EDITORIAL

CrossMark

A rose by any other name

Timothy F. Christian, MD, MPA[a]

[a] Division of Cardiovascular Disease and Hypertension, Robert Wood Johnson Medical School, Rutgers University, New Brunswick, NJ

See related article, pp.1480–1488

A rose by any other name would smell as sweet
William Shakespeare, Romeo and Juliet[1]

Equivalence of tests is an elusive proof. We pursue it in order to gain assurance that we can substitute one test for another. The reasons to do this are multiple but revolve around cost containment or ease of access. In this month's Journal article by de Haan et al,[2] an example of this type of substitution is presented, comparing the modalities of positron emission tomography (PET) and magnetic resonance imaging (MRI) for the identification of the penumbra of denervated viable myocardial tissue following myocardial infarction. Such regions have been postulated to be fertile ground for ventricular arrhythmia development due to extended refractory periods found within during EP testing.[3] These extend beyond the conventional boundaries of risk area and infarct size. The hope is that our precision in identifying those patients who benefit from implantable defibrillators (ICD) will improve if a reliable noninvasive test, aimed at quantification of this unstable region, can be introduced into the selection algorithm.

The stakes for correct ICD therapy are high. Consequently, sensitivity must be maximized with virtually no false negatives acceptable. Because of this, the present entry criteria are quite unselective. While there are historical variables that are fairly strong in identifying patients in greatest need of such therapy,[4] the default criterion has been implantation in all patients with a

reduced LVEF following MI.[5] Many ICDs are placed in patients who will not require them in order to catch the small pool for those where they will be life-saving. From a Bayesian stand-point, what is required is a test that will allow us to more selectively identify those with LV dysfunction post-MI who will benefit from aggressive therapy without losing any of those already captured in our low LVEF net (see Figure 1).

C-11 hydroxy ephedrine PET imaging is an attractive tool because it gets at the heart of the proposed mechanism of arrhythmia. However, due to the short half-life of C-11, an onsite cyclotron is required, which puts it out of reach as a routine test. Consequently, a more readily available substitute is attractive. Cardiac MRI is a much more widely available modality and there is evidence that identification of a ''grey zone'' on gadolinium-enhanced inversion recovery sequences correlates with ventricular arrhythmias in patients post-MI.[6] This sets the stage for potentially substituting MRI as a surrogate for the actual physiology described in the PET scan.

C-11 HED is a norepinephrine analog and so is attractive in getting at the heart of the neural-based hypothesis of ventricular re-entry tachycardia. Gadolinium-enhanced MRI is a nonspecific measure of extracellular volume. Gadolinium is a paramagnetic agent that accumulates in the extracellular space and changes the relaxation properties of the hydrogen protons that surround it. Infarcted tissue has more extracellular space than viable myocardium and consequently will have different T1 properties compared to viable myocardium after a Gadolinium-based contrast agent is given. ''Grey zone'' tissue is hypothesized to constitute islands of viable myocardium within the perimeters of the infarct zone. As infarcted tissue is conventionally imaged to produce white infarcts against black viable myocardium, gray tissue is an average of pixel values producing intermediate signal in this penumbra zone. But islands of viability are but one source of intermediate gray signal. Partial volume effects can be important. Since cardiac MRI generally

Reprint requests: Timothy F. Christian, MD, MPA, Division of Cardiovascular Disease and Hypertension, Robert Wood Johnson Medical School, Rutgers University, New Brunswick, NJ 08901; *timothy.christian@rutgers.edu*

LVEF Distribution for Patients Post MI using 35% Threshold for ICD Placement

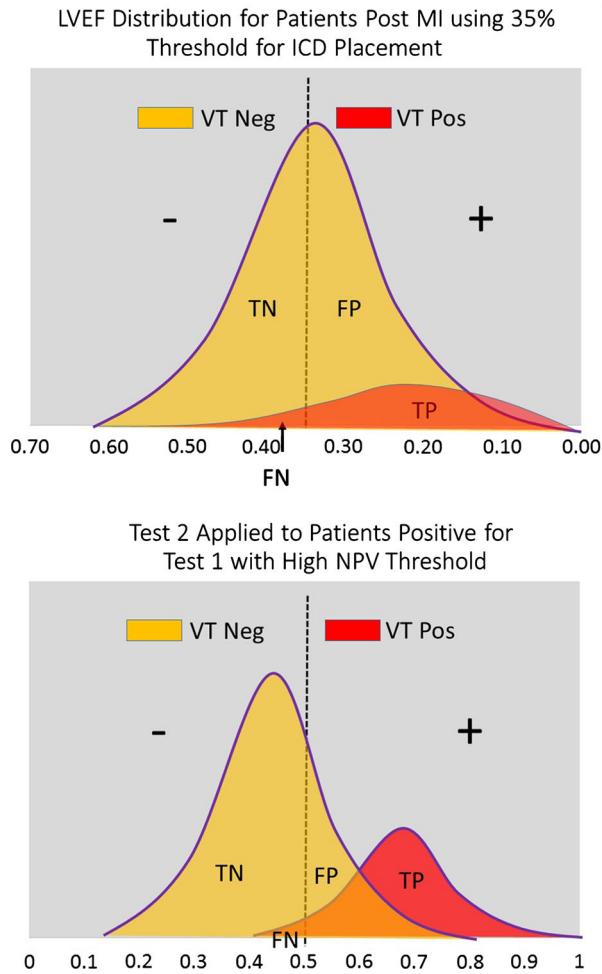Test 2 Applied to Patients Positive for Test 1 with High NPV Threshold

Figure 1. A hypothetical analysis of an ICD placement strategy using step-wise tests by to determine the probability of developing ventricular tachycardia (VT). *Panel 1* shows the distribution of LVEF values in all patients post-MI for those who will develop VT (*red*) and those that will not (*orange*). The threshold to determine ICD placement is set at 0.35 which produces a high negative predictive value but poor specificity as many false positives are generated using this threshold. However, approximately half the patients are eliminated from ICD placement. *Panel B* shows the distribution for the remaining patients with an LVEF <0.35 for those that will develop VT and those that will not. This group has a higher probability of VT. The threshold chosen has reasonable sensitivity and a very high negative predictive value. This degree of performance needs to be achieved by any noninvasive test in order to change the current ICD placement algorithm.

Figure 2. Potential mechanism to generate a peri-infarct grey zone based solely on partial volume effects. When an infarct is not wholly contained within a tomographic imaging slice, the resultant image will be an average of the infarct zone plus the normal tissue adjacent to it. This will give an intermediate signal value that could be misinterpreted as a denervated grey zone with arrhythmic potential.

requires a fairly thick myocardial slice to generate enough signal for adequate quality and since infarcts come in many shapes, partial volume effects related to the morphology of the infarct can be a factor. The signal for an entire 3-D voxel is averaged to produce a 2D value. Figure 2 is a simplistic example of how two infarcts of the same size but different positions w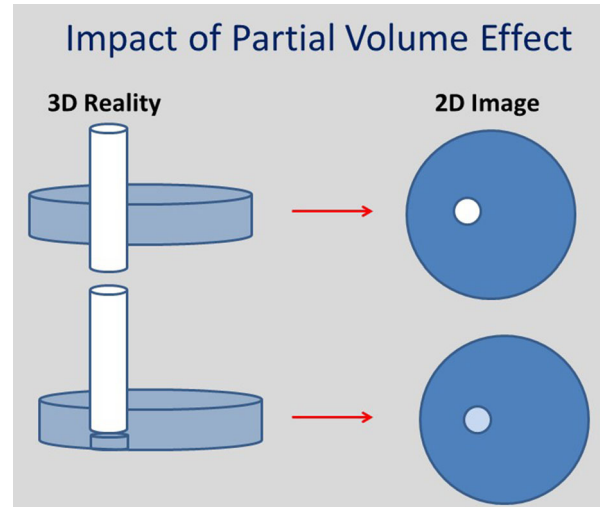ithin the tomographic slice can produce variable signal intensities. In the present study, slice thickness was set at 5 mm, which is reasonable but can still generate partial volume effects. Furthermore, the authors used a 5-mm gap between slices which saves time, but adds to the variability. The PET scan, on the other hand, appears to have thinner slices (though this is not entirely clear in the paper) and would be less prone to partial volume effects.

Equivalence of tests becomes of interest when one test is felt to reflect a physiologic process reliably but may be difficult to apply and a second test has the advantage of being less expensive, or more widely available or more easily performed, with the assumption that the information is of nearly equal accuracy. Consequently, the cheaper, easier test can be brought to scale in a manner not possible for the original test. But equivalence is in the eye of the beholder. We could define it as providing mean values that lie within each others' 95% confidence intervals (CI). It could be defined as a significant linear correlation between the two measures. In both instances, the number of measure samples is a critical component. Small sample sizes lead to wide confidence intervals opening the door for equivalence when none exists (type 2 error—a full discussion of the statistical approach to establishing test equivalence can be found here[7]). While no statistical proof of equivalence is possible, we can set statistical parameters of agreement that improve the probability that they are.

The authors are careful not to conclude equivalence in the present study between C-11 HED and contrast MRI but use the more ambiguous term of ''related.'' Indeed, the mean penumbra size was actually statistically different between techniques and the correlation was moderate. This agreement is based on denervated zone size rather than any physiologic parameter but it is not clear that size matters. It may be that a threshold quantity of gray zone is enough to increase risk. More problematic for our goal of better selecting patients for ICD therapy is the broad confidence interval in the correlation. The mean gray zone size of 6% of the LV produces a C-11 HED mismatch defect of anywhere from 7 to nearly 30% LV. It would be hard to rely on this information when there is no room for false negatives if size is important. Neither the PET nor the MRI results were associated with the induction of VT on electrophysiologic (EP) testing. This should not be of significant concern given the very small numbers of patients studied and the fact that it is not a direct endpoint of interest. We want to predict who will develop VT post-MI.

For all these types of studies, association with events and prediction are often confused. We should not mix associations of test results with outcomes and the ability to predict such outcomes—they are not equivalent. The association with an imaging characteristic for an event that has already happened is very different from using the imaging variable to predict the event. For example, showing that gray zone scar of a certain size is significantly more prevalent in a cohort of patients undergoing an EP study for inducible VT than without is not the same as predicting which patients with a low LVEF in the general population will have inducible VT prospectively based on the gray zone variable. This gets at the concept of conditional probability. The probability of B given A is not the same as the probability of A given B. A nonmedical example may be useful here. Equating conditional probabilities was made in the OJ Simpson trial by the defending attorney.[8] He equated A given B with B given A in the following manner: He stated that the probability of a wife being murdered by an abusive husband was 1:2500, or given the husband is abusive, what is the probability she will be murdered. That purported fact was presented to the jury. But the actual question was, given that a wife is murdered, what is the probability it was by an abusive husband? That condition has a probability of 8:9. Consequently, the probability of a peri-infarct gray zone being present in a patient with VT is not the same as the probability that patients with a significant gray zone develop VT. The authors make no such claim but it is an easy trap to fall into.

What should we make of the present study as it impacts the ICD implantation algorithm? There is a correlation between the PET physiology and the MRI anatomy. But in order for this to be useful three conditions must be met. First, the reproducibility over time of MRI gray zone quantitation (we have not discussed the potential dynamic nature of this zone) has to be established. Secondly, the negative predictive value to rule out future events in future trials has to be very high given the stakes. Since the default option has been to place an ICD in most post-MI patients with LV dysfunction, we are looking for opportunities to safely withhold that therapy in some patients to better target the population that will benefit. Consequently, we are not looking to noninvasively predict who will develop VT, we need to predict confidently who will not. Finally, viability imaging by MRI, if shown to meet the first two criteria, has to be taken to scale. This is a big population of patients.[9] At present, cardiac MRI is too expensive to serve this role. But payment reform through the use of a limited MR scan is a viable option. Since the majority of these patients have an assessment of LV function already, a simple injection of Gadolinium during patient set-up followed by viability imaging only could be performed within 20 minutes at low cost.

We should not be deterred in searching for equivalence. The present study chose a technique that describes the assumed mechanism using the best methods available as a gold standard and compared those results to a test that, though not cheap, is more widely available to the medical community. At present, it is unclear whether either test is accurate enough to rule out ICD placement in patients with a low LVEF post-MI. This needs to be established. But there remain opportunities in cardiac imaging where the stakes are less and an equivalent test that can be taken to scale based on economy, availability, and/or simplicity of use is an optimal solution. In those instances, we can set boundaries for what defines equivalence, taking note of the substituting test's weaknesses and biases. In this manner, that test may smell as sweet as the original rose.

Disclosure

T christian has nothing to disclose.

References

1. Shakespeare W. Romeo and Juliet, Act II, scene ll.
2. de Haan S, Rijnierse MT, Harms HJ, Verberne HJ, Lammertsma A, Huisman MC, et al. Myocardial denervation coincides with scar heterogeneity in ischemic cardiomyopathy: A PET and CMR study. J Nucl Cardiol 2015. doi:10.1007/s12350-015-0316-z.

3. Calkins H, Allman K, Bolling S, Kirsch M, Wieland D, Morady F, et al. Correlation between scintigraphic evidence of regional sympathetic neuronal dysfunction and ventricular refractoriness in the human heart. Circulation 1993;88:172-9.

4. Buxton AE, Lee KL, Hafley GE, Pires LA, Fisher JD, Gold MR, et al. Limitations of ejection fraction for prediction of sudden death risk in patients with coronary artery disease: Lessons from the MUSTT study. J Am Coll Cardiol 2007;50:1150-7.

5. Moss AJ, Zareba W, Hall WJ, Klein H, Wilber DJ, Cannom DS, et al. Prophylactic implantation of a defibrillator in patients with myocardial infarction and reduced ejection fraction. N Engl J Med 2002;346:877-83.

6. Schmidt A, Azevedo CF, Cheng A, Gupta SN, Bluemke DA, Foo TK, et al. Infarct tissue heterogeneity by magnetic resonance imaging identifies enhanced cardiac arrhythmia susceptibility in patients with left ventricular dysfunction. Circulation 2007;115:2006-14.

7. Obuchowski NA. Testing for equivalence of diagnostic tests. Am J Radiol 1997;168:13-7.

8. Strogatz S. Chances are. NY Times, April 25, 2010.

9. Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Borden WB, et al. Executive summary: heart disease and stroke statistics—2013 update: A report from the American Heart Association. Circulation 2013;127:143-52.