

## Comparing ROC Curves Derived from Nested Models

Mithat Gönen\*, Memorial Sloan-Kettering Cancer Center  
Venkatraman Seshan, Memorial Sloan-Kettering Cancer Center  
Colin Begg, Memorial Sloan-Kettering Cancer Center

### Abstract

In constructing predictive models, investigators frequently assess the value of an incremental predictive marker by comparing the ROC curve generated from the predictive model including the new marker with the ROC curve from the model excluding the new marker. Many commentators have noticed empirically that a test of the two ROC areas often produces a non-significant result when a corresponding Wald test from the underlying regression model is significant. A recent article showed using simulations that the widely-used ROC area test (DeLong et al, 1988) produces exceptionally conservative test size and extremely low power (Vickers et al, 2011). In this article we show why the DeLong et al. test is invalid in this context. We demonstrate how a valid test of the ROC areas can be constructed that has comparable statistical properties to the Wald test.

*Keywords:* Receiver operating characteristic curve; Biomarker; Area under the ROC curve; Predictive accuracy; Discrimination.

---

\* Presenting author