Notice of Retraction: Final Resolution

In October 2003, I issued a Notice of Retraction (1) for an article by Eric Poehlman (2). I was acting on a letter from the provost of the University of Vermont describing the outcome of the university’s investigation of allegations of fraudulent conduct of research. This investigation found that Dr. Poehlman had published fabricated data in Annals of Internal Medicine and 2 other journals. At the time, Dr. Poehlman claimed that he had not published fabricated data, but I felt that the University of Vermont findings were enough to warrant retraction. Subsequently, the Office of Research Integrity of the National Institutes of Health conducted an investigation that corroborated the earlier investigation and found that Dr. Poehlman had also included fabricated data in several federal grant proposals. As part of a negotiated settlement of criminal charges related to the fraudulent grant proposals, Dr. Poehlman wrote a letter in which he stated that he had fabricated the data in his 1995 Annals article. His letter, which appears here, shows beyond any possible doubt that the findings in the Annals article are deceptive and false.

Harold C. Sox, MD
Editor

References

TO THE EDITOR: I wish to acknowledge that the Editor of Annals of Internal Medicine appropriately retracted (1) my paper on changes in energy balance and body composition at menopause (2) after receiving notification from the University of Vermont that an investigation had determined that I had falsified and fabricated the reported data.

In addition, I request that this letter be published to ensure that the scientific community is notified that I accept full responsibility for this falsification and fabrication, and that my coauthors are publicly exonerated.

Eric T. Poehlman, PhD
Montréal, Québec H2L 3R8, Canada

References

Comments and Responses

Computed Tomography and Ultrasonography To Detect Appendicitis

TO THE EDITOR: In their meta-analysis, Terasawa and colleagues (1) suggested that overall, computed tomography (CT) is better than ultrasonography (1) for diagnosing appendicitis in patients with atypical presentations. Their findings will clarify clinical decision making for many but will complicate it for those practicing in low-resource settings. In many low-income countries, CT often costs many times a family’s monthly salary. It is important to carefully answer the following questions before we recommend CT as a standard of care.

First, what about cost and cost-effectiveness? In my practice setting, a focal appendiceal CT, which would probably tell the clinician whether the patient has appendicitis, costs about 6 times more than ultrasonography. When confronted with the broader question of what else could be causing the abdominal pain, the cost differences are even greater. A complete abdominal and pelvic CT scan with oral and intravenous contrasts costs about 18 times more than ultrasonography. It may therefore be less expensive to admit or observe a patient than to order a CT. The only patients who would benefit from immediate CT use are those who would receive surgery earlier with CT than with ultrasonography, a relatively small group. Those with negative CT results would still need admission or close follow-up, since no test has 100% specificity.

Second, is CT a better diagnostic test than ultrasonography for all patients, or is CT better in some while ultrasonography is better in others? Current literature needs to better define sensitivities and specificities for subgroups of patients. It would help to know whether CT is as good for a 30-year-old thin man as it is for a 30-year-old woman or a 60-year-old diabetic patient with nonspecific abdominal pain. Another issue is more specifically describing the degree of atypicality or the level of pretest probability. The literature does not determine how decisions should be made when a clinician thinks that a patient may have appendicitis but is not sure versus when he does not think a patient has appendicitis and is not sure.

Third, since most of the literature examined by Terasawa and colleagues originated in high-income countries, would their conclusions hold true for relatively leaner populations in low- and middle-income countries?

Junaid A. Razzak, MBBS
The Aga Khan University
Karachi, Pakistan

Reference
that $D = 5.2 - 0.49 \times S$ for CT and $D = 3.8 - 0.005 \times S$ for ultrasonography.

When we compared Terasawa and colleagues' original figure with ours, it was apparent that the diagnostic performance of CT is superior to that of ultrasonography in cases of appendicitis. Terasawa and colleagues stated that in most of the studies they reviewed, appendiceal diameter greater than 6 mm was used as the positive diagnostic criterion for appendicitis. But, judging from their Appendix Table 2, most of the studies seem to have several criteria. However, both regression coefficients $b$ are between $-1.0$ and $1.0$ ($P > 0.05$), suggesting no significant variation in diagnostic performance with threshold. The diagnostic log odds ratios (logit transformation of the ratio of positive to negative likelihood obtained by the Mantel–Haenszel method) of CT and ultrasonography are estimated to be $5.0$ (95% CI, 1.3 to 8.7) and $3.4$ (CI, 0.9 to 5.9), respectively, indicating again that CT is superior to ultrasonography.

We also created a funnel plot to explore the publication bias. The funnel plot is asymmetric, indicating that publication bias is likely. It is possible that some studies of small samples that indicated low diagnostic accuracy for these tests were not published.

Our additional analysis does not mean that Terasawa and colleagues’ conclusion needs alteration; rather, it strengthens their conclusion. They provided an elegant systematic review.

Toshiya Shiga, MD, PhD
Zen’ichiro Wajima, MD, PhD
Tetsuo Inoue, MD, PhD
Nippon Medical School, Chiba Hokusoh Hospital
Chiba 270-1613, Japan

Potential Financial Conflicts of Interest: None disclosed.

References

TO THE EDITOR: According to Terasawa and colleagues’ review (1), the overall sensitivity and specificity of CT are 0.94 and 0.95, respectively, and those of ultrasonography are 0.86 and 0.81, respectively. Although the authors concluded that CT was probably more accurate than ultrasonography, they also indicated that all of the included studies in their analysis had verification bias. The only way to avoid this bias is to construct a prospective study in which all patients have the definitive verification procedure. In this instance, the definitive procedure, surgery, is too invasive to apply universally. Fortunately, it is still possible to correct verification bias if we can obtain appropriate data from previous publications. Kosinski and Barnhart (2) proposed a global sensitivity analysis to study behavior of sensitivity and specificity while the numbers of patients with the disease among unverified groups vary over all possible values. This analysis can be performed if study authors report how many patients have their diagnosis verified by surgery. Only 3 of the 284 patients with unverified diagnoses had positive results on CT; the remaining 281 had negative CT results. For the 360 patients who had surgery, the sensitivity and specificity were 0.928 (298 of 321) and 0.667 (26 of 39), respectively. With regard to ultrasonography performance, I obtained appropriate information from 9 of 14 articles included in the systematic review. Of 620 patients in these 9 articles, 170 (27%) had unverified diagnoses. Six of these 170 patients had positive results on ultrasonography, and the remaining 164 had negative ultrasonography results. When I limited my analysis to the 450 patients with a verified diagnosis, sensitivity was 0.896 (354 of 395) and specificity was 0.582 (32 of 55). The crescent-shaped areas in the Figure indicate all possible values of sensitivity and specificity for each CT and ultrasonography scan, taking into account verification bias. The dashed lines in the crescent-shaped areas represent the possible combinations of sensitivity and specificity when the probability of acute appendicitis in the positive and negative unverified groups equals that in the corresponding verified groups. Since acute appendicitis is less prevalent among patients without surgery than among those with surgery, estimated sensitivity and specificity would lie within narrower intervals: 0.658 to 0.929 and 0.916 to 0.959, respectively, for CT and 0.726 to 0.897 and 0.773 to 0.889, respectively, for ultrasonography. As indicated in the Figure, all possible

Figure. Regions of all possible combinations of sensitivity and specificity as a result of a global sensitivity analysis.

CT = computed tomography; US = ultrasonography.
values of specificity for ultrasonography are lower than those for CT. On the other hand, the interval of CT sensitivity is too wide to determine which test is better without verification bias.

I can say with fair certainty that the overall CT sensitivity reported in Terasawa and colleagues’ systematic review is overestimated. To take a hypothetical example, if the prevalences of acute appendicitis are 0.5 and 0.1 among patients with unverified positive test results and those with unverified negative results, respectively, I could obtain the exact sensitivities and specificities of 0.854 and 0.955 for CT and 0.861 and 0.876 for ultrasonography, respectively. According to my analysis, CT had better specificity than ultrasonography even after correction for verification bias. However, there is still room for discussion regarding sensitivity because the sensitivity of CT in Terasawa and colleagues’ study was overestimated.

Masashi Goto, MD, MPH
Kyoto University Graduate School of Medicine
Kyoto 606-8507, Japan

Potential Financial Conflicts of Interest: None disclosed.

References

IN RESPONSE: We agree with Dr. Razzak that cost-effectiveness is critical to selection of appropriate imaging, as is local availability of imaging technologies. Furthermore, different imaging strategies may be appropriate in different subsets of patients. These areas would benefit from further investigation.

We agree with Drs. Shiga, Wajima, and Inoue that summary ROC curves (1) and funnel plots may be useful in meta-analysis. We had performed similar analyses but did not include the results because of several concerns. Summary ROC curves may be affected by the assumptions used in curve calculation and are less useful at the bedside than likelihood ratios, which enable understanding of disease probabilities. Asymmetrical funnel plots are affected by both publication bias and other factors that cause heterogeneity, including small study effects (2).

The global sensitivity analysis that Dr. Goto used (3) can apply to studies with partial verification bias, where not all participants’ diagnoses are confirmed by the single reference standard, but not to differential verification bias, as in the case of appendicitis. Under the global sensitivity analysis approach, the range of disease prevalence in patients with unverified diagnoses is applied to define the possible range of disease specificity for the test under consideration. This is appropriate if nothing is known about the patients who did not receive the reference standard (surgery). However, in the studies included in our review, most of the negative imaging results were verified by a secondary clinical reference standard (4). This differential verification bias will lead to some overestimation of overall accuracy, as we discussed in our paper. However, most patients with appendicitis should have developed worsening symptoms and therefore would have been identified through clinical follow-up. Dr. Goto’s assumption that the prevalence of undiagnosed appendicitis is 0.10 to 0.50 in patients who do not have surgery and do not have appendicitis identified on clinical follow-up is almost certainly not clinically sensible. The use of clinical follow-up is a valuable if imperfect reference standard and is clearly superior to the global sensitivity analysis assumption that nothing is known about these patients.

In summary, on the basis of the data, although some differential verification bias leads to overestimation of diagnostic accuracy for both CT and ultrasonography, CT scanning is a more sensitive and specific imaging method to detect appendicitis in patients with indeterminate clinical suspicion.

Teruhiko Terasawa, MD
National Hospital Organization and Nagoya Medical Center
Nagoya, Japan 460-0001

C. Craig Blackmore, MD, MPH
Harborview Medical Center
Seattle, WA 98104

Potential Financial Conflicts of Interest: None disclosed.

References

Defining the Role of Computed Tomographic Pulmonary Angiography in Suspected Pulmonary Embolism

TO THE EDITOR: I am concerned that the article by Moores and colleagues (1) conveys an ambiguous message about the utility of spiral computed tomographic pulmonary angiography (CTPA) for diagnosing pulmonary embolism (PE). The ambiguity results from seemingly different conclusions in the abstract, discussion, and Editors’ synopsis sections of the manuscript. The authors concluded their discussion by stating that withholding anticoagulation seems to be safe in patients managed by negative results on CTPA performed concurrently with imaging of the lower extremities (either ultrasonography or computed tomographic venography). The abstract that accompanies the manuscript, however, fails to mention the importance of concurrent lower-extremity imaging, stating only “it appears to be safe to withhold anticoagulation after negative CTPA results.” The Editors’ synopsis also fails to mention lower-extremity imaging, yet adds that concurrent clinical probability assessment is necessary by stating “withholding anticoagulation from patients with low to moderate probability of PE and negative results on CTPA appears reasonable.”

These incongruities are not trivial. Given that busy clinicians...
often only read the abstract and Editors’ synopsis sections of a manuscript, every effort should be made to ensure that these sections accurately reflect the authors’ conclusions. In the article by Moores and colleagues, clinicians who read only the abstract may conclude that negative results on CTPA can function as the sole diagnostic test for ruling out PE. Diagnostic strategies based on this erroneous conclusion will exclude lower-extremity imaging and clinical probability estimation, and subsequently, in some patients, venous thrombosis may not be diagnosed or treated.

Shaun D. Frost, MD
HealthPartners Medical Group and Clinics
Minneapolis, MN 55101

Potential Financial Conflicts of Interest: None disclosed.

References

TO THE EDITOR: In their careful and commendable meta-analysis of outcome studies that used CTPA in patients with suspected PE, Moores and colleagues (1) concluded that “it appears to be safe to withhold anticoagulation after negative CTPA results.” I believe that such a message goes beyond the conclusions that can be drawn from the authors’ data.

As Moores and colleagues acknowledged, among the 23 outcome studies they analyzed, only 1 used CTPA as the sole diagnostic test. Furthermore, in the 3 largest prospective studies (at least 350 patients in each study remained untreated), anticoagulation was withheld only when both CTPA and compression ultrasonography of the lower limbs yielded negative results (2–4). When it could be calculated, the prevalence of deep venous thrombosis detected by compression ultrasonography in patients with negative CTPA results was very low (0.8%) in 1 study (4), but it was significantly higher in 2 other studies (8.4% and 6.3%, respectively) (2, 5) and reached 18.8% in a study of 117 hospitalized patients (6). Therefore, until specific outcome studies demonstrate the opposite, using CTPA as the sole diagnostic test to withhold anticoagulation in patients with suspected PE should be considered unwise, not “safe.”

Philippe Girard, MD
Institut Mutualiste Montsouris
75014 Paris, France

Potential Financial Conflicts of Interest: None disclosed.

References


TO THE EDITOR: In many hospitals, CTPA is becoming the preferred diagnostic test for suspected acute PE. On the basis of their meta-analysis, Moores and colleagues (1) concluded that it appears to be safe to withhold anticoagulant therapy after negative CTPA results in patients with suspected PE. This may imply that CTPA could be used as the first and sole diagnostic test in these patients. We believe that this conclusion is premature for 2 reasons.

First, in consecutive patients with suspected PE, the sensitivity of CTPA is only 70% and the negative likelihood ratio is 0.3 (2). This likelihood ratio is comparable to that of a low-probability lung scan, a diagnostic result considered insufficient to safely exclude PE. The sensitivity of CTPA is only 30% for detection of isolated subsegmental emboli (3). These isolated subsegmental emboli are present in up to 30% of patients with documented PE, and their clinical significance cannot be excluded.

Second, CTPA should be part of a diagnostic strategy for the diagnosis of PE, not the sole diagnostic test. The combination of a low clinical probability score and a normal D-dimer level has been shown to safely exclude PE in 20% to 30% of patients (4). Therefore, using such an approach will prevent 20% to 30% of patients from receiving further diagnostic testing. Placing CTPA in the second round of a diagnostic strategy will reduce costs and radiation exposure. The low-risk patients will be filtered by a highly accurate, simple, and inexpensive strategy. In the remaining high-risk patients, including patients with high D-dimer levels or high clinical probability, CTPA can be performed to confirm the diagnosis of PE or establish an alternative diagnosis.

Thus, since the diagnostic accuracy of excluding PE using CTPA seems suboptimal and since this test is relatively expensive, CTPA should not be used as the first diagnostic test in patients with suspected PE. Instead, it should be implemented in a diagnostic strategy, which begins with clinical probability and D-dimer assessment.

Pieter W. Kamphuisen, MD
Giancarlo Agnelli, MD
University of Perugia
I-06100 Perugia, Italy

Potential Financial Conflicts of Interest: None disclosed.

References
In response to Drs. Kamphuisen and Agnelli, we did not endorse CTPA as an initial diagnostic test. On the contrary, we acknowledged that “pretest probability assessments should be used to select patients for CTPA, since recent publications support the exclusion of acute venous thrombosis and pulmonary embolism: a systematic review.” Ann Intern Med. 2004;140:589-602. [PMID: 15096330]

In response: We appreciate the comments regarding our meta-analysis and the opportunity to clarify our conclusions. As stated in the Discussion, “the role of CTPA without concomitant lower-extremity imaging is still undefined.” We agree with both Dr. Frost and Dr. Girard that the literature to date does not support definitive exclusion of venous thromboembolism with CTPA alone, and we recommend concurrent lower-extremity imaging before withholding anticoagulation in patients with suspected clots. While advancements in imaging technology may eventually improve diagnostic accuracy of CTPA and allow for a single test, in our opinion further study is required before this can be recommended. Moreover, the emergence of simultaneous computed tomographic venography might ultimately permit rapid, accurate assessment of the lower extremities in a combined, single-modality study (1–4), rendering moot the role of simultaneous computed tomographic venography might ultimately permit rapid, accurate assessment of the lower extremities in a combined, single-modality study (1–4), rendering moot the role of isolated imaging of the chest in venous thromboembolism.

We agree with both Dr. Frost and Dr. Girard that our abstract should have been written more explicitly to match our ultimate conclusions.

In response to Drs. Kamphuisen and Agnelli, we did not endorse CTPA as an initial diagnostic test. On the contrary, we acknowledged that “pretest probability assessments should be used to select patients for CTPA, since recent publications support the exclusion of PE based on a low pretest probability and negative results on D-dimer testing.”

Lisa K. Moores, MD
William L. Jackson Jr., MD
Walter Reed Army Medical Center
Washington, DC 20307

Disclaimer: The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the U.S. Department of the Army or the U.S. Department of Defense.

Potential Financial Conflicts of Interest: None disclosed.

References

CLINICAL OBSERVATION

Resolution of Severe Digital Ulceration during a Course of Bosentan Therapy

TO THE EDITOR: Background: Bosentan, the dual endothelin-receptor antagonist, has proven efficacy for treatment of secondary pulmonary arterial hypertension in patients with systemic sclerosis (1). Preliminary observations suggest a role for endothelin antagonist therapy in patients with systemic sclerosis who experience the Raynaud phenomenon and associated digital ulceration (2).

Objective: To describe a patient with systemic sclerosis and pulmonary arterial hypertension who had complete resolution of previously refractory digital ulcerations 4 months after initiation of bosentan therapy.

Case Report: A 50-year-old man with a 10-year history of progressive systemic sclerosis developed worsening dyspnea and debilitating digital ulcers. The patient had chronic dyspnea due to longstanding pulmonary fibrosis and required oxygen therapy at rest. In addition to scleroderma-associated pulmonary fibrosis, he had severe esophageal dysfunction and skin changes that restricted his thoracic cage. His previous treatment regimens included 20 cycles of pulse cyclophosphamide over approximately 2 years, methotrexate, steroids, and penicillamine. Secondary pulmonary hypertension was diagnosed as the cause of his recent dyspnea and increasing oxygen requirements, and bosentan therapy was initiated. At that time, the patient’s digital ulcers were extremely painful and had rapidly progressed over several months. Initially, his fingers were black and necrotic with bone exposed (no photograph available). After 2 weeks of treatment with bosentan, 62.5 mg twice daily, his digital ulcers show almost complete resolution after 18 weeks of therapy.

Figure. Digital ulcers after initiating treatment with bosentan.
initially healed and pain was substantially relieved (Figure 2, parts A and B). After 4 months of therapy, the ulcers had nearly resolved (Figure, parts C and D).

Discussion: The Raynaud phenomenon and digital ulcerations are frequent occurrences in patients with systemic sclerosis. Although the mechanism responsible for digital ulceration in these patients is not well defined, both vasoconstriction and vascular remodeling are important. In this regard, contributors to the remodeling seen in the pulmonary vasculature and in the peripheral digital arterioles include increased endothelin-1 levels, impaired prostacyclin release, coagulation and platelet dysfunction, inflammation and cytokine release, smooth-muscle proliferation, and extracellular matrix deposition (3). The result is occlusive vascular disease leading to tissue ischemia and ulceraltion.

In our patient, clinical improvement correlated with the onset of bosentan therapy, suggesting that endothelin blockade contributed to the healing of digital ulceration. The exact mechanism responsible for this effect in systemic sclerosis–related Raynaud phenomenon is not known. It is known that in patients with systemic sclerosis, endothelin-1 levels are elevated in the blood, dermis, and affected organs (4). Moreover, elevated levels of endothelin-1 have been identified in lung tissue of patients with primary pulmonary hypertension. Two receptors (endothelin A and endothelin B) are responsible for the downstream effects of endothelin. On vascular endothelium, endothelin B mediates vasodilation. In vascular smooth muscles, endothelin A receptors and endothelin B receptors mediate vasoconstriction and proliferation of smooth-muscle cells (3). The rapid rate of growth and repair in our patient is probably due to a combination of these effects on endothelial remodeling and vascular reactivity.

Current therapy for the Raynaud phenomenon and digital ulcers includes calcium-channel blockers, prostacyclins, and other vasodilatory agents. Our report suggests that bosentan should be added to this list of treatment options. One other case report suggests efficacy of this drug in digital ulceration related to systemic sclerosis (2), but photographs of the digital ulcers were not provided for comparison; it is unclear whether they were as severe as those seen in our patient. Finally, a recent prospective study designed specifically for patients with systemic sclerosis and digital ulcers showed a significant decrease in the number of new ulcers during the 16-week treatment period (5).

Conclusions: Bosentan has other clinical effects besides decreased pulmonary arterial pressure and increased exercise capacity in patients with systemic sclerosis who are being treated for pulmonary hypertension. This additional benefit was demonstrated in our patient by the rapid healing of digital ulcers. The effects are probably mediated through endothelin A receptors and endothelin B receptors on vascular smooth-muscle cells, resulting in vascular remodeling. In select patients with systemic sclerosis, bosentan will probably be a useful therapy for the Raynaud phenomenon and associated digital ulcers.

Michael J. Snyder, MD
Mary R. Jacobs, MD
Rafael G. Grau, MD
David S. Wilkes, MD
Kenneth S. Knox, MD
Indiana University School of Medicine
Indianapolis, IN

Potential Financial Conflicts of Interest: None disclosed.

References


The abstract of a review on methods for diagnosing intravascular device–related bloodstream infection (1) contained an error. In the third sentence of the Data Synthesis section, it should have read, “followed by IVD-drawn quantitative blood culture,” not “qualitative blood culture.”

References