Radioiodine Ablation Outcomes After Imaging with ^{123}I or ^{131}I: Is No News Good News?

TO THE EDITOR: In a recent paper (*1*), Silberstein reported data from his study assessing outcomes of radioiodine ablation in patients with differentiated thyroid carcinoma after imaging with 2 different isotopes. This study analyzed the results from 49 patients, 26 of whom received ^{123}I before ablation and 23 of whom received ^{131}I before ablation. Acknowledging the difficulties of adequately defining successful ablation, Silberstein reported that 81% of the patients receiving ^{123}I had a successful ablation, compared with 74% of the patients receiving ^{131}I, and that this difference was not statistically significant.

However, we would suggest that the author has overextrapolated from this result to the statement that "the same" ablation rate was achieved, irrespective of diagnostic agent. The logical conclusion of such a statement is that either agent could be used for the purpose, with no loss of patient benefit. Even if true, that conclusion is not demonstrated by Silberstein's study, as it is underpowered to detect what may be clinically significant differences between the techniques. What constitutes such a difference is always difficult to judge, but one might argue that a reduction in the ablation failure rate from 26% to 19% (i.e., nearly a 27% reduction in failures) is clinically significant. A simple power calculation (*2*) would have revealed that to detect the difference between 74% and 81% would require 479 patients for each diagnostic agent. Even if Silberstein had powered his study to look for a bigger difference of 15%, which we believe that most in the oncology community would agree represents a clinical improvement, achieving this difference would have required 71 patients for each diagnostic agent. The power calculations assume a 1-sided χ^2 test, 80% power, and a 0.05 significance level. Conversely, for the patient numbers Silberstein reported, the rate of successful ablations would have needed to rise to 100% for ^{123}I (compared with ^{131}I) for the difference between the techniques to reach statistical significance (Fisher exact test, $P = 0.014$).

The danger of interpreting absence of evidence as absence of negative effects has recently been highlighted in this journal by a letter in which Walter et al. (*3*) made a plea for adequately powered trials. We would add our voice to that plea: Silberstein's study set out to answer an important question that was never going to be answered with the number of patients recruited. When studies are limited by the small number of patients referred through a single hospital or unit, a multicenter approach is the option of choice. Small-scale studies not only represent a waste of resources but also can lead to incorrect conclusions.

REFERENCES

1. Silberstein EB. Comparison of outcomes after ^{123}I versus ^{131}I preablation imaging before radioiodine ablation in differentiated thyroid carcinoma. *J Nucl Med.* 2007; 48:1043–1046.
2. Machin D, Campbell M, Fayers P, et al. *Sample Size Tables for Clinical Studies.* Oxford, U.K.: Blackwell Science; 1997:19.
3. Walter MA, Muller-Brand J, Muller B. Antithyroid drugs and radioiodine and the absence of evidence [letter]. *J Nucl Med.* 2007;48:1403.

Maria T. Burniston
Daniel J. Wilson
Leeds Teaching Hospitals NHS Trust
Leeds, United Kingdom

DOI: 10.2967/jnumed.107.047076

REPLY: I appreciate the thoughtful comments directed to my paper (*1*), wherein I attempted to examine patient outcomes when imaging with ^{123}I versus ^{131}I before ^{131}I ablation. As they point out, it is always important to recognize the possibility, and actually the probability, of type I, or alpha, and type II, or beta, errors in any scientific inquiry.

I must first disagree with the authors' statement that I acknowledged any difficulty in adequately defining successful ablation. As stated in my paper, the determination of successful ablation required both negative diagnostic ^{131}I follow-up findings 6 mo after ablation, performed with the serum thyroid-stimulating hormone elevated in excess of 30 μIU/mL, and, simultaneously, an undetectable level of serum thyroglobulin in the absence of antithyroglobulin antibodies. I examined our data with, and without, the serum thyroglobulin requirement, in order to compare our findings to the majority of data on ^{131}I-induced thyroid ablation—data accumulated over many years and acquired without use of the thyroglobulin assay.

I found that complete ablation, assessed without determining the level of serum thyroglobulin, occurred in 88% of patients initially scanned with ^{123}I and 91% of patients initially scanned with ^{131}I. It is a dubious premise that adding 10 times more patients to the number examined, as suggested by Burniston and Wilson, would have led to the conclusion that the 3.4% difference I described between the 2 groups was real or significant. In the context of the question asked, the same comment applies to the 8.6% difference found in ablation rates between the ^{123}I and ^{131}I groups when these were assessed by both scintigraphic and thyroglobulin criteria. The ablation rates we found are well within the range of those noted in the unflawed studies cited in my paper and have been replicated many times, as a trip to PubMed or any textbook on the topic documents.

However, let us suppose, just for the moment, that we had the time and resources to study 10 times as many patients, as Burniston and Wilson would have us do, and discovered that the 8.6% difference we described between outcomes in the 2 groups was significant. What would be the implications? I believe that clinicians who do not have access to ^{123}I would not be surprised to find that even if stunning were to occur at the dosages used (a concept with which our data and those of our cited references disagree), it would hardly be clinically relevant, because the phenomenon would produce such a small decrease in ablation rates. Of course, we believe, in accordance with the data, that this stunning does not occur.

It is impossible, of course, to disagree with the point made by Burniston and Wilson that with 10 times as many patients studied in this, or any, scientific endeavor, there may occasionally be slightly different outcomes and data interpretations.