A meta-analysis of the accuracy of a neuroendocrine tumor mRNA genomic biomarker (NETest) in blood

K. Öberg[1*], A. Califano[2], J. R. Strosberg[3], S. Ma[4], U. Pape[5], L. Bodei[6], G. Kaltsas[7], C. Toumpanakis[8], J. R. Goldenring[9], A. Frilling[10] & S. Paulson[11]

[1]Department of Endocrine Oncology, University Hospital, Uppsala, Sweden; [2]Department of Systems Biology, Columbia University, New York; [3]Department of Gastrointestinal Oncology, Moffitt Cancer Center, Tampa; [4]Department of Biostatistics, Yale University, New Haven, USA; [5]Department of Internal Medicine and Gastroenterology, Asklepios Kliniken Hamburg, Germany; [6]Department of Radiology, Memorial Sloan Kettering Cancer Center, New York, USA; [7]Division of Endocrinology, Department of Pathophysiology, University of Athens, Athens, Greece; [8]Department of Gastroenterology, University College of London, London, UK; [9]Section of Surgical Sciences, Vanderbilt University Medical Center, Nashville, USA; [10]Department of Endocrine Surgery, Imperial College London, London, UK; [11]Division of Medical Oncology, Baylor Charles A Sammons Cancer Center, Dallas, USA

Background: The lack of an accurate blood biomarker in neuroendocrine tumor (NET) disease has hindered management. The advance of genomic medicine and the development of molecular biomarkers has provided a strategy—liquid biopsy—to facilitate real-time management. We reviewed the role of a blood mRNA-based NET biomarker, the NETest, as an *in vitro* diagnostic (IVD).

Patients and methods: A systematic review of the literature using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines was undertaken. The methodological quality was evaluated using the QUADAS-2 tool. We identified ten original scientific papers that met the inclusion criteria. These were assessed by qualitative analysis and thereafter meta-analysis. Data were pooled and a median [95% confidence interval (CI)] diagnostic odds ratio (DOR), positive likelihood ratio (+LR), and negative likelihood ratio (−LR) were calculated. For the meta-analysis, a generic inverse variance method was undertaken using the accuracy and area under the curve (AUC) data.

Results: The ten studies exhibited moderate to high methodological quality. They evaluated NETest usage both as a diagnostic and as a monitoring tool. The meta-analysis identified the diagnostic accuracy of the NETest to be 95%−96% with a mean DOR of 5 853, +LR of 195, and −LR of 0.06. The NETest was 84.5%−85.5% accurate in differentiating stable disease from progressive disease. As a marker of natural history, the accuracy was 91.5%−97.8%. As an interventional/response biomarker, the accuracy was 93.7%−97.4%. The pooled AUC for the NETest was 0.954 ± 0.005, with a z-statistic of 175.06 ($P < 0.001$).

Conclusions: The NETest is an accurate biomarker suitable for clinical use in NET disease management. The meta-analysis supports the utility of the NETest as an IVD to establish a diagnosis and monitor therapeutic efficacy. The use of this as a biomarker provides information relevant to NET management consistent with observations regarding utility of liquid biopsies in other oncological disciplines.

Key words: NETest, IVD, meta-analysis, biomarker, liquid biopsy

INTRODUCTION

A major factor in facilitating the management of disease is the ability to understand, on a real-time basis, its natural course and the effect of therapy upon it.[1] The three key methods for assessing disease status are clinical assessment, imaging, and biomarkers.[1−3] A clinical review has significant accuracy limitations and is often only effective once changes are substantial. Imaging is complex, requires sophisticated technology, and often fails to detect early changes in disease status. Furthermore, imaging involves repetitive exposure to radiation and can be costly. Both clinical and imaging strategies require subjective interpretation, are operator-dependent, and have high intra-observer variability.[4] Blood biomarkers are simpler to evaluate since measurements are objective and information can be acquired on a real-time basis by the relatively innocuous process of venipuncture.[1−3] The latter is substantially different from tissue biomarker assessment, which represents a one-time invasive random evaluation of disease status in tumors that are often heterogeneous.[5,6] In the evolving landscape of scientific advances that offer

Correspondence to: Prof. Kjell Öberg, Department of Endocrine Oncology, University Hospital, S-75185 Uppsala, Sweden. Tel: +46 18 6114917
E-mail: kjell.oberg@medsci.uu.se (K. Öberg).

powerful tools to facilitate disease management, the development of molecular biomarkers and the application of genomic medicine, for example multianalyte blood biomarkers to neuroendocrine tumor (NET) disease, will represent a major advance likely to substantially improve its clinical management.[5,7]

In general, biomarkers encompass all tools and technologies that facilitate the prediction and diagnosis of disease and assess progression, regression, and treatment outcome. Information can be derived from but not limited to tissue, body fluids, and imaging.[8] While most biomarkers represent the cell surface or secreted proteins, discovery pipelines now focus on molecular (genetic/RNA) targets that provide multilevel data that define cell behavior.[9]

The US National Cancer Institute convened a NET summit meeting in 2007 to prioritize key research areas and clinical issues necessary to advance the field.[10] Of the ten recommendations, one was to develop tumor and plasma markers to facilitate early diagnosis and augment information necessary to monitor disease management.[10] The extended delay, typically several years, associated with a NET diagnosis[11] limits treatment effectiveness since tumor burden is often advanced. In addition, the absence of sensitive and specific biomarkers results in the inaccurate monitoring of disease progression and the imprecise evaluation of treatment responses.[12] This principally reflects the overreliance on monoanalyte biomarkers that define cell/tumor secretory activity and provide minimal information about the metabolic, proliferative, and metastatic components that are critical regulators of tumor evolution.

The National Institute of Health (NIH) has classified biomarkers into three broad categories based on validation and clinical usage.[13] This comprises type 0 markers, 'indicators of the natural history of disease', which correlate with diagnosis, prognosis, and outcome; their relationship with disease can be indirect. Type I 'capture the effects of an intervention in accordance with the mechanism of action of the drug' and reflects the general efficacy of treatment through linkage to the specific mechanism being exploited. Type II markers are used as surrogates for clinical endpoints, the term reflecting patient health, functionality, or survival.

A blood-based multianalyte gene transcript (mRNA) biomarker developed for NETs is the *in vitro* diagnostic (IVD) NETest (supplementary material and supplementary Figure S1, available at *Annals of Oncology* online).[14,15] Per the Food and Drug Administration (FDA) definition, an IVD is any 'reagent, instrument, and/or system intended for use in diagnosis of disease or other conditions, including a determination of the state of health, in order to cure, mitigate, treat, or prevent disease or its sequelae.'[16] The multigene assay was designed specifically for NET disease and proposed as a multianalyte algorithmic biomarker tool for gastroenteropancreatic (GEP) and bronchopulmonary NET management. Its principal clinical utility is to monitor disease progress and assess prognosis.[7] It therefore functions both as type 0 and type II biomarkers. Early studies provided a score between 0 and 8[14,17,18] with a prediction of whether a sample was a NET or not. Thereafter, the machine-learned output was amplified by inclusion of information from specific gene clusters or 'omes' proposed as hallmarks of NET cancer.[15,19] This quotient provides a score between 0 and 100 that is segregated into broad clinically relevant categories of normal, low, and high risk.[15] The category risk assignment was derived empirically from clinical databases of patients with variable progression status.[7,15]

The genomic validity of this transcriptome-based algorithm-analyzed assay has been independently confirmed in a multi-center NIH-funded study.[20] The gene expression profiles of 10 244 tumors encompassing 32 different types of neoplasia, including NETs, and normal samples from The Cancer Genome Atlas database were evaluated.[20] The results identified that the NETest gene signature accurately captured the NET genotype.[20] The same study demonstrated the gene expression profile was not detectable in hematologic-derived cells, particularly 'white blood cells', which constitute the predominant source of blood-derived mRNA. Any detectable circulating levels therefore must be derived from the tumor.

Public domain assessment of the assay and its use as an IVD is supported by a substantial body of literature.[5,14,15,17,18] Despite positive academic assessment and numerous peer-reviewed publications, adoption of the assay by the neuroendocrine oncological community has been slow. Although numerous published studies demonstrate superiority over monoanalyte biomarkers [Chromogranin A (CgA)], the NETest has not yet been included in society guidelines, which remains controversial given the clear scientific evidence of its superiority. Obviously the NET field requires the best available tools to optimize the management of patients.

In 2014, a state-of-the-art Delphic conference with a focus on the status of NET biomarkers identified that 'circulating multianalyte biomarkers provide the highest sensitivity and specificity necessary for minimum disease detection and that this type of biomarker had sufficient information to predict treatment effectiveness and prognosis', and concluded that 'trials measuring multianalytes (e.g. neuroendocrine gene transcripts) should also identify how such information can optimize the management of patients with neuroendocrine tumours'.[1] We reviewed the literature in the 5 years since this meeting to investigate whether these initial considerations about the NETest have been borne out by clinical usage. The overall aim was to determine if current biomarkers should be replaced with the NETest to advance the management of NET patients.

METHODS

A systematic review of the literature based on the recommendations of systematic reviews and meta-analyses outlined by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)[21] and by the Cochrane Diagnostic Test Accuracy Working Group[22] was undertaken (May 2019) using the term 'NETest'. No other

terms were utilized as we focused only on assays thus named. Ovid, Medline (PubMed.gov), and Embase were evaluated. The search was undertaken without any filter. No language restriction was applied. The reference lists of the included manuscripts were also reviewed to ensure all the relevant studies were enrolled.

Twenty-three studies, of which 10 had appropriate NETest data for evaluation, were identified (Figure 1, Table 1). Samples included both GEP- and bronchopulmonary NETs and sample sizes of the individual studies ranged from 19 to 253 patients (supplementary Table S1, available at *Annals of Oncology* online). All studies included sufficient data to determine accuracy, sensitivity, and specificity and absolute numbers (true positive, false positive, true negative, and false negative) were reported or derivable from the published material. The methodological quality of

each of the studies was evaluated using the QUADAS-2 tool[23]; all 10 were identified to exhibit moderate to high methodological quality. Some studies evaluated multiple uses of the NETest, for example, as a diagnostic and monitoring tool. These were evaluated separately (see below).

In order to provide an unbiased review, a meta-analysis based upon Cochrane methodology was undertaken following the construction of 2×2 tables.[24] The diagnostic accuracy, sensitivity, and specificity were calculated and standard errors were derived from 95% confidence intervals (supplementary material and Table S2, available at *Annals of Oncology* online). Diagnostic odds ratios (DOR), the positive likelihood ratio (+LR), and the negative likelihood ratio (−LR) were calculated. Data from the different studies were pooled and mean [95% confidence interval (CI); median] DOR, +LR, and −LR are presented. The area under the curves (AUCs) were also calculated for each of the different studies.

For the meta-analysis, a generic inverse variance method was undertaken using the accuracy, sensitivity, specificity, and AUC data. We evaluated both random and fixed models and presented both. Cochran's *Q*-test and the Higgins' *I*-squared test were calculated for each of the two models (both are presented) (supplementary material and Table S3, available at *Annals of Oncology* online).[25] MedCalc statistical software version 16.2.1 (MedCalc Software bvba, Ostend, Belgium; http://www.medcalc.org; 2017) and meta-a4diag R package[26] were utilized.

Four areas of clinical utility were examined:

(i) Diagnostic: this included six studies[27–32]; the cut-off for diagnosis was 20.
(ii) Disease status: to define disease status as either stable or progressive at the time of the blood draw, six studies included NETest data that could be used to differentiate these two disease states.[27,28,31–34]
(iii) Natural history marker of NET disease: this corresponds to the NIH classification of a type 0 biomarker as 'indicators of the natural history of disease'.[13] Two are published[33,34]; we specifically evaluated the clinical value of two cut-offs.
- First, whether values ≤40 represented stable disease per category output.
- Second, whether an increase in NETest to >40 was indicative of progressive disease.
(iv) The utility of the NETest as a response or interventional biomarker to therapies is equivalent to the NIH classification of a type II biomarker. These are used as surrogates for clinical endpoints and typically reflect progression free survival (PFS).[13] Four studies were identified[31,33,35,36]; three studies examined the efficacy of somatostatin analogs and one, peptide receptor radionuclide therapy (PRRT). We examined
- First, whether a decrease in the NETest to ≤40 was associated with disease stabilization;
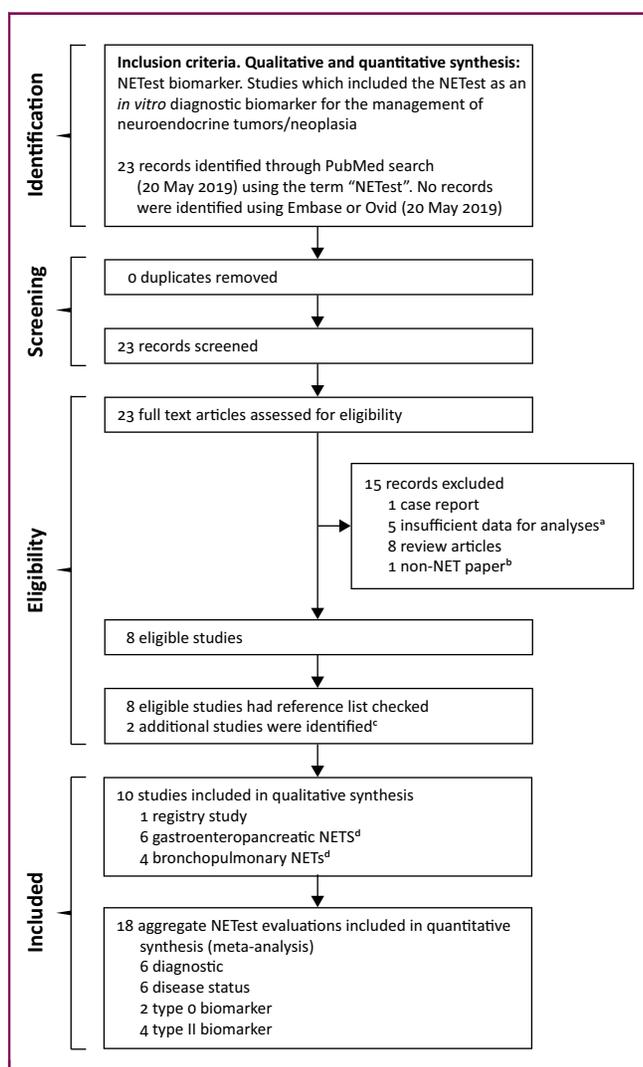- Second, whether scores of 41–100 were associated with treatment 'failure'.

Figure 1. PRISMA flow diagram.
[a] This included two surgical studies[63,64] and three studies that did not use the NETest 0–100 score.[17,47,52]
[b] This study focused on paragangliomas and pheochromocytomas.
[c] One study focused on the NETest in bronchopulmonary NETs, the second included an abstract[30] presented as ASCO-GI, January 2019.
[d] One study[36] included patients with both gastroenteropancreatic and bronchopulmonary NETs.

Table 1. List of NETest studies included in the meta-analysis

No.	Author (reference)	Sample size	NET types	Study type	Study category	Summary results	Level of evidence[60]
1	Kidd et al. 2017[27]	50	BP	Retrospective	Diagnostic research	Blood measurements accurately diagnosed bronchopulmonary carcinoids, distinguishing stable from progressive disease. This marker panel will have clinical utility as a diagnostic liquid biopsy able to define disease activity and progression in real-time.	IIB
2	Filosso et al. 2018[28]	208	BP	Retrospective	Diagnostic validation	Blood NET gene levels accurately identified BPNETs (100%) and differentiated these from controls and benign and malignant lung disease. Progressive disease could be identified and surgical resection verified. Monitoring NET transcript levels in blood will facilitate management by detecting residual tumor and identifying progressive disease.	IIA
3	Van Treijen et al. 2019[29]	253	GEP	Case control study	Diagnostic validation	The superior sensitivity of the NETest irrespective of the stage of the disease emphasizes its potential as a marker of disease presence in follow up as well as an indicator for residual disease after surgery.	IIA
4	Al Toubah et al. 2019[30]	75	GEP	Case control study	Diagnostic validation	The sensitivity of the NETest is exceptionally high in a population of metastatic well-differentiated NETs. Specificity within a healthy population of patients is exceptionally high when using a normal range of 0% to 20%.	IIA
5	Malczewska et al. 2019a[31]	201	BP	Case control study	Diagnostic validation	Elevated NETest levels are indicative of lung neuroendocrine neoplasia. NETest levels correlate with tumor tissue and imaging and accurately define clinical progression.	IIA
6	Malczewska et al. 2019b[32]	184	GEP	Case control study	Diagnostic validation	NETest is an effective diagnostic for PNETs and SINETs. Elevated NETest is as effective as imaging in diagnosis and accurately identifies progression.	IIA
7	Liu et al. 2018[33]	100	GEP, BP & CUP	Prospective observational cohort study	Longitudinal study	Blood NETest is an accurate diagnostic and can be of use in monitoring disease status and facilitating management change in both watch-and-wait and treatment cohorts.	IIA
8	Pavel et al. 2017[34]	34	GEP	Cross-sectional study	Longitudinal study	The NETest correlated with a well-differentiated GEP-NET clinical status. The NETest has predictive and prognostic utility for GEP-NETs identifying clinically actionable alterations ∼1 year before image-based evidence of progression.	IIA
9	Cwikla et al. 2015[35]	28	GEP	Controlled clinical study	Treatment research	NETest values (80% to 100%) were more accurate and occurred at a significantly earlier time point than CgA and predicted SSA treatment response.	IB
10	Bodei et al. 2016[36]	54	GEP & BP	Controlled clinical study	Treatment research	Blood NET transcript levels accurately predicted PRRT efficacy.	IB

BP, bronchopulmonary; CgA, chromogranin A; CUP, carcinoid of unknown primary; GEP, gastroenteropancreatic; IB, randomized prospective controlled trials or prospective-retrospective trials; IIA, prospective, observational studies; IIB, retrospective data modeling; NET, neuroendocrine tumor; PNET, pancreatic NET; PRRT, peptide receptor radio-therapy; SINET, small intestinal NET.

A total of 18 separate NETest evaluations (diagnostic $n =$ 6; disease status $n =$ 6; type 0 $n =$ 2; and type II $n =$ 4) were available for analysis. Inclusion criteria are included in Figure 1.

Finally, to check for publication bias, we undertook a Deeks' funnel plot asymmetry test, which is an effective method for detecting and excluding bias.[37] The Deeks' test evaluates the regression of the lnDOR against $1/\sqrt{}$(effective sample size). A P value <0.05 indicates significant asymmetry (or study bias).[38]

RESULTS

The results of the meta-analysis are included in Figures 2—4 and the diagnostic test utility is included in Table 2.

Diagnostic

As a diagnostic, the accuracy of the NETest was 95% (random) to 96% (fixed) (Figure 2A). The sensitivity was 94.4% and 89.4% and the specificity was 95.4% and 98.7%. The Q-test ranged between 10.8 and 41.5 ($P < 0.0001$ to $P = 0.054$) and I^2 was 54%—88%, which is variably suggestive of heterogeneity. The DOR was 5583 (167—25 580; median 5319, Table 2, Figure 3). The +LR and −LR were 195 (12—986; median 335) and 0.06 (0.02—0.1; median 0.06), respectively. These data are consistent with a biomarker demonstrating excellent diagnostic utility. The NETest had value both as a rule-in (a positive score that is >20 makes a significant contribution to a diagnosis by identifying those with a NET) and as a rule-out biomarker (subject has a
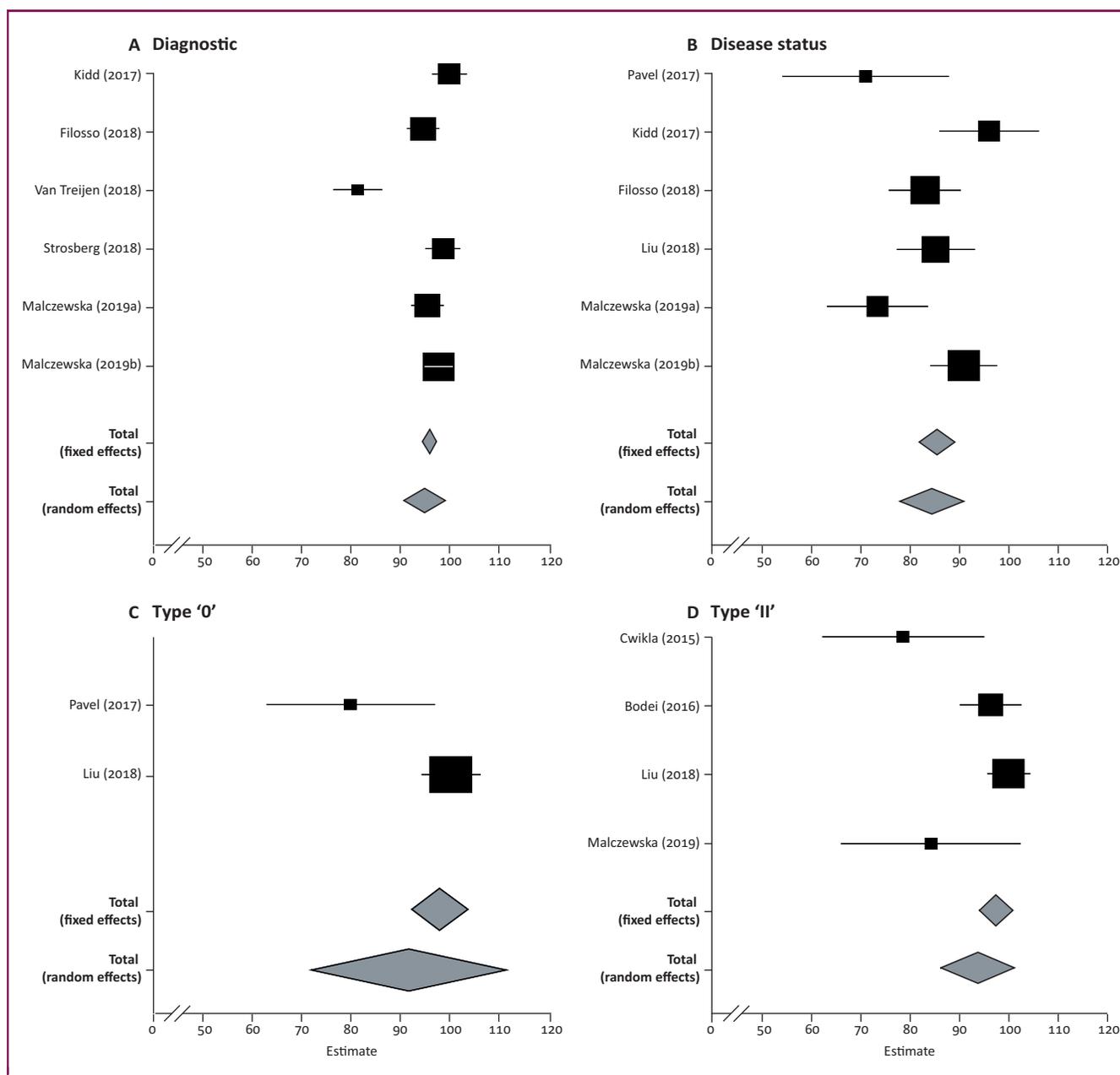
Figure 2. The clinical utility of the NETest using Forest plot analysis.
NETest utility as a diagnostic (A); the estimated accuracy and 95% confidence intervals (black boxes/lines). NETest utility for differentiating stable and progressive disease ('Disease Status') (B), NETest utility as a type '0' biomarker (C), NETest utility as a type 'II' biomarker (D).
The total effects (fixed and random) are represented by grey diamonds. Individual studies are designated by author name. The calculated accuracies range from 85% to 97% (all, $P < 0.001$).

significantly low probability of having the disease if the score falls within the normal range).

Disease status

The NETest was 84.5% and 85.6% accurate in differentiating stable disease from progressive disease (Figure 2B). The sensitivity was 79.6% and 83.1% and the specificity was 86.3% and 89.4%. The Q was 15—15.2 ($P = 0.0097$ to $P = 0.011$) and I^2 66.5%—67% (moderate heterogeneity). The DOR was 46.3 (3—216; median 30, Figure 3). The $+LR$ was 9.4 (1.6—35.2; median 7.1) and $-LR$ was 0.27 (0.12—0.54; median 0.28) (Table 2). A low NETest (≤ 40) or a

NETest that ranges from 41 to 100 provides significant information for determining clinical status and is consistent with image-based categorizations. A score ≤ 40 can be used to rule in 'stable disease' while a score >40 rules out stable disease.

Type 0 biomarker

As a marker of natural history, NIH category '0' biomarker,[13] the accuracy was 90.2% and 93.6% (Figure 2C). The sensitivity was 96.9% and 99.6% and specificity was 84.2% and 99.4%. Q was 3.6—18.4 ($P < 0.001$ to $P = 0.07$) and I^2 is 69.3%—94.6%. The DOR was 5520 (5—22 318; median
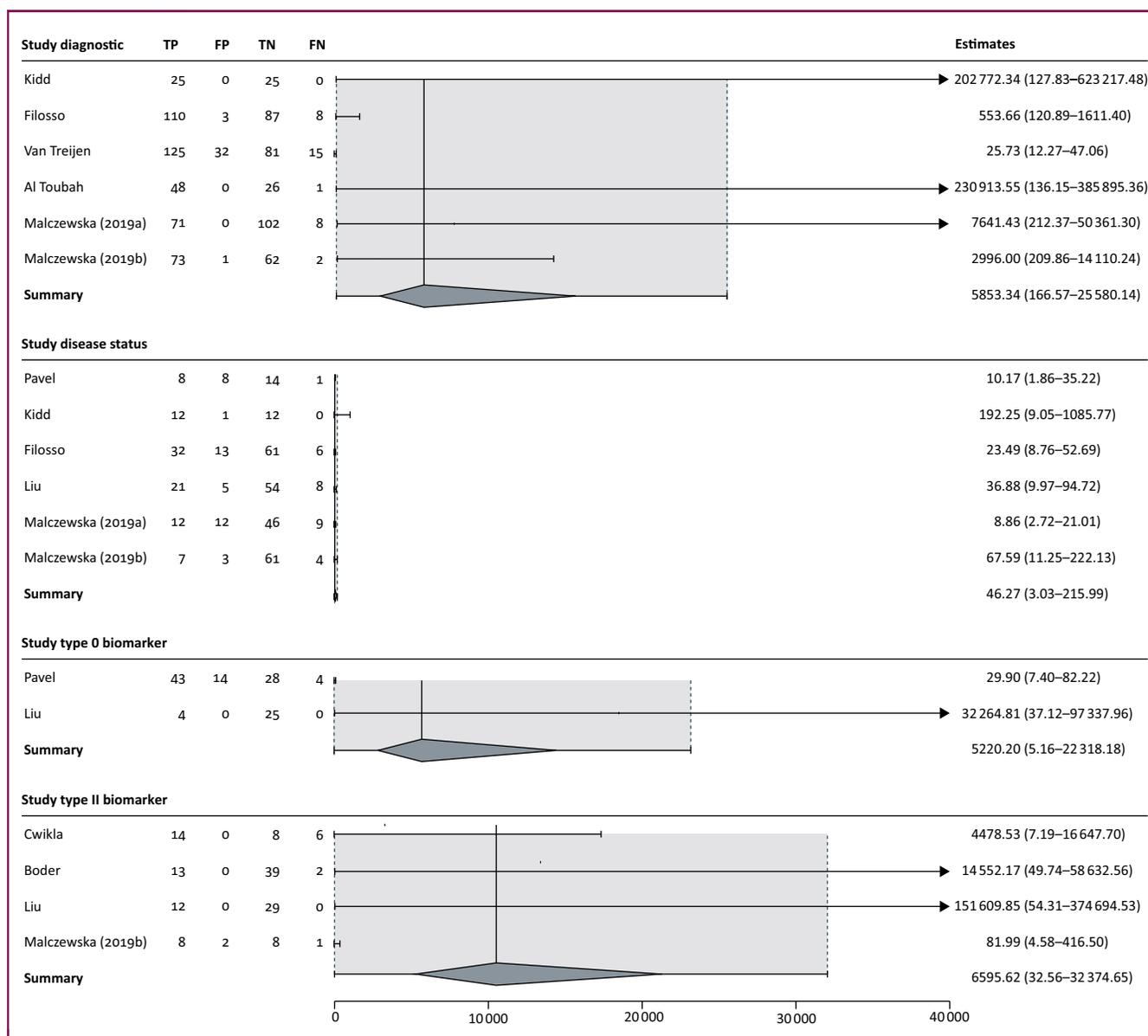
Figure 3. Forest plot of the diagnostic odds ratio (DOR) for each of the four groups of studies.
Individual studies are included as well as numbers for true positive (TP), false positive (FP), true negative (TN), and false negative (FN). The summary (dark grey diamonds) and ranges (light grey including the upper and lower estimates). The mean estimated DORs were 5853 (diagnostic), 46.3 (disease status), 5220 (type 0), and 6596 (type II).

16 147, Figure 3), the +LR was 127 (1.4—720, median 163) and the −LR was 0.10 (0.01—0.35, median 0.1, Table 2). A low NETest (≤40) or a NETest with a range of 41—100 correlates significantly with the natural history of the disease. A score ≤40 provides information that there is no evidence of progressive disease while a score >40 strongly rules out stable disease.

Type II biomarker

As an interventional or response biomarker (NIH type II), the accuracy was 93.7% and 97.4% (Figure 2D). The sensitivity was 88% and 90.1% and specificity was 99.6% and 99.7%. The Q was 2.1—8.7 (P = 0.034 to P = 0.54) and I^2 was 0%—65.4% (low heterogeneity) and the DOR was 6596 (33—32 375, median 9515, Table 2, Figure 3). The +LR was 588 (7—3073 and median 2501) and the −LR was 0.15

(0.05—0.31, median 0.17). This is consistent with an excellent interventional biomarker as a rule-in (score >40 identifies treatment failure) and a good rule-out (score ≤40 demonstrates treatment utility) biomarker.

Finally, to evaluate utility as a multilevel NET biomarker, we undertook a meta-analysis of the AUCs from all individual studies (n = 18). The pooled AUC (fixed model) was 0.954 ± 0.005, with a z-statistic of 175.06 (P < 0.001) (Figure 4A). Utilizing a random effects model, the pooled AUC was 0.897 ± 0.02; the z-statistic was 41.73 (P < 0.001). The Q was 131.5 (P < 0.0001) and the I^2 was 87% (95% CI: 81% to 91%).

The individual P values for the slope coefficients derived from each of the Deeks' funnel plot asymmetry tests were 0.398—0.647 (Figure 4B), confirming a low likelihood of publication bias.
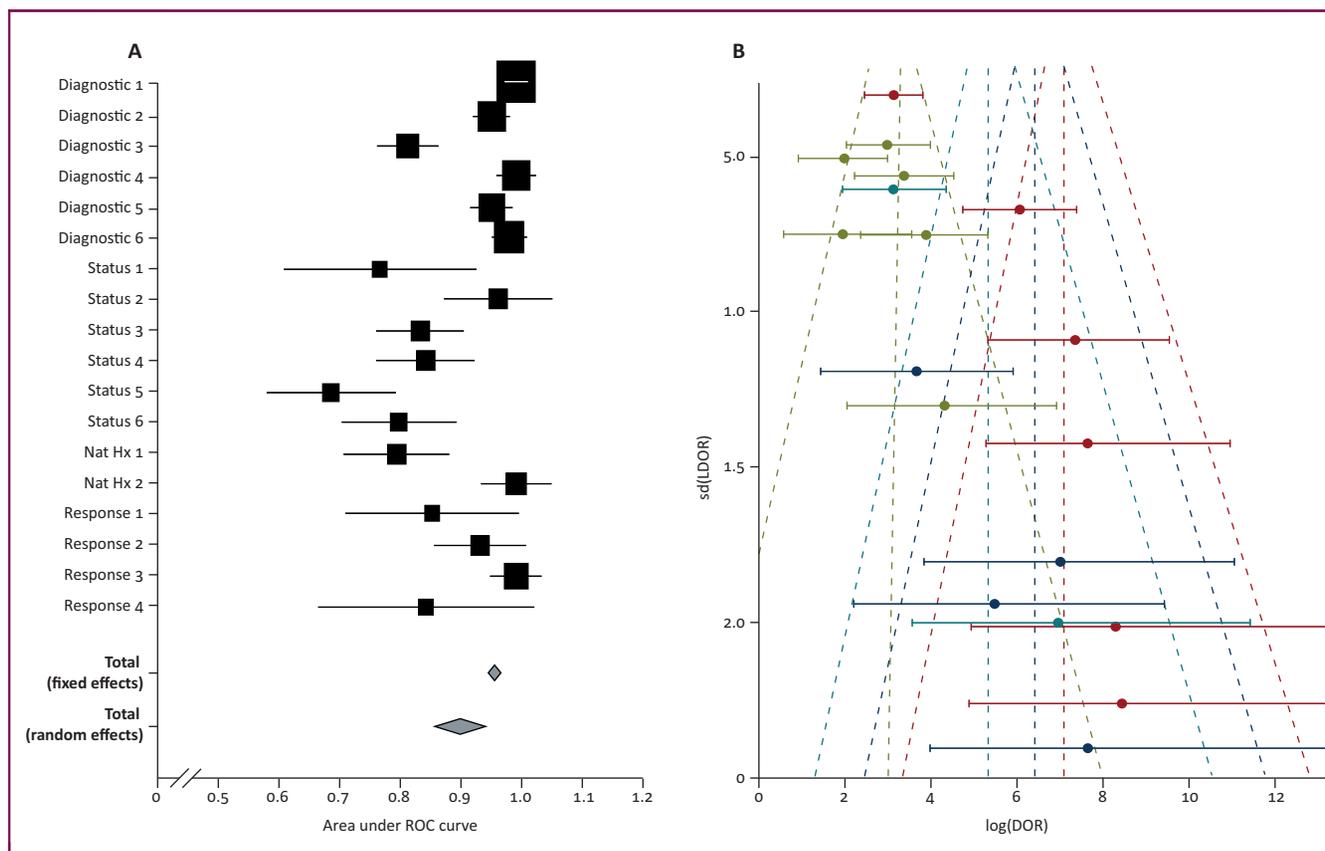
Figure 4. Pooled area under the curve (AUC) of all individual studies (*n* = 18) and Deeks' plot for bias.
(A) Forest plot summating the estimated accuracy and 95% confidence intervals (black boxes/lines) for each of the individual study AUCs. The total effects (fixed and random) are represented by grey diamonds. Individual studies are designated by study type [diagnostic, disease status identification, natural history (Nat Hx), and response biomarker].
The calculated AUCs range from 0.9 to 0.95 (all, *P* < 0.001). (B) Deeks' funnel plots to evaluate for publication bias. The *P* values in all cases were >0.39, commensurate with an absence of bias.

DISCUSSION

The premise for this study was based upon the need to assess whether a multianalyte blood marker for NET disease exhibited metrics that allowed it to be considered an accurate biomarker suitable for clinical use in NET disease. A meta-analysis demonstrated that the diagnostic accuracy of the NETest is extremely high: 95% with a specificity of 95%—98%. The DOR was 5853 and the +LR was 195, with a −LR of 0.06. These data are consistent with the functional utility of an IVD to establish a diagnosis and determine either the presence (rule-in) or absence (rule-out) of a disease. The low to moderate study heterogeneity (from the *Q* and *I*[2] results) supports the utility, since this indicates that studies were directly comparable.

Contrary to IVDs, screening tests are expected to have a high sensitivity for the role of identifying potential disease.[39] These 'low-complexity' tests are studied in asymptomatic but at risk individuals. The NETest is used in individuals considered to potentially have NET disease or who have an established diagnosis.[7] The Clinical Laboratory Improvement Amendments (CLIA) and the New York Department of Health (NYDOH) have accredited the test and consider it a high complexity testing tool. The accuracy/

precision metrics from six independent, multi-institutional studies confirm it as a diagnostic not a screening tool.

The accuracy of the NETest to differentiate progressive and stable disease at the time of blood draw was ~85% with a specificity of ~90%. The estimated mean DOR was 46.3 and the +LR and −LR were 9.4 and 0.28, respectively. These values are based on the Response Evaluation Criteria in Solid Tumors (RECIST) classification of disease. They surpass the 80% concordance proposed by the NIH[40] and exceed the performance of other NET biomarkers (typically ≤50% concordance with disease status).[7] This raises the question as to why the NETest is not 100% concordant with image-based disease status? RECIST criteria are used in clinical studies to ensure comparability but standard serial CT/MRI imaging have well-described sensitivity limitations.[5] RECIST criteria are well-recognized to not be consistently accurate, which reflects difficulties with accurately assessing indolent disease.[4] In addition, CT imaging may even provide false negative output in comparison to functional imaging with ^{68}Ga-somatostatin analog (SSA)-PET/CT.[41,42] Malczewska et al. in the assessment of 111 individuals reported the NETest accurately related to functional imaging (98%) than CT/MRI alone (92%).[32] This observation identified false

Table 2. Summarized DOR and likelihood ratios for the 10 individual studies in each of the different categories

	DOR	+LR	−LR
Diagnostic (n = 6)			
Kidd et al.[27]	202 772	3 707.1	0.04
Filosso et al.[28]	553.6	34.6	0.07
Van Treijen et al.[29]	25.7	3.3	0.14
Al Toubah et al.[30]	230 913	4 247.1	0.04
Malczewska et al. (a)[31]	7 641.4	566.7	0.08
Malczewska et al. (b)[32]	2 996	103.6	0.05
Mean (95% CI)	5 853 (167−25 580)	195 (12−986)	0.06 (0.02−0.1)
Median (range)	5 319 (26−230 913)	335 (3−3707)	0.06 (0.04−0.14)
Disease status (n = 6)			
Pavel et al.[34]	10.2	2.5	0.34
Kidd et al.[27]	192.3	17.8	0.17
Filosso et al.[28]	23.5	4.9	0.23
Liu et al.[33]	36.9	9.2	0.28
Malczewska et al. (a)[31]	8.9	3.3	0.43
Malczewska et al. (b)[32]	67.6	15.7	0.28
Mean (95% CI)	46.3 (3−216)	9.4 (1.6−35.2)	0.27 (0.12−0.54)
Median (range)	30.2 (8.9−192.3)	7.1 (2.5−15.7)	0.28 (0.17−0.43)
Type 0 biomarker (n = 2)			
Pavel et al.[34]	29.9	2.96	0.13
Liu et al.[33]	32 265	322.3	0.07
Mean (95% CI)	5 520 (5.2−22 318)	127 (1.4−720)	0. (0.01−0.35)
Median (range)	16 147 (30−32 265)	163 (3−322)	0.1 (0.07−0.13)
Type II biomarker (n = 4)			
Cwikla et al.[35]	4 478.5	855.9	0.24
Bodei et al.[36]	14 552.2	1 422.5	0.14
Liu et al.[33]	151 609.9	4 146.8	0.11
Malczewska et al. (b)[32]	81.99	10.4	0.21
Mean (95% CI)	6 596 (33−32 375)	588 (6.8−3073)	0.15 (0.05−0.31)
Median (range)	9 515 (82−151 610)	2 501 (10.4−4147)	0.17 (0.11−0.24)

DOR, diagnostic odds ratio; +LR, positive likelihood ratio; −LR, negative likelihood ratio.

negative CT results and indicated the NETest very accurately correlated with somatostatin receptor-based functional disease. These data suggests the NETest not only has utility as an IVD but may have adjunctive value with somatostatin receptor-based imaging. Further studies are required to assess whether a blood test can supplant imaging as a monitoring tool, potentially reducing health care expenses as well as limiting radiation exposure.

We identified two type 0 biomarker studies.[33,34] The accuracy of the biomarker was high (92%−98%) with a sensitivity >97%, a DOR of 5520, a +LR of 127, and a −LR of 0.1. In our study we assessed if a score ≤40 identified stable disease and whether a score >40 was associated with progression. Our analysis indicated that the NETest unequivocally separated stable disease from progression. The NETest therefore clearly meets the criteria of an NIH type 0 biomarker as a surrogate for the natural history of the disease. The NETest also meets the criteria as a biomarker for disease progression. The very high DOR and the effective positive and negative likelihood ratio values confirm that the categories of the NETest (≤40 versus 40−100) strongly and significantly correlated with the natural history of the disease over the 12−18 months evaluated.

We also compared the previously used NET biomarker CgA to assess comparability and to answer the question of how other NET biomarkers compared to NETest. Meta-analyses have been published for CgA in both GEP- and bronchopulmonary NETs. In GEP-NETs, the CgA diagnostic sensitivity was 73% with a 95% specificity.[43] For lung NETs,

the diagnostic sensitivity was 35% with a specificity of 94%.[44] The inability of CgA to attain acceptable metrics (sensitivity >80% and specificity >90%) reflects variations in the different assay methodologies, false negative measurements, as well as numerous non-neuroendocrine conditions or common drugs (proton pump inhibitors [PPIs]) that alter CgA blood levels.[45] Although biomarkers such as pancreastatin and neurokinin A (NKA) have been suggested, they have been demonstrated to function significantly more poorly than CgA.[2,31−33,44,46] Studies involving direct comparisons with the NETest and CgA and pancreastatin and NKA have reported their metrics are substantially lower than measuring circulating mRNA.[17,18,47] Furthermore, each of these markers are monoanalytes that capture the secretory capacity of a NET and do not reflect biological features such as proliferation, metabolism, or growth factor signaling.[15,48]

Pharmacodynamic or response biomarkers[49] are used to demonstrate a biological response has occurred in an individual exposed to a treatment. If they are not mechanistically based, they can function as a type II or 'surrogate' endpoint marker. Typically, levels that change are used to provide early evidence that a treatment is having an effect on a clinical endpoint, for example, PFS. Such a biomarker provides meaningful information whether an intervention is biologically effective and may provide a measure of the durability of response.[49] Three studies evaluated the NETest in the assessment of the effect of somatostatin analogs and one examined the effect of PRRT on PFS as an endpoint.

Using the same strategy as employed for the type 0 biomarker, we examined whether a score ≤40 identified disease stabilization (response to therapy), whereas a score that increased (>40) was associated with treatment failure. This tested whether the identification of disease progression by an increasing NETest score functioned as a surrogate marker for treatment failure. Our analysis demonstrated the accuracy was 94%—97% with a specificity of 99%. The DOR was 6596, with a +LR of 588 and a −LR of 0.15. These data strongly support that changes in the NETest can be used as a biomarker of treatment response. The test can therefore be regarded as an effective surrogate endpoint for treatment responsiveness.

In the US, the NETest is undertaken at an accredited clinical laboratory certified by CLIA (Connecticut, Pennsylvania, Florida, New York, and California) and is NYDOH-accredited. Similarly, it is available in Europe at a certified molecular testing facility in London, UK (HCA Healthcare UK). Validation evaluation includes confirmation that it measures specific target analytes in a highly reproducible fashion (coefficient of variation <5%). The assay includes external controls, and the precision/accuracy (inter- and intra-assay metrics) range between 0.4% and 4.8%.[50] The assay is positive when spiked samples (NET cell lines) are included in normal blood. It has a standard operating protocol and has been effectively calibrated between different laboratories [Wren Laboratories, Branford, USA, and Sarah Cannon Molecular Diagnostics (HCA), London].[51] The assay is stable (>95% correlation between scores and follow-up imaging-based status)[33] and its diagnostic accuracy confirmed as comparable to imaging as a gold standard.[32,33,52] Various studies undertaken over time (some for as long as 5 years)[34] assessing interventions and natural history, identify that it has biological relevance as a biomarker. The assay uses a cut-off of 20 (on a scale of 0—100) to exclude the low-level expression detectable in different circulating blood cell populations under normal or pathological conditions (supplementary material and Figures S2 and S3, available at Annals of Oncology online).

Biomarkers are characterized as tools that facilitate diagnosis, disease prognosis, define clinical disease status, predict response to treatment, or indicate the outcome of treatment.[53] An ideal biomarker should possess all these features. The scientific power of a biomarker is defined by its metrics, including sensitivity and specificity. Sensitivity is the ability of a test to correctly classify an individual as 'diseased' (positive in disease), while specificity is the ability to correctly classify an individual as disease-free (negative in health).[54] There is agreement that NET biomarkers should exceed 90% for specificity and 80% for sensitivity, negative predictive value, or positive predictive value (PPV).[1] Clearly, the NETest meets these criteria.

The United States National Cancer Co-operative Network (NCCN) characterizes CgA as a category 3 biomarker. This is defined as: 'Based upon any level of evidence, there is major NCCN disagreement that the intervention is appropriate.'[55] CgA is not accepted as a NET biomarker based on its metrics. Despite this, CgA remains a component of published guidelines in Europe and the USA.[56,57] One

reason for this is that until recently a multianalyte genomic biomarker has not been available for clinical use in NET disease.

The precedent for the introduction into clinical use of gene expression-based biomarkers has been established. Other disciplines such as breast and prostate cancer have embraced the clinical utility of molecular assays; MammaPrint and Oncotype Dx are two well-known tests.[58] They direct therapy and were included in the American Society for Clinical Oncology (ASCO) and NCCN guidelines before definitive random controlled studies were published.[59] Inclusion was based on prospective-retrospective studies (level of evidence IA and IB[60]). The NETest has been similarly evaluated (Table 1) and functions in an analogous fashion to both these assays.

The NETest has the same output and functions in a similar fashion. It has been reported in a real-world study that patients in a watch-and-wait program that have a low NETest score can continue to be monitored without intervention.[33] This is akin to breast cancer patients who fall into a 'low risk category' per MammaPrint output and are cautioned against using chemotherapy.[61] In this case, the score is used as a rule-out. The NETest risk score provides a comparable strategy for disease monitoring. Since therapeutic agents used for NET disease are relatively ineffective and very expensive, the health economic rationale as well as the patient quality of life issues are worthy of consideration.[7,14,15]

From an economic standpoint, the NETest has advantages. The identification of patients with molecularly stable disease, 40% of whom required fewer follow-up interventions than using standard imaging procedures, has obvious clinical and economic implications.[33] A blood biomarker that identifies disease status earlier than imaging and facilitates earlier cessation of an ineffective therapy also has an obvious cost benefit.[62] As an IVD, the NETest can confirm the diagnosis of a NET and has applications in monitoring and assessing the prognosis of NET disease. A meta-analysis confirms this and that the performance criteria are all >90%. In an evaluation of costs, the positive impact of the NETest on health care expenditure would be significant. We estimate that $240 million to $1.69 billion could be saved annually with a further additional reduction in annual costs of $205 million to $1.1 billion if imaging was reduced by the biomarker (supplementary material, Figure S4 and Table S4, available at Annals of Oncology online). It is likely that the NETest can serve as a non-invasive, easily repeatable investigation that serves as a real-time guide about the status of an individual NET and facilitates implementation of optimal care. The adoption of novel biomarker information that depicts in real-time the molecular biological status of a tumor would benefit NET patients and facilitate the advance of the field in parallel with other oncological disciplines.

DISCLOSURES

The authors have declared no conflicts of interest.

REFERENCES

1. Oberg K, Modlin IM, De Herder W, et al. Consensus on biomarkers for neuroendocrine tumour disease. *Lancet Oncol*. 2015;16:e435—e446.
2. Oberg K, Modlin I, DeHerder W, et al. Biomarkers for neuroendocrine tumor disease: a Delphic consensus assessment of multianalytes, genomics, circulating cells and monoanalytes. *Lancet Oncol*. 2015;16:e435046.
3. Oberg K, Krenning E, Sundin A, et al. A Delphic consensus assessment: imaging and biomarkers in gastroenteropancreatic neuroendocrine tumor disease management. *Endocr Connect*. 2016;5:174—187.
4. de Mestier L, Dromain C, d'Assignies G, et al. Evaluating digestive neuroendocrine tumor progression and therapeutic responses in the era of targeted therapies: state of the art. *Endocr Relat Cancer*. 2014;21:R105—R120.
5. Capdevila J, Casanovas O, Salazar R, et al. Translational research in neuroendocrine tumors: pitfalls and opportunities. *Oncogene*. 2017;36:1899—1907.
6. Yang Z, Tang LH, Klimstra DS. Effect of tumor heterogeneity on the assessment of Ki67 labeling index in well-differentiated neuroendocrine tumors metastatic to the liver: implications for prognostic stratification. *Am J Surg Pathol*. 2011;35:853—860.
7. Modlin IM, Kidd M, Malczewska A, et al. The NETest: the clinical utility of multigene blood analysis in the diagnosis and management of neuroendocrine tumors. *Endocrinol Metab Clin North Am*. 2018;47:485—504.
8. Hulka B. Overview of biological markers. In: Hulka B, Griffith J, Wilcosky T, eds. *Biological Markers in Epidemiology*. New York, New York: Oxford University Press; 1990:3—15.
9. Zhao X, Modur V, Carayannopoulos LN, Laterza OF. Biomarkers in pharmaceutical research. *Clin Chem*. 2015;61:1343—1353.
10. Modlin IM, Moss SF, Chung DC, et al. Priorities for improving the management of gastroenteropancreatic neuroendocrine tumors. *J Natl Cancer Inst*. 2008;100:1282—1289.
11. Modlin IM, Oberg K, Chung DC, et al. Gastroenteropancreatic neuroendocrine tumours. *Lancet Oncol*. 2008;9:61—72.
12. Modlin I, Kidd M, Taylor A, et al. Neuroendocrine tumor biomarkers: current status and perspectives. *Neuroendocrinology*. 2014;100:265—277.
13. Frank R, Hargreaves R. Clinical biomarkers in drug discovery and development. *Nat Rev Drug Discov*. 2003;2:566—580.
14. Modlin I, Drozdov I, Kidd M. The identification of gut neuroendocrine tumor disease by multiple synchronous transcript analysis in blood. *PLoS One*. 2013;8:e63364.
15. Kidd M, Drozdov I, Modlin I. Blood and tissue neuroendocrine tumor gene cluster analysis correlate, define hallmarks and predict disease status. *Endocr Relat Cancer*. 2015;22:561—575.
16. Genzen JR, Mohlman JS, Lynch JL, et al. Laboratory-developed tests: a legislative and regulatory review. *Clin Chem*. 2017;63:1575—1584.
17. Modlin I, Drozdov I, Alaimo D, et al. A multianalyte PCR blood test outperforms single analyte ELISAs for neuroendocrine tumor detection. *Endocr Relat Cancer*. 2014;21:615—628.
18. Modlin IM, Aslanian H, Bodei L, et al. A PCR blood test outperforms chromogranin A in carcinoid detection and is unaffected by PPIs. *Endocr Connect*. 2014;3:215—223.
19. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144:646—674.
20. Chen F, Zhang Y, Gibbons DL, et al. Pan-cancer molecular classes transcending tumor lineage across 32 cancer types, multiple data platforms, and over 10,000 cases. *Clin Cancer Res*. 2018;24:2182—2193.
21. Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ*. 2015;350:g7647.
22. Leeflang MM, Deeks JJ, Takwoingi Y, Macaskill P. Cochrane diagnostic test accuracy reviews. *Syst Rev*. 2013;2:82.
23. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155:529—536.
24. Cochrane. Cochrane Screening and Diagnostic Tests Methods. In Part 2: General Methods for Cochrane Reviews of Diagnostic Test Accuracy. Available at https://methods.cochrane.org/sdt; 2019. The Cochrane Collaboration. Accessed 15 July, 2019.
25. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327:557—560.
26. Guo J, Riebler A. meta4diag: Bayesian bivariate meta-analysis of diagnostic test studies for routine practice. *J Stat Softw*. 2018;83:1—31.
27. Kidd M, Modlin I, Drozdov I, et al. A liquid biopsy for bronchopulmonary/lung carcinoid diagnosis. *Oncotarget*. 2017;9:7182—7196.
28. Filosso P, Kidd M, Roffinella M, et al. The utility of blood neuroendocrine gene transcript measurement in the diagnosis of bronchopulmonary neuroendocrine tumors (BPNET) and as a tool to evaluate surgical resection and disease progression. *Euro J Cardiothorac Surg*. 2018;53:631—639.
29. van Treijen MJC, Korse CM, van Leeuwaarde RS, et al. Blood transcript profiling for the detection of neuroendocrine tumors: results of a large independent validation study. *Front Endocrinol (Lausanne)*. 2018;9:740.
30. Al-Toubah T, Cives M, Valone T, et al. Sensitivity and specificity of the NETest: a validation study. *J Clin Oncol*. 2019;37:222.
31. Malczewska A, Oberg K, Bodei L, et al. NETest liquid biopsy is diagnostic of lung neuroendocrine tumors and identifies progressive disease. *Neuroendocrinology*. 2019a;108:219—231.
32. Malczewska A, Makulik K, Witkowska M, et al. NETest liquid biopsy is diagnostic of small intestine and pancreatic neuroendocrine tumors and accurately correlates with anatomical and functional imaging. *Endocr Connect*. 2019. https://doi.org/10.1530/EC-19-0030.
33. Liu E, Paulson S, Gulati A, et al. Assessment of NETest clinical utility in a US registry-based study. *Oncologist*. 2019;24:783—790.
34. Pavel M, Jann H, Prasad V, et al. NET blood transcript analysis defines the crossing of the clinical rubicon: when stable disease becomes progressive. *Neuroendocrinology*. 2017;104:170—182.
35. Cwikla JB, Bodei L, Kolasinska-Cwikla A, et al. Circulating transcript analysis (NETest) in GEP-NETs treated with somatostatin analogs defines therapy. *J Clin Endocrinol Metab*. 2015;100:E1437—E1445.
36. Bodei L, Kidd M, Modlin IM, et al. Measurement of circulating transcripts and gene cluster analysis predicts and defines therapeutic efficacy of peptide receptor radionuclide therapy (PRRT) in neuroendocrine tumors. *Eur J Nucl Med Mol Imaging*. 2016;43:839—851.
37. van Enst WA, Ochodo E, Scholten RJ, et al. Investigation of publication bias in meta-analyses of diagnostic test accuracy: a meta-epidemiological study. *BMC Med Res Methodol*. 2014;14:70.
38. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol*. 2005;58:882—893.
39. Siravegna G, Marsoni S, Siena S, et al. Integrating liquid biopsies into the management of cancer. *Nat Rev Clin Oncol*. 2017;2:14.
40. Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med*. 2003;138:W1—W12.
41. Bodei L, Sundin A, Kidd M, et al. The status of neuroendocrine tumor imaging: from darkness to light? *Neuroendocrinology*. 2015;101:1—17.
42. Toumpanakis C, Kim MK, Rinke A, et al. Combination of cross-sectional and molecular imaging studies in the localization of gastroenteropancreatic neuroendocrine tumors. *Neuroendocrinology*. 2014;21:21.
43. Yang X, Yang Y, Li Z, et al. Diagnostic value of circulating chromogranin A for neuroendocrine tumors: a systematic review and meta-analysis. *PLoS One*. 2015;10:e0124884.
44. Malczewska A, Kidd M, Matar S, et al. An assessment of circulating chromogranin A as a biomarker of bronchopulmonary neuroendocrine neoplasia: a systematic review and meta-analysis. *Neuroendocrinology*. 2019. https://doi.org/10.1159/000500525.
45. Gkolfinopoulos S, Tsapakidis K, Papadimitriou K, et al. Chromogranin A as a valid marker in oncology: clinical application or false hopes? *World J Methodol*. 2017;7:9—15.

46. Matar S, Malczewska A, Oberg K, et al. Blood chromogranin A is not effective as a biomarker for diagnosis or management of broncho-pulmonary carcinoids. *Neuroendocrinology.* 2019. https://doi.org/10.1159/000500202.

47. Modlin IM, Kidd M, Bodei L, et al. The clinical utility of a novel blood-based multi-transcriptome assay for the diagnosis of neuroendocrine tumors of the gastrointestinal tract. *Am J Gastroenterol.* 2015;110:1223—1232.

48. Walenkamp A, Crespo G, Fierro Maya F, et al. Hallmarks of gastrointestinal neuroendocrine tumours: implications for treatment. *Endocr Relat Cancer.* 2014;21:R445—R460.

49. Amin S, Bathe OF. Response biomarkers: re-envisioning the approach to tailoring drug therapy for cancer. *BMC Cancer.* 2016;16:850.

50. Modlin I, Drozdov I, Kidd M. Gut neuroendocrine tumor blood qPCR fingerprint assay: characteristics and reproducibility. *Clin Chem.* 2014;52:419—429.

51. Kidd M, Drozdov IA, Matar S, et al. Utility of a ready-to-use PCR system for neuroendocrine tumor diagnosis. *PLoS One.* 2019;14:e0218592.

52. Bodei L, Kidd M, Modlin IM, et al. Gene transcript analysis blood values correlate with (68)Ga-DOTA-somatostatin analog (SSA) PET/CT imaging in neuroendocrine tumors and can define disease status. *Eur J Nucl Med Mol Imaging.* 2015;42:1341—1352.

53. Modlin IM, Oberg K, Taylor A, et al. Neuroendocrine tumor biomarkers: current status and perspectives. *Neuroendocrinology.* 2014;100: 265—277.

54. Parikh R, Mathai A, Parikh S, et al. Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol.* 2008;56:45—50.

55. Shah MH, Goldner WS, Halfdanarson TR, et al. NCCN guidelines insights: neuroendocrine and adrenal tumors, version 2.2018. *J Natl Compr Canc Netw.* 2018;16:693—702.

56. Oberg K, Couvelard A, Delle Fave G, et al. ENETS consensus guidelines for standard of care in neuroendocrine tumours: biochemical markers. *Neuroendocrinology.* 2017;105:201—211.

57. Shah MH, Kulke MH, Goldner WS, et al. NCCN guidelines: neuroendocrine and adrenal tumors, version 3.2018 -Sept 11. *J Natl Compr Canc Netw.* 2018:693—702.

58. Krop I, Ismaila N, Andre F, et al. Use of biomarkers to guide decisions on adjuvant systemic therapy for women with early-stage invasive breast cancer: American Society of Clinical Oncology clinical practice guideline focused update. *J Clin Oncol.* 2017;35:2838—2847.

59. Gradishar WJ, Anderson BO, Balassanian R, et al. NCCN guidelines insights: breast cancer, version 1.2017. *J Natl Compr Canc Netw.* 2017;15: 433—451.

60. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst.* 2009;101: 1446—1452.

61. Cardoso F, van't Veer LJ, Bogaerts J, et al. 70-gene signature as an aid to treatment decisions in early-stage breast cancer. *N Engl J Med.* 2016;375:717—729.

62. Strosberg J, Casciano R, Stern L, et al. United States-based practice patterns and resource utilization in advanced neuroendocrine tumor treatment. *World J Gastroenterol.* 2013;19:2348—2354.

63. Genç CG, Jilesen APJ, Nieveen van Dijkum EJM, et al. Measurement of circulating transcript levels (NETest) to detect disease recurrence and improve follow-up after curative surgical resection of well-differentiated pancreatic neuroendocrine tumors. *J Surg Oncol.* 2018;118:37—48.

64. Modlin IM, Frilling A, Salem RR, et al. Blood measurement of neuroendocrine gene transcripts defines the effectiveness of operative resection and ablation strategies. *Surgery.* 2016;159:336—347.