Is It Time to Retire PIOPED?

Lionel S. Zuckier1 and Sean Logan Boone2

1Division of Nuclear Medicine, Department of Radiology, Montefiore Medical Center and Albert Einstein College of Medicine, Bronx, New York; and 2Department of Radiology, St. Joseph’s Hospital and Medical Center, Phoenix, Arizona

Since its publication in 1990 (1), the Prospective Investigation of Pulmonary Embolism Diagnosis (PIOPED) study has played a central role in informing algorithms used to diagnose pulmonary embolism (PE). Indeed, PIOPED-based algorithms maintain a central role in current best practices and procedure standards (2). Given that most early-career practitioners and trainees were born after the PIOPED results were released in 1990, its chronology bears retelling.

PIOPED was a National Institutes of Health–funded prospective, multiinstitutional study that analyzed the diagnostic usefulness of ventilation–perfusion lung scintigraphy in acute PE (1,3–5). Symptomatic adult subjects were enrolled and imaged by planar scintigraphy after administration of 133Xe gas and 99mTc-macroaggregated albumin. PIOPED was notable for its prospective interpretation criteria, large cohort of patients, efforts to avoid selection bias, and rigorous gold standard, including pulmonary catheter angiography, which was performed on most subjects. Though not the first to do so, PIOPED used a probabilistic model of reporting, casting the lung scan results as normal/near normal or as low, intermediate (indeterminate), or high probability for PE.

The original PIOPED investigation was flawed from the start. Because it was a prospective trial, the criteria for scintigraphic interpretation were assigned before initiation; unfortunately, these were ultimately determined to be suboptimal. This Achilles’ heel led to poor correlation between scintigraphic interpretation and interventional angiography, the standard of truth used in the trial (1). A lackluster outcome contributed to impugning of lung scintigraphy’s value in the minds of many clinicians, bringing about its near demise (6). The PIOPED investigators subsequently moved beyond their initial error by retrospective reanalysis of the study’s large data pool, giving rise to revised (7) or modified (8) PIOPED criteria, which were then prospectively tested in new patient cohorts, though generally with a weaker, composite, clinical gold standard (9). These revised criteria have been incorporated into various diagnostic protocols (2,10). After the original PIOPED study, PIOPED II and III were conducted, which were National Institutes of Health–funded trials of spiral CT angiography (11) and gadolinium-enhanced MR angiography (12) for the diagnosis of PE, which bear only tangential relevance to our current discussion.

Incredibly, accrual of patients in the PIOPED study occurred over 37 y ago; at that time the term evidence-based medicine had not yet been coined (13), Technegas (Cyclomedica) was a new product available in only limited markets (14), SPECT cameras were being initially introduced in the clinic, and SPECT/CT did not yet exist (15). In essence, the landscape of clinical nuclear medicine bore little resemblance to the current terrain. Is the venerable PIOPED too dated and dissonant to be applicable in the contemporary environment? It is telling that a similar question was raised in this journal some 15 y ago (6). We will first reflect on the contributions made by PIOPED to lung scintigraphy and consider which of these features, if any, retain currency in the modern era, over 30 y since their introduction.

Two types of validity are required for a research study to support clinical practice (16,17). Internal validity (or study quality) refers to the confidence we have that the study incorporates minimal bias, based on best research practices such as randomization and masking, leading to conclusions that are internally consistent and accurate. External validity (or generalizability) refers to whether the conclusions derived from the sample of subjects studied can be extended to other broader populations of patients. This is often achieved by recruiting subjects from multiple institutions and ensuring that they reflect a wide variety of demographic backgrounds. The PIOPED study excelled in internal validity, based on data that were robust, complete, extensive, and validated, including an exceptional gold standard. These data were harnessed to generate new and optimized revised interpretation criteria, which de facto converted the lackluster prospective trial into a powerful retrospective study. In its day, PIOPED also reflected excellent external validity, based on contemporary best imaging practices that were performed on more than 1,400 study participants across 6 different institutions. As population, equipment, radiopharmaceuticals, and techniques have changed over the ensuing 30 y of practice, the study’s external validity has been gradually eroded. Patients undergoing lung scintigraphy today are markedly different from those studied during PIOPED, with a much lower prevalence of PE. From a technical perspective, only a minority of practitioners still use 133Xe gas for ventilation, instead substituting aerosol ventilation methods (18), and this fraction may further decrease now that Technegas has been approved by the United States Food and Drug Administration and will be adopted into the market. γ-cameras have progressed from analog acquisition and display to fully digital systems, with superior resolution and larger fields of view than in the time of PIOPED. Numerous practitioners have also moved beyond planar imaging to embrace tomography (especially in Canada and Europe (18,19)), whereas many more physicians would be amenable to this change if reflected in updated guidelines. Reinartz has succinctly pointed out that in no other realm of scintigraphy do we limit ourselves to nontomographic imaging (6). The concern that tomography will lead to visualization
and overcalling of small, insignificant defects would be best allayed by updated criteria and education, not by throttling imaging data. In toto, it seems clear that changes in practice patterns have led to an insidious decline in external validity that has eclipsed any advantage gained from the original superior internal validity of the PIOPED data.

A further feature of the PIOPED interpretation schema is their Bayesian or probabilistic reporting nomenclature, although these, in fact, were introduced by other investigators predating PIOPED (20). It is a mathematical truism that calculation of posttest probability of disease must take into account the a priori probabilities (21,22). Furthermore, clinical diagnostic imaging has been moving toward—not away from—standardized reporting, use of clearly defined criteria, and probabilistic interpretation, as evidenced by the proliferation of “-RADS” systems of reporting throughout radiology (23–26). For these idealized reasons, the PIOPED criteria were prescient, incorporating medical decision making into the science of diagnostic imaging. Nonetheless, on a practical level, the Bayesian categorization of test results is judged by many as tedious, misunderstood, and impractical. Categorization of the images into 3 or 4 categories ranging from normal/near normal through high probability differs radically from binary interpretations customarily applied in much of medical imaging, including CT pulmonary angiography, which is currently the dominant radiographic method of evaluating PE. If clinicians do not comprehend the nuances of a probabilistic diagnosis, more harm than benefit may result. Has the complexity of PIOPED been shown to really improve outcomes in the field or is it in fact unhelpful and poorly understood? Previous research has shown that there is significant variability in how referring and even interpreting physicians understand the probability categories, particularly intermediate- and low-probability results (27–29).

How can we move beyond PIOPED? Can we develop new criteria, replete with both internal and external validity, that will incorporate a Bayesian framework of diagnosis but will also be manageable and understandable? Can the principles of evidence-based medicine inherent in the PIOPED design be ported to our current practice paradigms? In fact, a universal methodology to replace PIOPED has not emerged in the intervening 33 y since it was developed because of the difficulty of replicating the high-quality data, the extensive clinical experience, and the need to embed scan findings into an integrated diagnostic strategy (10). For example, the European Association of Nuclear Medicine criteria (19), although widely used in Canada and Europe, have not been universally embraced in the United States, at least in part due to concern that the acquisition technique and diagnostic criteria for reporting tomographic (SPECT) ventilation–perfusion scans are variable and have not been sufficiently validated (30,31).

It seems conceivable that artificial intelligence (AI) techniques have the potential to inherit the mantle of PIOPED. Many of the rigorous concepts that were embodied in the PIOPED approach can now be applied within AI interpretation of lung scintigraphy, including harvesting of extensive pretest, test, and validated outcome data, correlated by complex deep learning models (32–34). Many features enter into an expert’s evaluation of lung scintigraphy, often exceeding the performance of published diagnostic algorithms (35). The improved performance of expert evaluation has been attributed to the use of intangible and unique Gestalt facets (36,37), versus additional personal, though not codified, rules of interpretation (38). This is clearly the province of AI. Lung scintigraphy was in fact one of the earliest medical imaging applications of AI (39–42), with a flurry of activity in the 1990s and early 2000s (43–45), though as CT pulmonary angiography became the dominant clinical diagnostic modality in PE, it also became the primary focus of AI research (46). The senescence of PIOPED should be countered by development of powerful techniques of AI interpretation. In that manner, we can enhance the role of scintigraphy in patients with suspected PE while simultaneously improving diagnostic outcomes.

DISCLOSURE
No potential conflict of interest relevant to this article was reported.

REFERENCES
