

Surveillance of Cancer Patients with Imaging: Self-Evident or Evidence-Based?

One of the few universal findings in human cancers is the close relationship between tumor stage at diagnosis and outcome. This relationship is the basis of the TNM system, which is used worldwide to guide cancer therapy. Many solid tumors can be cured if they have not formed metastases, but only very few when there are distant metastases. In addition, the size of the tumor at diagnosis, the degree of infiltration of neighboring structures, and the presence, location, and number of lymph node metastases are strong prognostic factors in virtually all solid tumors. Similarly, stage at diagnosis correlates with outcome in malignant lymphomas. On the basis of these data, many consider early detection as perhaps the most important goal to improve the outcome of

See page 1518

cancer patients. Screening of asymptomatic subjects has been successful for some cancers, such as cervical cancer, and has shown promise in screening high-risk patients for lung cancer (1,2), but the high false-positive rate of screening tests in unselected populations has made widespread screening impractical for many malignancies. Patients with a history of cancer not only are at risk for disease recurrence but also have a substantially higher risk of developing second primary tumors. Because of the higher incidence rates, false-positive test results are expected to be significantly lower than in patients without a history of cancer. As a consequence, it is intuitive to

use surveillance in asymptomatic cancer patients to detect recurrence or a second primary tumor at a potentially curable stage. In fact, virtually all cancer treatment guidelines include some form of surveillance as part of their recommendations. The primary goal is to detect recurrent tumors or second primary tumors when they are small and can be treated with curative intent using local therapies such as surgery or radiotherapy. Delayed treatment of recurrence may result in failure to cure the patient or excessive morbidity from the treatment (e.g., in head and neck cancer recurrence). However, early detection of incurable metastatic disease potentially may also be beneficial, if complications such as fractures due to bone metastases can be prevented by palliative therapy. Finally, surveillance is psychologically important for many patients to assure them that they are free of disease and that some unspecific symptoms do not indicate recurrent cancer.

The use of surveillance appears therefore so obvious that surprisingly few studies have tested its actual benefits. Most of these studies have unexpectedly indicated that there is no definite benefit to surveillance after primary treatment of several cancers (3,4). The perhaps best-studied example is the use of bone scans in patients with a history of breast cancer. Bone metastases are common in patients with breast cancer, and untreated bone metastases can cause debilitating complications. Bone scans can detect asymptomatic bone metastases in the whole body. Radiotherapy is a well-established palliative therapy for bone metastases and can effectively treat the symptoms and complications. However, 2 randomized trials—each including more than 1,000 patients—found no evidence that bone scintigraphy and chest radiography in asymptomatic patients improved overall survival or quality of life (5,6). Progression-free survival was shorter in patients receiving routine follow-up with chest radiography and bone scanning but did not lead to improved overall survival (6). A more recent Finish randomized trial including 472 breast cancer patients found that neither the frequency of visits nor the

intensity of diagnostic examinations (including blood counts, erythrocyte sedimentation rate, liver enzymes, tumor marker CA 15-3, chest radiography, liver ultrasound, and bone scanning) had any effect on disease-free or overall survival of patients (7). This unexpected finding has been explained by the limited diagnostic accuracy of the studied imaging tests and by the fact that metastatic breast cancer is incurable and that, therefore, early detection of metastatic disease does not lead to improved survival (8).

This example illustrates that the clinical impact of surveillance imaging depends not only on the diagnostic accuracy of the imaging test but also on the availability of effective therapies for recurrent disease. Furthermore, the clinical utility of surveillance imaging may be limited by the pattern of metastatic spread. If metastases develop simultaneously at several sites, surveillance imaging may have less impact on patient outcome than in the case of recurrences at a limited number of sites (“oligometastatic disease”). In the latter case, surgery or radiotherapy may be curative, whereas in the former case, palliative systemic chemotherapy may be the only therapeutic option. Therefore, the impact of surveillance imaging is expected to be more significant in oligometastatic disease.

Another factor is the time to recurrence and the growth rate of recurrent tumors. If the risk for recurrent disease declines slowly over time, surveillance imaging will have to be performed over long periods and surveillance may not be cost-effective. Conversely, if recurrent tumors grow rapidly, they may not be detected early enough to be treated with curative intent. Finally, use of surveillance imaging is unlikely to be cost-effective and may even be harmful to the patient if the overall risk for recurrence is low. In this case, false-positive results may cause many unnecessary invasive procedures for verification of the imaging findings. Furthermore, radiation exposure by repeated follow-up imaging may not be negligible, at least in children and young adults (9). Thus, choosing the right population for surveillance would seem highly

Received Jul. 5, 2013; revision accepted Jul. 16, 2013.

For correspondence or reprints contact: Wolfgang A. Weber, Molecular Imaging and Therapy Service, Memorial Sloan-Kettering Cancer Center, 1275 York Ave., New York, NY 10065.

E-mail: weberw@mskcc.org

Published online Aug. 8, 2013.

COPYRIGHT © 2013 by the Society of Nuclear Medicine and Molecular Imaging, Inc.

DOI: 10.2967/jnumed.113.127183

appropriate. Effective therapies for recurrent disease should be available, the risk of recurrence should be substantial, and the impact of failing to detect recurrence early should be major.

Considering these various factors, colorectal cancer appears well suited for surveillance imaging. Recurrences are frequently local or in the liver, and the risk of recurrence increases with stage at diagnosis. If recurrent tumors or metastases are completely resected, long-term progression-free survival and even cure can be achieved. Consequently, surveillance using tumor markers, CT, ultrasound, and endoscopy is commonly applied clinically after curative resection of colorectal cancer. However, even for this disease, data on the benefits of routine surveillance are conflicting (3). A recent systematic review has concluded that the status of "evidence is still weak" for the benefits of surveillance after curative resection of colorectal cancer (3): "the efficacy of each of the various elements of surveillance is not well supported in isolation, and the [randomized controlled trials] performed to date have significant heterogeneity in terms of the follow-up programs applied." Thus, there is a lack of evidence not only for PET/CT imaging for surveillance but also for other imaging modalities and, strictly speaking, the concept of surveillance in general. However, it is important to note that in the single prospective randomized trial of PET in colorectal cancer surveillance, the recurrences were detected earlier in the group having imaging surveillance, and surgery for recurrent disease was performed more frequently on those in the PET group (15/23 [65%] vs. 2/21 [9.5%], $P < 0.0001$). The frequency of curative resection of recurrences was higher in the PET group (43.8% vs. 9.5%, $P < 0.01$). These data suggest that PET is useful in the follow-up of patients with colorectal cancer, but the data are limited by study sample size ($n = 65$ per group) (10).

At first glance, the obvious approach to studying the clinical benefits of surveillance imaging is a randomized controlled trial with overall survival as the endpoint or with important intermediate endpoints such as those in the colorectal cancer trial. More careful consideration, however, indicates that randomized trials will be challenging and must be carefully designed to provide relevant information. The impact of surveillance imaging will depend not only on the accuracy of the imaging test but also on the pattern of recurrence and

the availability of effective treatments in cases of recurrence. All 3 parameters can easily change over time. Imaging modalities undergo rapid technologic improvements, new approaches to treatment of the primary tumor may affect the pattern of recurrence, and therapy for recurrence may become more effective or less toxic. Thus, at the completion of a randomized trial the results may no longer be considered as relevant to current practice. Furthermore, patients may not accept random assignment to a follow-up strategy that involves only limited surveillance. This points to the need to do adequately sized multicenter trials that accrue rapidly while imaging and treatments are relatively stable. Finally, it is not clear who would fund randomized trials of surveillance imaging, because large patient populations would be required for each cancer type and these trials may have to be repeated when there are changes in the treatment of the disease at the time of diagnosis or at the time of recurrence. However, given the investment that society makes in cancer therapy and imaging, having well-designed and -powered trials in higher-risk groups of asymptomatic patients in diseases for which effective salvage therapies exist would be a rational societal investment. Head and neck cancer and colorectal cancer at high risk of recurrence are prime candidates for adequately powered prospective trials that can be completed in a reasonable period.

Expecting a single answer regarding the value of surveillance imaging, and concluding that it is overall not useful, may be akin to throwing the baby out with the bathwater. It is quite possible that in several diseases, such as advanced head and neck cancer and colorectal cancer, surveillance is highly rational and appropriate. By contrast, it may not be useful in low-risk breast cancer or high-risk lung cancers. Answers from the past, however, must continue to be reassessed as treatments improve.

Clearly, we cannot afford to do prospective randomized trials of screening in all patient groups, and only a small number of trials may be possible. A more realistic approach is to study the diagnostic accuracy of surveillance imaging and to determine how often a positive finding on surveillance imaging was true-positive, how often disease recurrence was diagnosed by imaging only, and how often the recurrent disease was amenable to (potentially) curative therapy. These kinds of questions can be addressed in significantly smaller randomized or nonrandomized trials. However,

as pointed out by Patel et al. (11) in this issue of *The Journal of Nuclear Medicine*, as well as by others (3,4), it is necessary to better standardize the design and endpoints of these studies. The conventional design of diagnostic accuracy studies mandates a single unbiased reference test that is evaluated independently from the index test (12). By definition, such a test does not exist for whole-body imaging of cancer. It is not possible to do an autopsy in a living patient to secure a gold standard. Strictly speaking, it is therefore impossible to perform a high-quality study on the diagnostic accuracy of surveillance imaging with whole-body PET/CT (13).

Although the terms *sensitivity* and *specificity* are frequently applied to whole-body staging, they are not well defined in this setting because there is no reference test to exclude metastases. Furthermore, there is frequently the problem that a study is both true-positive and false-positive. For example, an imaging test may indicate the presence of a lung lesion and a liver lesion, both of which are suggestive of metastatic disease. If histologic evaluation confirms the presence of a liver metastasis but the lung lesion turns out to be a granuloma, it becomes arbitrary whether the result for this patient should be classified as true-positive or false-positive. Researchers have used various approaches toward dealing with this problem and have added region-based or lesion-based analyses for whole-body imaging studies. However, these are not standardized, and the reported values for sensitivity and specificity of different studies may therefore not be comparable. Sometimes, as well, the additional finding on PET is a new primary tumor unrelated to the original.

A further problem is the reference standard used to exclude metastatic disease. Frequently, the findings of all available imaging modalities and follow-up are used to exclude metastatic disease. However, the "available imaging modalities" and their quality frequently vary across studies, and the length of follow-up may be different. Thus, the reported sensitivities cannot be compared. Perhaps even more importantly, the sensitivities become dependent on the diagnostic performance of the other imaging modalities. For example, somatostatin receptor scintigraphy was reported to provide a sensitivity of more than 90% for detection and staging of various neuroendocrine tumors when it was introduced in the early 1990s (14). More recent studies have found a much lower sensitivity, because

CT and MR imaging have made considerable progress and the total number of lesions has become much higher (15).

Because of these fundamental problems, it may be better to avoid the terms *sensitivity* and *specificity* in imaging studies for whole-body cancer staging or surveillance. A more robust approach to data analysis could be the systematic validation of discrepant findings of 2 tests for whole-body staging or surveillance. In this approach, all cases that were positive with one test but negative with the other would be verified with a reference test, preferably a biopsy. Using this approach, the relative diagnostic accuracy of the 2 tests can be compared, even if the true sensitivity and specificity are unknown (16).

In conclusion, the use of imaging for surveillance clearly needs further study. Future trials evaluating ¹⁸F-FDG PET/CT or other imaging modalities for surveillance should take into account the probability of recurrence, the distribution of metastatic disease, and the availability of effective therapies for recurrent disease. In addition, the risks of a delay in appropriate therapy must be considered. Patients with head and neck cancer who lose their larynx or tongue, when an earlier detection of recurrence could have conserved their organ, do not necessarily show up in studies in which survival is the only outcome assessed. In appropriately selected high-risk populations, ¹⁸F-FDG PET/CT may improve outcome and be cost-effective, as is suggested by the limited data in colorectal cancer. It will be challenging and prohibitively expensive to prove this hypothesis for all cancers in randomized trials. However, several randomized trials on patients in whom the risk of recurrence is high and effective alternative therapies are available would seem appropriate for consideration. It will

be also important to systematically compare the accuracy of ¹⁸F-FDG PET/CT-based surveillance with surveillance strategies that are generally accepted as being the standard of care. For these trials to be successful, it will be crucial to establish a standardized methodology to assess the diagnostic performance and impact of ¹⁸F-FDG PET/CT, as the commonly used paradigms of diagnostic accuracy studies are not well applicable to whole-body staging and surveillance. Ultimately, such data can lead to risk-of-recurrence/benefit-of-early-therapy/risk-of-delayed-diagnosis-adapted algorithms in which surveillance is used for precisely defined patient groups who likely may benefit.

Wolfgang A. Weber

*Memorial Sloan-Kettering Cancer Center
New York, New York*

Richard L. Wahl

*Johns Hopkins University School of
Medicine
Baltimore, Maryland*

REFERENCES

1. Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365:395–409.
2. Church TR, Black WC, Aberle DR, et al. Results of initial low-dose computed tomographic screening for lung cancer. *N Engl J Med*. 2013;368:1980–1991.
3. Baca B, Beart RW Jr, Etzioni DA. Surveillance after colorectal cancer resection: a systematic review. *Dis Colon Rectum*. 2011;54:1036–1048.
4. Srikantharajah D, Ghuman A, Nagendran M, Maruthappu M. Is computed tomography follow-up of patients after lobectomy for non-small cell lung cancer of benefit in terms of survival? *Interact Cardiovasc Thorac Surg*. 2012;15:893–898.
5. Givio Investigators. Impact of follow-up testing on survival and health-related quality of life in breast

cancer patients: a multicenter randomized controlled trial. The GIVIO Investigators. *JAMA*. 1994;271:1587–1592.

6. Rosselli Del Turco M, Palli D, Cariddi A, Ciatto S, Pacini P, Distante V. Intensive diagnostic follow-up after treatment of primary breast cancer: a randomized trial. National Research Council Project on Breast Cancer Follow-up. *JAMA*. 1994;271:1593–1597.
7. Kokko R, Hakama M, Holli K. Follow-up cost of breast cancer patients with localized disease after primary treatment: a randomized trial. *Breast Cancer Res Treat*. 2005;93:255–260.
8. Jochelson M, Hayes DF, Ganz PA. Surveillance and monitoring in breast cancer survivors: maximizing benefit and minimizing harm. *Am Soc Clin Oncol Educ Book*. 2013:13–18.
9. Shenoy P, Sinha R, Tumeh JW, Lechowicz MJ, Flowers CR. Surveillance computed tomography scans for patients with lymphoma: is the risk worth the benefits? *Clin Lymphoma Myeloma Leuk*. 2010;10:270–277.
10. Sobhani I, Tiret E, Lebtahi R, et al. Early detection of recurrence by ¹⁸F-FDG-PET in the follow-up of patients with colorectal cancer. *Br J Cancer*. 2008;98:875–880.
11. Patel K, Hadar N, Lee J, Siegel BA, Hillner BE, Lau J. The lack of evidence for PET or PET/CT surveillance of patients with treated lymphoma, colorectal cancer, and head and neck cancer: a systematic review. *J Nucl Med*. 2013;54:1518–1527.
12. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol*. 2003;3:25.
13. Weber WA. Is there evidence for evidence-based medical imaging? *J Nucl Med*. 2011;52(suppl 2):74S–76S.
14. Schirmer WJ, Melvin WS, Rush RM, et al. Indium-111-pentetreotide scanning versus conventional imaging techniques for the localization of gastrinoma. *Surgery*. 1995;118:1105–1113.
15. Reidy-Lagunes DL, Gollub MJ, Saltz LB. Addition of octreotide functional imaging to cross-sectional computed tomography or magnetic resonance imaging for the detection of neuroendocrine tumors: added value or an anachronism? *J Clin Oncol*. 2011;29:e74–e75.
16. Vach W, Hoiland-Carlsen PF, Gerke O, Weber WA. Generating evidence for clinical benefit of PET/CT in diagnosing cancer patients. *J Nucl Med*. 2011;52(suppl 2):77S–85S.