as well as sensitivity. Some of the alternatives to PET have been evaluated using more rigorous study designs.

Some studies analyzed results separately according to the clinical role of PET for subsets of patients; these studies appear in the table more than once, and may have received different methodologic quality grades for each subset analysis.

Role	Study	N	Operating characteristics*				Evidence-based medicine criteria**			Methodologic quality grade***
			PET	CT	MRI	Other	controls	standard	blinding	
Detecting or staging primary or recurrent disease	Falk, et al., 1994	16 patients: 15 malignant lesions; 3 benign lesions	Se = 87% Sp = 67%	Se = 47% Sp = 100%			+ (internal)	+	partial	D
	Nattinger, et al., 1991 (ACP review)				colonoscopy Se = 94% Sp = 100%	(review)	(review)	(review)	(review)	
	Hernandez-Socorro, et al., 1995	40 cases 64 controls				colonoscopy Se = 94% Sp = 100% hydrocolonic ultrasound Se = 97% Sp = 97%	+	+	+	B
Diagnosing recurrent tumor vs scar	Strauss, et al., 1989	29 patients: 21 malignant lesions; 8 scar	Se = 95% Sp = 100%				+ (internal)	+	+ (quantitative analysis)	C
	Schlag, et al., 1989	18 patients: 11 malignant lesions; 6 scar	Se = 92% Sp = 100%			immunoscintigraphy Se = 40% Sp = 50%	+ (internal)	+	+ (quantitative analysis)	C
	Ito, et al., 1992	15 patients: 11 malignant lesions; 4 scar	Se = 100% Sp = 100%		Se = 91% Sp = 100%		+ (internal)	+	+ (quantitative analysis)	C
	Schiepers, et al., 1994	6 patients: 5 malignant lesions; 1 scar	Se = 100% Sp = 100%				+ (internal)	+	-	D
Diagnosing recurrent tumor vs scar	Hawes, et al., 1993	85 with disease 408 without disease (review with weighted average of results from 7 studies)				endoscopic ultrasound Se = 99% Sp = 88%	(review)	(review)	(review)	(review)

Role	Study	N	Operating characteristics*				Evidence-based medicine criteria**			Methodologic quality grade***
			PET	CT	MRI	Other	controls	standard	blinding	
Diagnosing liver metastases	Schiepers, et al., 1994	80 studies: 34 malignant lesions; 46 benign lesions	Se = 94% Sp = 100%			CT and/or ultrasound Se = 85% Sp = 98%	+	+	-	C
	Vitola, et al., 1996	55 sites: 39 malignant; 16 benign 24 patients: 19 malignant disease; 5 benign	Se = 90% Sp = 100% Se = 95% Sp = 100%	Se = 86% Sp = 58%		CT portography Se = 97% Sp = 9% Se = 100% Sp = 33%	+(internal)	+	+(semiquantitative analysis)	C
	Lai, et al., 1996	34 patients: 27 with malignant disease; 7 benign or no disease	Se = 93% Sp = 57%	Se = 100% Sp = 14%	Se = 100% Sp = 80%		+(internal)	+	+	C
	Stark, et al., 1987	57 cases; 72 controls: 21 benign liver disease; 51 with normal livers		Se = 80% Sp = 94%	Se = 82% Sp = 99%		+	+	+	B
	Panzer, et al., 1991 (ACP review)	review		Se = 90% Sp = 90% LR + = 8 LR - = 0.11		ultrasound, adequate studies Se = 80% Sp = 90% LR + = 9 LR - = 0.22	(review)	(review)	(review)	(review)
Diagnosing liver metastases	Rafaelsen, et al., 1995	295 patients: 64 with liver metastases 231 without liver metastases				liver enzymes Se = 9-47% Sp = 92-98% preop US Se = 70% Sp = 94% surgical exploration Se = 84% Sp = 97% intraop US Se = 97% Sp = 98%	+(internal)	+	+	B

Abbreviations CT, computed tomography
MRI, magnetic resonance imaging
neg, negative for disease
pos, positive for disease
LR, likelihood ratio

PPV, positive predictive value
NPV, negative predictive value
US/FNA, ultrasound/fine needle aspiration
ACP, American College of Physicians

*operating characteristics defined in Appendix 2: Assessing Diagnostic Technologies, pages 5-7
**Appendix 2, page 8
*** Appendix 2, page 9

Table 3 Summary of the Literature: Diagnostic accuracy efficacy studies of PET and alternatives in breast cancer

Notes: All studies except Nieweg, et al., 1993b, which was a case-control study, were series of patients presenting for surgical evaluation of breast masses (a high index of suspicion of malignant disease) and included internal controls as the comparison group. Predictive values should be viewed accordingly. Studies assessing axillary node involvement included patients with malignant primary breast disease. Results from Avril, et al., 1996b were reported as ranges of data from all subgroup analyses. Results from Avril, et al., 1996a included all patients with benign and malignant primary disease and represent 95% confidence intervals; subgroup analyses were not reported because of their small study size. None of these studies met strict evidence-based medicine criteria for blinding, but all studies provided data on the comprehensiveness of blinding of test interpreters to the gold standard.

Where substantial duplication in purpose of study, patients studied, and results in multiple studies from the same institution could be inferred, only the most recent, largest, most rigorously designed, or most comprehensive was included in the table. Although data from both studies by Avril and associates (1996a and 1996b) represent the same patient population, these studies addressed different purposes; inclusion of both publications were felt to be warranted.

Abbreviations are listed at the end of the table.

Role <i>(Note: some studies assessed multiple roled)</i>	Study	N	Operating Characteristics*			Evidence-Based Medicine Criteria**			Methodologic Quality Grade***
			PET	Clinical Exam	Mammography	comparison group	histologic gold standard	blinding	
Defining primary disease	Adler, et al., 1993	27 positive lesions 8 negative lesions	Se=96% Sp=100%			+ internal	+	+	C
	Nieweg, et al., 1993b	11 cases 8 controls	Se=91% Sp=100%			+	+	+	C
	Avril, et al., 1996b	41 positive lesions 31 negative lesions	Se=68%-94% Sp=84%-100% PPV=87%-97% NPV=70%-93%			+ internal	+	partial	D
	Scheidhauer, et al., 1996	23 malignant cases 7 benign cases	Se=91% Sp=86%	Se=74% Sp=71%	Se=86%	+ internal	+	partial	D
Defining axillary node involvement	Adler, et al., 1993	9 positive axillae 10 negative axillae	Se=90% Sp=100%			+ internal	+	+	C
	Avril, et al., 1996a	24 positive axillae 27 negative axillae	Se=57%-93% Sp=81%-100% PPV=75%-100% NPV=66%-100%	Se=36%-78% Sp=66%-96% PPV=30%-70% NPV=51%-85%		+ internal	+	+	C
	Scheidhauer, et al., 1996	9 malignant cases 9 benign cases	Se=100% Sp=89%			+ internal	+	partial	D
Detecting distant metastases	Scheidhauer, et al., 1996	8 positive lesions 15 negative lesions	Se=100% Sp=100%			+ internal	+	partial	D

N, number of study subjects included in analysis; unless otherwise noted, data are analyzed by subject
 Se, sensitivity
 Sp, specificity
 PPV, positive predictive value
 NPV, negative predictive value

* operating characteristics defined in *Appendix 2: Assessing Diagnostic Technologies, pages 5-7*
 ** *Appendix 2, page 8*
 *** *Appendix 2, page 9*

Table 4 Summary of the Literature: Diagnostic accuracy efficacy studies of PET and alternatives in lung cancer

Notes: All of the studies in the table are case series (Level V evidence) with internal controls (i.e. those with benign masses) used as a comparison group. All patients in these studies had suspected or biopsy-proven lung cancer (i.e. the pre-test probability of disease in the study populations was very high). Results from Knight, et al., 1996 and Inoue, et al., 1995 were reported as ranges to include data from all subgroup analyses.

None of these studies met strict evidence-based medicine criteria for blinding, but all studies presented information on blinding of the test interpreters to the biopsy gold standard. Blinding of the PET interpreters to other clinical and radiologic data varied across studies and is reflected in the columns designated "Operating Characteristics"; "PET + CT" indicates a complementary role of PET with CT, and PET alone indicates a substitutive role of PET for CT.

Where substantial duplication in purpose of study, patients studied, and results in multiple studies from the same institution could be inferred, only the most recent, largest, most rigorously designed, or most comprehensive was included in the table. Studies reviewed but not included are listed under "References".

Abbreviations are listed at the end of the table.

Role <i>(Note: Some studies assessed multiple roles)</i>	Study	N	Operating Characteristics*			Evidence-Based Medicine Criteria**			Methodologic Quality Grade***
			PET	PET + CT	CT	comparison group	histologic gold standard	blinding	
Defining unknown primary disease	Kubota, et al., 1990	12 malignant cases 10 benign cases		Se=83% Sp=90% accuracy=86%	no data reported	+ internal	+	+	C
	Scott, et al., 1994	47 malignant cases 15 benign cases		Se=94% Sp=80%	no data reported	+ internal	+	+	C
	Slosman, et al., 1994	31 malignant cases 5 benign cases		Se=93.5%	no data reported	+ internal	+ & follow-up	+	C
	Wahl, et al., 1994	19 malignant cases 4 benign cases	Se=100%		Se=100%	+ internal	+	+	C
	Sazon, et al., 1996	82 malignant cases 25 benign cases	Se=100% Sp=52%		no data reported	+ internal	+	+	C
	Knight, et al., 1996	32 malignant cases 16 benign cases		Se=100% Sp=58%-63% PPV=75% NPV=100%	Se=33%-41% Sp=52% PPV=83% NPV=52%	+ internal	+	+	D

Role <i>(Note: Some studies assessed multiple roles)</i>	Study	N	Operating Characteristics*			Evidence-Based Medicine Criteria**			Methodologic Quality Grade***
			PET	PET + CT	CT	comparison group	histologic gold standard	blinding	
Detecting overall lymph adenopathy	Patz, et al., 1995	42 patients with: 23 malignant nodes 39 benign nodes	Se=83% Sp=82%		Se=43% Sp=85%	+ internal	+	+	D
Detecting hilar/lobar lymph adenopathy	Patz, et al., 1995	42 patients with : 11 malignant nodes 29 benign nodes	Se=73% Sp=76%		Se=27% Sp=86%	+ internal	+	+	D
Detecting mediastinal lymph adenopathy	Patz, et al., 1995	42 patients with: 12 malignant nodes 10 benign nodes	Se=92% Sp=100%		Se=58% Sp=80%	+ internal	+	+	D
	Wahl, et al., 1994	23 patients with: 11 malignant sides 16 benign sides	Se=82% Sp=81% accuracy=81%		Se=64% Sp=44% accuracy=52%	+ internal	+	+	C
	Chin, et al., 1995	9 malignant cases 21 benign cases		Se=70% Sp=81% accuracy=80%	Se=56% Sp=86% accuracy=77%	+ internal	+	+	D
	Valk, et al., 1995	24 malignant sides 52 benign sides		Se=83% Sp=94% accuracy=91%	Se=63% Sp=73% accuracy=70%	+ internal	+ & follow-up	+	D
	Sazon, et al., 1996	32 patients with: 16 malignant sides 16 benign sides	Se=100% Sp=100%		Se=81% Sp=56%	+ internal	+	+	C
	Scott, et al., 1996	10 malignant nodes 65 negative nodes within: 9 malignant cases 18 benign cases		Se=100% Sp=98%-100%	Se=60% Sp=83%-94%	+ internal	+	+	D
Distinguishing local cancer recurrence from fibrosis	Patz, et al., 1994	35 recurrence cases 8 fibrosis cases		Se=97.1% Sp=100%	no data reported	+ internal	+ & follow-up	+	D
	Inoue, et al., 1995	23 recurrence cases 13 fibrosis cases		PET + x-ray, CT, MRI Se=100% Sp=56%-78% accuracy=86%	no x-ray, CT, or MRI data reported	+ internal	+ & follow-up	+	D

N, number of total study subjects included in analysis; unless otherwise noted, data are analyzed by subject
 Se, sensitivity
 Sp, specificity
 PPV, positive predictive value
 NPV, negative predictive value
 CT, computed tomography
 MRI, magnetic resonance imaging

* operating characteristics defined in Appendix 2: Assessing Diagnostic Technologies, Pages 5-7
 ** Appendix 2, page 8
 *** Appendix 2, page 9

Table 5 Summary of the Literature: Diagnostic accuracy efficacy studies of PET in solitary pulmonary nodules

Notes: All of the studies in the table are case series (Level V evidence) and met most of the evidence-based medicine criteria for diagnostic test evaluations. None of the studies met strict evidence-based medicine criteria for blinding, but all studies except Gupta, et al., 1996 provided information on the comprehensiveness of blinding of test interpreters to the biopsy gold standard. Blinding of PET interpreters to other clinical and radiologic data varied across studies.

Internal controls (i.e. those with benign masses) were used in each study, and it was possible to calculate sensitivity and specificity for PET in those studies. The pre-test probability of disease in these study populations was very high, and predictive values were not reported. Each study varied in inclusion criteria with respect to maximal lesion size and image characteristics (pulmonary masses, ill-defined infiltrates, focal lesions). Operating characteristics from these studies should be interpreted with caution.

Where substantial duplication in purpose of study, patients studied, and results in multiple studies from the same institution could be inferred, only the most recent, largest, most rigorously designed, or most comprehensive was included in the table. While data from Dewan, et al., 1995 and Gupta, et al., 1996 are likely derived from the same patient population, these studies addressed different purposes, and inclusion of both was felt to be warranted. Studies reviewed but not included are listed under "References".

Abbreviations are listed at the end of the table.

Role	Study	N	Operating Characteristics*			Evidence-Based Medicine Criteria**			Methodologic Quality Grade***
			PET	TTNA	other	comparison group	histologic gold standard	blinding	
Defining unknown SPN	Dewan, et al., 1995	26 malignant lesions 9 benign lesions	Se=100% Sp=78% accuracy=94%	Se=81% Sp=100% accuracy=86%		internal	+	partial	D
	Bury, et al., 1996	33 malignant cases 17 benign cases	Se=100% Sp=88%			internal	+	+	C
	Duhaylongsod, et al., 1995b	59 malignant cases 28 benign cases	for lesions < 4 cm Se=97% Sp=81% accuracy=92%			internal	+	+	C
	Gupta, et al., 1996	45 malignant cases 16 benign cases	Se=93% Sp=88% accuracy=92%			internal	+	unclear	C

Se, sensitivity
Sp, specificity
TTNA, transthoracic needle aspiration biopsy

* operating characteristics defined in Appendix 2: Assessing Diagnostic Technologies, pages 5-7
** Appendix 2, page 8
*** Appendix 2, page 9

Table 6 Diagnostic accuracy and diagnostic thinking efficacy of PET and its neuroimaging alternatives

Notes: All of the studies that evaluated a diagnostic test against the standard of histopathology fully met evidence based medicine and other methodologic quality criteria (i.e., received methodology grades of A or B). PET and other studies evaluating the new technology against clinical criteria for dementia of the Alzheimer s type would receive methodology grades of A or B, with the exception of the absence of histopathologic diagnosis.

Neuroimaging Test	Diagnostic Standard Used in Evaluation Studies		Characteristics
	Histopathology	Clinical criteria	
CT	x		Se = 94%; Sp = 93.5% (AD-specific orientation; AD vs normal controls and other dementias) <i>Jobst, et al., 1994</i>
SPECT	x	x	Se = 96%; Sp = 89% (AD vs normal controls and other dementias) <i>Jobst, et al., 1994</i> Sp = 89% • all probable AD, Se = 43% • probable AD < 80 years, Se = 56% • probable AD > 80 years, Se = 29% • SPECT contributed to 8% of final diagnoses <i>Van Gool, et al., 1995</i>
CT + SPECT	x		Se = 90%; Sp = 97% (AD vs normal controls and other dementias) <i>Jobst, et al., 1994</i>
PET		x	Se = 94.6; Sp = 97% ("robust ratio"; DAT vs normal controls) <i>Herholz, et al., 1993</i> Post test probability of disease, positive test = 90%; posttest probability, negative test = 10% in patients with pretest probability of disease = 50% (neural net; DAT vs normal controls) <i>Kippenhan, et al., 1994</i> Se = 94%; Sp = 79% (4 image patterns typical of DAT; DAT vs normal controls) <i>Salmon, et al., 1994</i> Se = 94%; Sp = 53% (4 image patterns typical of DAT; DAT vs non-DAT dementia controls) <i>Salmon, et al., 1994</i> Se = 94%; Sp = 99% (stereotactic surface projections; DAT vs non-DAT controls) <i>Burdette, et al., 1996</i>
PET vs CT		x	PET: Se = 97%; Sp = 84% (qualitative) CT: Se = 86%; Sp = 28% (cortical atrophy) (DAT vs normal controls) <i>Fazekas, et al., 1989</i>
PET vs MRI		x	PET: Se = 97%; Sp = 84% (qualitative) MRI: Se = 92%; Sp = 60% (ventricular atrophy) (DAT vs normal controls) <i>Fazekas, et al., 1989</i>
PET vs SPECT		x	PET: Se = 80%; Sp = 100% (typical functional pattern) SPECT: Se = 80%; Sp = 65% (typical functional pattern) (DAT vs normal controls and vascular dementia) <i>Mielke, et al., 1994</i>

Se = sensitivity; Sp = specificity
AD, Alzheimers disease
MRI, magnetic resonance imaging
CT, computed tomography
SPECT, single photon emission computed tomography
DAT, dementia of the Alzheimers type

VIII. REFERENCES

A. General

Banta HD, Luce BR. A system for health care technology assessment. in: *Health Care Technology and its Assessment*. Oxford University Press, 1993, p. 61-82.

Chalmers, TC. PET scans and technology assessment. *Journal of the American Medical Association* 1988;260:2713-5.

Cooper LS, Chalmers TC, McCally M, Berrier J, Sacks HS. The poor quality of early evaluations of magnetic resonance imaging. *Journal of the American Medical Association* 1988;259:3277-80.

Egglin TKP, Feinstein AR. Context bias: a problem in diagnostic radiology. *JAMA* 1996; 276:1752-5.

Eysenck HJ. Meta-analysis and its problems. *British Medical Journal* 1994;309:789-92.

Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Medical Decision Making* 1991; 11:88-94.

Goodman C, Snider GL, Flynn K: *Primer: Technology Assessment In VA*. Management Decision and Research, Boston MA; VA Health Services Research and Development Service, Washington DC, 1996

Hasselblad V, Hedges LV. Meta-analysis of screening and diagnostic tests. *Psychological Bulletin* 1995;117(1):167-78.

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

Hoffman RM, Kent DL, Deyo RA. Diagnostic accuracy and clinical utility of thermography for lumbar radiculopathy: a meta-analysis. *Spine* 1992;16:623-8.

Huston P, Moher D. Redundancy, disaggregation, and the integrity of medical research. *Lancet* 1996;347:1024-26.

Institute of Medicine. *Assessing Medical Technologies*. National Academy Press, Washington DC, 1985.

Kent DL, Haynor DR, Longstreth WT, Larson EB. The clinical efficacy of magnetic resonance imaging in neuroimaging. *Annals of Internal Medicine* 1994;120:856-71.

Kent DL, Larson EB. Disease, level of impact, and quality of research methods: three dimensions of clinical efficacy assessment applied to magnetic resonance imaging. *Investigative Radiology* 1992;27:245-54.

Kippenhan JS, Barker WW, Pascal S, Nagel J, Duara R. Evaluation of a neural-network classifier for PET scans of normal and Alzheimer's disease subjects. *Journal of Nuclear Medicine* 1992; 33:1459-67.

Kumar A, Schapiro MB, Grady C, Haxby JV, Wagner E, Salerno JA, et al. High-resolution PET studies in Alzheimer's disease. *Neuropsychopharmacology* 1991;4:35-46.

Light RJ, Pillemer DB. *Summing Up: The Science of Reviewing Research*. Harvard University Press, Cambridge, Massachusetts 1984.

Links JM, Devous MD. Detection and comparison of patterns in images. *Journal of Nuclear Medicine* 1994;35:16-17.

McMaster University Health Sciences Centre Department of Clinical Epidemiology and Biostatistics. How to read clinical journals: II. To learn about a diagnostic test. *Canadian Medical Journal* 1981;124:703-10.

Mulrow CD. The medical review article: state of the science. *Annals of Internal Medicine* 1987; 106:485-8.

Mulrow CD. Rationale for systematic reviews. *British Medical Journal* 1994;309:597-9.

Office of Technology Assessment: *Strategies for Medical Technology Assessment*. Pub. No. OTA-H-181. U.S. Government Printing Office, Washington DC, 1982.

Powers WJ, Berg L, Perlmutter JS, Raichle ME. Technology assessment revisited: Does positron emission tomography have proven clinical efficacy? *Neurology* 1991;41:1339-40.

Slavin RE. Best-evidence synthesis: an alternative to meta-analytic and traditional reviews. *Educational Researcher* 1986;15:5-11.

Slavin RE. Best evidence synthesis: an intelligent alternative to meta-analysis. *Journal of Clinical Epidemiology* 1995;1:9-18.

Thornbury JR, Kido DK, Mushlin AI, Phelps CE, Mooney C, Fryback DG. Increasing the scientific quality of clinical efficacy studies of magnetic resonance imaging. *Investigative Radiology* 1991;26:829-35.

B. Head and Neck Cancer Studies (Table 1)

Benchaou M, Lehmann W, Slosman DO, Becker M, Lemoine R, et al. The role of FDG-PET in the preoperative assessment of N-staging in head and neck cancer. *Acta Otolaryngologica* 1996;116:332-5.

Braams JW, Pruijm J, Freling NJM, Nikkels PGJ, Roodenburg JLN, Boering G, et al. Detection of lymph node metastases of squamous cell cancer of the head and neck with FDG-PET and MRI. *Journal of Nuclear Medicine* 1995;36:211-6.

Lapela M, Grenman R, Kurki T, Joensuu H, Leskinen J, Lindholm P, et al. Head and neck cancer: detection of recurrence with PET and 2-[F-18]Fluoro-2-deoxy-D-glucose. *Radiology* 1995; 197:205-11.

Laubenbacher C, Saumweber D, Wagner-Manslau C, Kau RJ, Herz M, Avril N, et al. Comparison of fluorine-18-fluorodeoxyglucose PET, MRI and endoscopy for staging head and neck squamous-cell carcinomas. *Journal of Nuclear Medicine* 1995;36:1747-57.

McGuirt WF, Williams DW, Keyes JW, Greven KM, Watson NE, Geisinger KR, Cappellari, JO. A comparative diagnostic study of head and neck nodal metastases using positron emission tomography. *Laryngoscope* 1995;105:373-7.

Rege S, Maass A, Chaiken L, Hoh CK, Choi Y, Lufkin R, et al. Use of positron emission tomography with fluorodeoxyglucose in patients with extracranial head and neck cancers. *Cancer* 1994;73:3047-58.

C. Colorectal Cancer Studies (Table 2)

Corman ML, Galandiuk S, Block GE, Prager ED, Weiner GJ, Kahn D, et al. Immunoscintigraphy with ¹¹¹In-satumomab pendetide in patients with colorectal adenocarcinoma: performance and impact on clinical management. *Diseases of the Colon and Rectum* 1994;37:129-37.

Falk PM, Gupta NC, Thorson AG, Frick MP, Bowman BM, Christensen MA, et al. Positron emission tomography for preoperative staging of colorectal carcinoma. *Diseases of the Colon and Rectum* 1994;37:153-6.

Ito K, Kato T, Tadokoro M, Ishiguchi T, Oshima M, Ishigaki T, et al. Recurrent rectal cancer and scar: differentiation with PET and MR imaging. *Radiology* 1992;182:549-52.

Hawes RH. New staging techniques: endoscopic ultrasound. *Cancer* 1993;71:4207-13.

Hernandez-Socorro CR, Guerra C, Hernandez-Romero J, Rey A, Lopez-Facal P, Alvarez-Santullano V. Colorectal carcinomas: diagnosis and preoperative staging by hydrocolonic sonography. *Surgery* 1995;117:609-15

Lai DTM, Fulham M, Stephen MS, Chu K-M, Solomon M, Thompson JF, et al. The role of whole-body positron emission tomography with [¹⁸F]fluorodeoxyglucose in identifying operable colorectal cancer metastases to the liver. *Archives of Surgery* 1996;131:703-7.

Nattinger AB. Colon Cancer Screening and Detection. In Panzer RJ, Black ER, Griner PF, eds. *Diagnostic Strategies for Common Medical Problems*. American College of Physicians, Philadelphia, 1991.

Panzer RJ. Hepatic Metastases. In Panzer RJ, Black ER, Griner PF, eds. *Diagnostic Strategies for Common Medical Problems*. American College of Physicians, Philadelphia, 1991.

Rafaelsen SR, Kronborg O, Larsen C, Fenger C. Intraoperative ultrasonography in detection of hepatic metastases from colorectal cancer. *Diseases of the Colon and Rectum* 1995;38:355-60.

Schiepers C, Penninckx F, De Vadder N, Mercks E, Mortelmans L, Bormans G, et al. Contribution of PET in the diagnosis of recurrent colorectal cancer: comparison with conventional imaging. *European Journal of Surgical Oncology* 1995;21:517-22.

Schlag P, Lehner B, Strauss LG, Georgi P, Herfarth C. Scar or recurrent rectal cancer: positron emission tomography is more helpful than immunoscintigraphy. *Archives of Surgery* 1989; 124:197-200.

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

Strauss LG, Clorius JH, Schlag P, Lehner B, Kimmig B, Egenhart R, et al. Recurrence of colorectal tumors: PET evaluation. *Radiology* 1989;170:329-32.

Vitola JV, Delbeke D, Sandler MP, Campbell MG, Powers TA, Wright JK, et al. Positron emission tomography to stage suspected metastatic colorectal carcinoma to the liver. *American Journal of Surgery* 1996;171:21-6.

Yonakura Y, Benua RS, Brill AB, Som P, Yeh SDJ, Kemeny NE, et al. Increased accumulation of 2-deoxy-2-[¹⁸F]fluoro-D-glucose in liver metastases from colon cancer. *Journal of Nuclear Medicine* 1982;23:1133-7.

D. Breast Cancer Studies (Table 3)

Adler LP, Crowe JP, Al-Kaisi NK, Sunshine JL. Evaluation of breast masses and axillary lymph nodes with [¹⁸F] 2-deoxy-2-fluoro-D-glucose PET. *Radiology*. 1993;187:743-50.

Avril N, Dose J, Jänicke F, Ziegler S, Römer W, Weber W, et al. Assessment of axillary lymph node involvement in breast cancer patients with positron emission tomography using radiolabeled 2-(fluorine-18)-fluoro-2-deoxy-D-glucose. *Journal of the National Cancer Institute*. 1996;88(17):1204-9.

Avril N, Dose J, Jänicke F, Bense S, Ziegler S, Laubenbacher C, et al. Metabolic characterization of breast tumors with positron emission tomography using F-18 fluorodeoxyglucose. *Journal of Clinical Oncology*. 1996;14:1848-57.

Nieweg OE, Kim EE, Wong WH, Broussard WF, Singletary SE, Hortobagyi GN, et al. Positron emission tomography with fluorine-18-deoxyglucose in the detection and staging of breast cancer. *Cancer*. 1993;71:3920-5.

Scheidhauer K, Scharl A, Peitzyk U, Wagner R, Göhring UJ, Schomäcker K, et al. Qualitative [¹⁸F]FDG positron emission tomography in primary breast cancer: clinical relevance and practicability. *European Journal of Nuclear Medicine*. 1996;23(6):618-23.

E. Lung Cancer Studies (Table 4)

Chin R, Ward R, Keyes JW, Choplin RH, Reed JC, Wallenhaupt, et al. Mediastinal staging of non-small-cell lung cancer with positron emission tomography. *American Journal of Respiratory and Critical Care Medicine* 1995; 152:2090-6.

Inoue T, Kim EE, Komaki R, Wong FCL, Bassa P, Wong W, et al. Detecting recurrent or residual lung cancer with FDG-PET. *Journal of Nuclear Medicine* 1995; 36:788-93.

Knight SB, Delbeke D, Stewart JR, Sandler MP. Evaluation of pulmonary lesions with FDG-PET: Comparison of findings in patients with and without a history of prior malignancy. *Chest* 1996;109:982-8.

Kubota K, Matsuzawa T, Fujiwara T, Ito M, Hatazawa J, Ishiwata K, et al. Differential diagnosis of lung tumor with positron emission tomography: a prospective study. *Journal of Nuclear Medicine* 1990; 31:1927-33.

Patz EF, Lowe VJ, Hoffman JM, Paine SS, Harris LK, Goodman PC. Persistent or recurrent bronchogenic carcinoma: detection with PET and 2-[F-18]-2-deoxy-D-glucose. *Radiology* 1994; 191:379-82.

Patz EF, Lowe VJ, Goodman PC, Herndon J. Thoracic nodal staging with pet imaging with ¹⁸F-FDG in patients with bronchogenic carcinoma. *Chest* 1995; 108:1617-21.

Sazon DAD, Santiago SM, Soo Hoo GW, Khonsary A, Brown C, Mandelkern M, *et al.* Fluorodeoxyglucose-positron emission tomography in the detection and staging of lung cancer. *American Journal of Respiratory and Critical Care Medicine* 1996; 153:417-21.

Scott W, Schwabe JL, Gupta NC, Dewan NA, Reeb SD, Sugimoto JT, *et al.* Positron emission tomography of lung tumors and mediastinal lymph nodes using F-18-fluorodeoxyglucose. *Annals of Thoracic Surgery* 1994; 58:698-703.

Scott WJ, Gobar LS, Terry JD, Dewan NA, Sunderland JJ. Mediastinal lymph node staging of non-small-cell lung cancer: a prospective comparison of computed tomography and positron emission tomography. *The Journal of Thoracic and Cardiovascular Surgery* 1996; 111:642-8.

Slosman DO, Spiliopoulos A, Couson F, Nicod L, Louis O, Lemoine R, *et al.* Satellite PET and lung cancer: a prospective study in surgical patients. *Nuclear Medicine Communications* 1993; 14:955-61.

Valk PE, Pounds TR, Hopkins DM, Haseman MK, Hofer GA, Greiss HB, *et al.* Staging non-small cell lung cancer by whole-body positron emission tomographic imaging. *Annals of Thoracic Surgery* 1995; 60:1573-82.

Wahl R, Quint LE, Greenough RL, Meyer CR, White RI, Orringer MB. Staging of mediastinal non-small cell lung cancer with FDG PET, CT, and fusion images: preliminary prospective evaluation. *Radiology* 1994; 191:371-7.

F. Solitary Pulmonary Nodules (Table 5)

Bury T, Dowlati A, Paulus P, Corhay JL, Benoit T, Kayembe JM, *et al.* Evaluation of the solitary pulmonary nodule by positron emission tomography imaging. *European Respiratory Journal*. 1996;9:410-14.

Dewan NA, Reeb SD, Gupta NC, Gobar LS, Scott WJ. PET-FDG imaging and transthoracic needle lung aspiration biopsy in evaluation of pulmonary lesions: a comparative risk-benefit analysis. *Chest*. 1995;108:441-6.

Duhaylonsod FG, Lowe VJ, Patz EF, Vaughn AL, Coleman RE, Wolfe WG. Detection of primary and recurrent lung cancer by means of f-18 fluorodeoxyglucose positron emission tomography (FDG PET). *The Journal of Thoracic and Cardiovascular Surgery*. 1995;110:130-140. (b)

Gupta NC, Maloof J, Gunel E. Probability of malignancy in solitary pulmonary nodules using fluorine-18-FDG and PET. *The Journal of Nuclear Medicine*. 1996;37:943-8.

G. Alzheimer s Disease (Table 6)

Burdette JH, Minoshima S, Vander Borgh T, Tran DD, Kuhl DE. Alzheimer disease: improved visual interpretation of PET images by using three-dimensional stereotaxic surface projections. *Radiology* 1996;198:837-43.

Fazekas F, Alavi A, Chawluk JB, Zimmerman RA, Hackney D, Bilaniuk L, et al. Comparison of CT, MR, and PET in Alzheimer's dementia and normal aging. *Journal of Nuclear Medicine* 1989; 1607-15.

Herholz K, Perani D, Salmon E, Granck G, Fazio F, Heiss WD, et al. Comparability of FDG PET studies in probable Alzheimer's disease. *Journal of Nuclear Medicine* 1993;34:1460-6.

Jobst KA, Hindley NJ, King E, Smith AD. The diagnosis of Alzheimer's disease: a question of image? *Journal of Clinical Psychiatry* 1994;55(11, suppl):22-31.

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. *Journal of Nuclear Medicine* 1994;35:7-15.

Mielke R, Pietrzyk U, Jacobs A, Fink, GR, Ichimiya A, Kessler J, et al. HMPAO SPET and FDG PET in Alzheimer's disease and vascular dementia: comparison of perfusion and metabolic pattern. *European Journal of Nuclear Medicine* 1994;1053-60.

Salmon E, Sadzot B, Mazuet P, Degueldre C, Lemaire C, Rigo P, et al. Differential diagnosis of Alzheimer's disease with PET. *Journal of Nuclear Medicine* 1994;35:391-98.

Van Gool WA, Walstra GJM, Teunisse S, Van der Zant FM, Weinstein HC, Van Royen EA. Diagnosing Alzheimer's disease in elderly, mildly demented patients: the impact of routine single photon emission computed tomography. *Journal of Neurology* 1995;242:401-5.