A Delphic consensus assessment: imaging and biomarkers in gastroenteropancreatic neuroendocrine tumor disease management

Kjell Oberg1, Eric Krenning2, Anders Sundin1, Lisa Bodei3, Mark Kidd4, Margot Tesselaar5, Valentina Ambrosini6, Richard P Baum7, Matthew Kulke8, Marianne Pavel9, Jaroslaw Cwikla10, Ignat Drozdov4, Massimo Falconi11, Nicola Fazio12, Andrea Frilling13, Robert Jensen14, Klaus Koopmans15, Tiny Korse9, Dik Kwekkeboom2, Helmut Maecke16, Giovanni Paganelli17, Ramon Salazar18, Stefano Severi17, Jonathan Strosberg19, Vikas Prasad2, Aldo Scarpa20, Ashley Grossman21, Annemeik Walenkamp22, Mauro Cives19, Irene Virgolini23, Andreas Kjaer24 and Irvin M Modlin25

1Uppsala University, Uppsala, Sweden
2Erasmus Medical Center, Rotterdam, Netherlands
3Memorial Sloan Kettering Cancer Center, New York, New York, USA
4Wren Laboratories, Branford, Connecticut, USA
5Netherlands Cancer Institute, Amsterdam, Netherlands
6University of Bologna, Bologna, Italy
7Zentralklinik Bad Berka, Bad Berka, Germany
8Dana Farber Cancer Institute, Boston, Massachusetts, USA
9Charite Hospital, Berlin, Germany
10University of Warmia and Mazury, Olsztyn, Poland
11Ospedale San Raffaele, Milan, Italy
12IEO (European Institute of Oncology), Milan, Italy
13Imperial College London, London, UK
14National Institutes of Health, Bethesda, Maryland, USA
15Martini Ziekenhuis, Groningen, Netherlands
16University Hospital Freiburg, Freiburg, Germany
17Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori, Meldola, Italy
18Instituto Catala d’Oncologia, Barcelona, Spain
19H. Lee Moffitt Cancer Center, Tampa, Florida, USA
20University of Verona, Verona, Italy
21University of Oxford, Oxford, UK
22University of Groningen, Groningen, Netherlands
23Medical University Innsbruck, Innsbruck, Austria
24Copenhagen University, Copenhagen, Denmark
25Yale University, New Haven, Connecticut, USA

Abstract

The complexity of the clinical management of neuroendocrine neoplasia (NEN) is exacerbated by limitations in imaging modalities and a paucity of clinically useful biomarkers. Limitations in currently available imaging modalities reflect difficulties in measuring an intrinsically indolent disease, resolution inadequacies and inter-/intra-facility device variability and that RECIST (Response Evaluation Criteria in Solid Tumors) criteria are not optimal for NEN. Limitations of currently used biomarkers...
are that they are secretory biomarkers (chromogranin A, serotonin, neuron-specific enolase and pancreastatin); monoanalyte measurements; and lack sensitivity, specificity and predictive capacity. None of them meet the NIH metrics for clinical usage. A multinational, multidisciplinary Delphi consensus meeting of NEN experts \((n = 33)\) assessed current imaging strategies and biomarkers in NEN management. Consensus (>75%) was achieved for 78% of the 142 questions. The panel concluded that morphological imaging has a diagnostic value. However, both imaging and current single-analyte biomarkers exhibit substantial limitations in measuring the disease status and predicting the therapeutic efficacy. RECIST remains suboptimal as a metric. A critical unmet need is the development of a clinico-biological tool to provide enhanced information regarding precise disease status and treatment response. The group considered that circulating RNA was better than current general NEN biomarkers and preliminary clinical data were considered promising. It was resolved that circulating multianalyte mRNA (NETtest) had clinical utility in both diagnosis and monitoring disease status and therapeutic efficacy. Overall, it was concluded that a combination of tumor spatial and functional imaging with circulating transcripts (mRNA) would represent the future strategy for real-time monitoring of disease progress and therapeutic efficacy.

Introduction

The management of neuroendocrine neoplasms (NENs, also called ‘NETs’) remains clinically challenging despite advances in classification systems (1), inauguration of novel therapies and innovations in imaging and the introduction of multidisciplinary management strategies (2). In particular, the management of NEN reflects diverse approaches often based on empiric pronouncements, local practical experience or the availability of certain therapies. Despite the promulgation of effective and applicable guidelines (e.g., WHO/ENETs classification of 2010 (3, 4) and their regular reassessment, a critical limitation is the dearth of large, randomized prospective trials. The precise delineation of definable strategies is further constrained by the tumor heterogeneity (diverse cell types, disparate molecular regulatory mechanisms and ill-understood oncogenic drivers) (5, 6). As a consequence, five-year survival rates diverge widely (15–95%), depending on the primary site, variable tumor biology, disease extent at diagnosis, available therapeutic options and designated centers of care (7, 8, 9). Therapeutic options remain diverse and run the full gamut from mechanistic excision to pharmacological intervention and the infusion of radioactive somatostatin analogs (10). Strategies include somatostatin receptor agonists, ‘targeted’ agents (mTOR inhibitors and VEGF antagonists), immunotherapy (interferon), cytotoxic chemotherapy, peptide receptor radionuclide therapy (PRRT), external radiation and interventional radiological or probe-directed ablation (11). In those with ‘indolent tumor behavior’, a watch-and-wait-strategy is considered appropriate in certain selected cases (12). Apart from ‘early identified’ (usually serendipitous) appendiceal, rectal or gastric NETs, cure is uncommon, and overwhelmingly, the majority of treatment includes diverse combinations of strategies to delay local or metastatic disease progression (13). Given their relatively slow growth, continual assessment by imaging, biomarker levels and overall survival represents the fundamental basis for all management strategies. The need to monitor tumor responsiveness, both in clinical trials and in routine practice, is mandatory given the range of expensive, empirical and often times toxic treatment choices used (14).

For many non-neuroendocrine neoplasms, therapeutic responsiveness is assessed through imaging, but for NENs, this has well-described limitations (15, 16, 17). Anatomic imaging using the Response Evaluation Criteria in Solid Tumors (RECIST) criteria exhibits well-documented limitations (18, 19, 20). These include issues with lesion dimensionality and measurements thereof, effects of therapy on lesion appearance itself, difficulties with reproducibility and accurate delineation of metastatic disease, particularly extra-liver disease. The development of new lesions is probably the most powerful indicator of disease progression. Functional imaging with somatostatin receptor-based strategies, for example, \(^{68}\text{Ga}-\text{SSA}-\text{PET/CT}\), has proved of considerable value (21); however, limited
spatial resolution (6–8 mm for PET scanners) and partial volume effects constrain the ability to delineate small lesions. As a consequence, timely, clinically reproducible assessments of progression remain unattainable (22, 23). Changes in the 68Ga-SSA tumor standardized uptake value (SUV) during treatment have not been a reliable measure for therapy monitoring (24, 25). 18F-DG-PET, although useful prognostically, is not established as an early harbinger of tumor progression (26). Despite significant advances, current imaging strategies in NENs remain suboptimal (27, 28) and exhibit significant limitations. In particular, the identification and delineation of residual (and occult) disease is difficult.

Credible general biomarkers with broad clinical utility for gastroenteropancreatic (GEP) NENs remain unavailable although chromogranin A (CgA) and urinary 5-hydroxyindoleacetic acid (5-HIAA; in serotonin-secreting tumors) have been used in this capacity (29). Secretory (monoaalyte) biomarkers for specific tumor types (insulinoma: insulin, gastrinoma: gastrin, glucagonoma: glucagon and VIPoma: VIP) are effective serum indicators of tumor activity, but as this group of lesions represents a minority of NENs (<3–5%), its broad utility is limited. CgA is a constitutive product of the neuroendocrine cell secretory granule and is measurable in serum or plasma. It has been variably reported to correlate with tumor biology and mass and prognosticate survival (30, 31). Despite initial enthusiasm, the limitations of CgA have become increasingly evident. There is considerable discrepancy as to whether alterations in CgA have clinical utility in the identification of progressive disease. Although there has been some improvement in comparable unit use, there is no reference CgA standard, and wide variations exist in the assay measurements in different laboratories (30). Furthermore, the sensitivity of CgA ranges from 60% to 90% with a specificity <50% (depending on the population studied) (32). This reflects the CgA elevations associated with numerous non-NEN-related conditions including renal failure, cardiac disease, other neoplasia and PPI administration (30).

The complexity and diversity of the biological behavior of a cancer or its response to therapy have been effectively addressed in scientific publications (33, 34). The limitations of secretory products to define the permutations of oncogenic genomic regulators are apparent and have led to the development of molecular technologies to better delineate cancer biology (35, 36). This biological research has identified extensive interfacing mechanisms that delineate GEP-NEN neoplastic development (37). A key unmet need is the identification of what constitutes the driver of neoplastic development (i.e., driver mutations) and whether this is clinically actionable i.e., targetable, and can be used as a predictive biomarker.

The majority of tumors (~95%) do not exhibit germline mutations (6, 38). Although genomic studies have revealed a number of sporadic genomic alterations, particularly in pancreatic NENs, the relationship between specific genes and tumor pathobiology remains unclear (5). Unlike the majority of cancers, activating mutations are infrequent if not largely unknown in GEP-NEN (5) with most tumors exhibiting mutations (when identified) in tumor suppressor genes. Although genomic studies seeking underlying driver mutations have proven disappointing (39, 40), transcriptome assessments have been useful in identifying and differentiating the different subtypes of NENs (based on origin e.g., pancreatic vs small intestinal, and aggressiveness e.g., non-progressive vs malignant/metastatic) (41, 42) and have demonstrable predictive utility at a tissue level (43). More recently, blood-based assays (CTCs, miRNA and circulating mRNA) have been developed. The most extensively investigated biomarker tool is blood-based multianalyte transcript analysis (44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54). Blood gene expression of tumor biomarkers closely correlates with tumor tissue expression levels, and analysis of relevant clusters captures NEN biology facilitating accurate definition of clinical status (37). The clinical application of such blood-based information to the management of NEN disease has therefore become a subject for investigation. Likewise, the concept of fusing such data with functional imaging to provide a synergistic monitoring platform is worthy of consideration, especially given the current limitations in accurate monitoring.

Although biomarkers have been used in conjunction with imaging as adjuncts to inform clinical decision making, ‘biochemical’ responses using monoaalytes are often non-concordant with image-based assessments (10, 55). The detailed analysis of other neoplastic diseases has led to the recognition that evaluation of monoaalyte secretory products (exocytotic or secreted proteins) alone fails to adequately describe the diversity of neoplastic pathobiology (56). Thus, complex analytic strategies measuring diverse regulators of neoplastic cell biology interfaced with mathematical algorithms to facilitate interpretation have been developed for breast, lung and hematological malignancies (57, 58, 59, 60). Therefore, a key unmet need that remains is the development of a clinically applicable, multianalyte biomarker that captures NEN behavior and can be used
to guide clinical management strategies. The use of such blood-based molecular information in combination with functional imaging would provide non-invasive real-time multidimensional information regarding tumor behavior.

Based on the need for a better understanding of the relationship between imaging and therapeutic assessment in NEN disease and the emergence of molecular-based biomarkers that have utility in assessing disease status, e.g., blood-based multianalyte transcript analysis NETest (37), a meeting of multidisciplinary experts in the field was convened in Castelldefels, Spain in March 2015. The goals of this forum were twofold. First, to establish a consensus on the state of the art of imaging and biomarkers in NEN, and secondly, to identify how these two information disciplines could be interfaced to provide added value in clinical decision making and therapeutic response assessment. This meeting represents a follow-up of a previous, more biomarker-focused Delphi consensus meeting that specifically examined the current status of circulating analytes in the management of GEP-NETs with respect to their individual metrics and clinical utility (61).

Materials and methods

Thirty-three multinational experts in the field of NEN disease diagnosis and management were identified including nuclear medicine physicians (n=12; A Kjaer, E Krenning, D Kwekkeboom, L Bodei, V Ambrosini, R Baum, J Cwikla, G Paganelli, S Severi, H Maecke, V Prasad, I Virgolini), radiologists (n=2; A Sundin, K Koopmans), endocrinologists (n=2; M Pavel, A Grossman), gastroenterologists (n=1, R Jensen), oncologists (n=9, K Oberg, M Tesselaar, M Kulke, N Fazio, R Salazar, J Strosberg, A Walenkamp, M Cives, T Meyer (see Authors contributions)), pathologists (n=1, A Scarpa), basic scientists (n=3, M Kidd, I Drozdov, T Korse) and surgeons (n=3: M Falconi, A Frilling, I Modlin). The Delphi method (62) was used to achieve consensus on 142 questions, using a 75% agreement threshold as the basis for achieving consensus (61). Questions were categorized into four major groups (therapeutic management, imaging, molecular status of NETs and biomarkers). The first iteration of the statements to be discussed was developed by a core group (KO, EK, LB, IMM) and distributed to all participants eight weeks before the conference. This first round of electronic assessment was undertaken to eliminate or redefine inconsistencies or ambiguous statements (61). After integration of the primary assessment comments from all participants, this second list (revised) of statements/questions (yes or no responses) was electronically distributed one month ahead of the consensus meeting. All participants provided answers to this interrogatory. The collated results of the entire group responses were made available to all participants at the initiation of the meeting. The meeting format comprised two co-moderators for each discussion session. Any question with less than 75% prior agreement (either consensus: yes or consensus: no) was then reviewed and discussed by the entire panel and re-voted on. Voting was anonymous (electronic touch pad) with re-wording of ambiguous, controversial or non-consensus statements as proposed by participants with the objective of attaining a 75% agreement threshold (61). Up to five re-iterations of a proposal were undertaken before considering an issue resolved. Resolution was achieved in 78%. Not all questions (22%) resulted in a consensus.

Results

A total of 142 questions and sub-questions were posed. The first round of electronic consensus was achieved before the March 2015 meeting in 69 (48.5%). At the meeting, after statement/question reformulation and repeat voting, final consensus was achieved on 111 (78%). The full lists of statements and voting results are documented in the Supplementary Results, see section on supplementary data given at the end of this article. Three participants (ID, HM and DK) were unable to attend the meeting and participate in the final round of voting. The final consensus therefore includes input from these members at rounds 1 and 2 but not round 3.

Therapeutic management

Consensus was achieved on 30 questions (47%) before the meeting. A further 16 (total of 72%) met consensus after discussion and re-voting. The panelists agreed that optimal management strategies required assessment of information based on histology, grade and stage, specific and non-specific symptoms, as well as knowledge regarding the patient’s overall condition. However, they also decided that clinical knowledge alone was inadequate for predicting whether a NEN would be progressive or exhibit a stable disease. Although a wait-and-see strategy was considered an acceptable management strategy, there was full concurrence that current diagnostic parameters were neither of adequate sensitivity nor specificity for defining progress. Moreover, currently available randomized controlled trial (RCT) data were considered insufficient.
to accurately delineate the optimal therapeutic sequence strategy in NEN disease. Overall, the group concluded that there was a paucity of rigorous data available to facilitate objective, clinical decision making.

With respect to imaging, current standard diagnostic parameters are neither sensitive nor specific enough to define progress. Additional predictors of the individual course of disease are therefore required to identify individuals in whom early treatment may be of benefit. This would include additional imaging parameters. Limitations in the assessment of therapeutic responses with current imaging have a negative impact on patient management. Limitations in the discriminant index of both anatomic and functional imaging diminished the accuracy of assessment of therapeutic response. Somatostatin receptor (SSR) density was considered a relevant parameter but knowing the liver tumor load and pretreatment growth rate were considered important predictors of disease course. It was agreed that additional predictors of the individual course of a specific tumor are required to define those in whom early treatment may be of benefit. Biomarkers including but not limited to tissue gene signatures, circulating genetic information and mutational events were considered critical requirements for such a strategy.

The thresholds and cutoffs for defining Ki67 were considered problematic for defining when chemotherapy should be considered. No consensus could be reached upon the precise applicable cutoff. Ki67 was not considered a relevant parameter for predicting SSA response. Surgery was considered the only curative treatment, and a blood signature that could predict disease relapse after R0/R1 (primary or liver) resection was agreed upon as an important requirement. It was identified that selective internal radiation therapy (SIRT), radio frequency ablation (RFA) and trans-arterial (chemo-) embolization (TACE/TAE) were all effective in metastatic liver disease, though individual modalities differed in efficacy based upon patient selection and disease status (63). Individual interventions were noted to have adverse events although lack of comparable data prevented rigorous comparison (63). No consensus was reached regarding associations with adverse events. The use of somatostatin analogs (SSAs) should not only be limited to midgut and pancreatic NENs with Ki67 <10%. No consensus, however, could be reached as to whether SSAs were effective early in the disease course to prevent disease progression. Likewise, it was not accepted that there was evidence that above-label doses should be used in non-functioning progressive disease. There were also no sufficient data to support the use of SSAs as anti-proliferative agents in patients with significant metastatic burden, e.g., >50% neuroendocrine tumor liver metastases (NELM) and/or extra-hepatic metastases. The panel was unsure whether everolimus had a role in non-pancreatic NEN disease (it should be noted that this meeting occurred before the publication of the Radiant-4 study (64)). Controversy was also apparent regarding the initial therapeutic use of chemotherapy. The group was of the opinion that PRRT might warrant consideration at an earlier time point in the therapeutic strategy for management of NETs (it should be noted that this meeting occurred before the availability of the NETTER-1 study results (65)). It was, however, deemed appropriate to consider the use of PRRT before other targeted therapies. Overall, a substantial lack of consensus (~28%) was evident for GEP-NEN therapeutic management. This likely reflects the individualized, empiric-based approaches and the divergent views of European and US experts.

Imaging

Consensus was achieved in 72% of questions (Fig. 1). There was agreement that CT or MRI should be used in conjunction with functional imaging. 68Ga-SSA-PET/CT was preferred to 111In-pentetreotide scintigraphy for functional imaging. 68Ga-SSA-PET/CT was considered the preferred approach compared with 18F-DOPA imaging for pancreatic and small intestinal NEN diagnosis. 18F-FDG-PET/CT was considered useful for differentiating high- from low-grade tumors, which might have future implications for staging. The technique, however, has prognostic implications although this requires validation in larger series. No consensus, however, was reached regarding combining 18F-FDG- and 68Ga-SSA-PET/CT or the timing of imaging for use of each of these modalities in a diagnostic setting.

Imaging was considered the best current modality for measuring treatment efficacy, but no consensus was achieved regarding the optimal strategy, PET/CT or CT or MRI. It was agreed that RECIST criteria were not appropriate for defining the therapeutic responses in NETs at least for biological therapy, and furthermore, inclusion of morphologic parameters, e.g., attenuation measurements, were not considered useful. No consensus was reached regarding whether ‘cold’ analogs e.g., Sandostatin or lanreotide (non-radioactive without bound isotopes) should be discontinued before somatostatin receptor
imaging (SRI). Overall, the heterogeneity in SSR expression was considered a potential sensitivity limitation to this approach because current ligands are SSR2/5 avid. Similarly, the SUV\textsubscript{\text{max}} was also not considered an entirely reliable parameter for assessing patient management based on current ligand–receptor affinities (66). Based on currently available studies, different \textsuperscript{68}Ga-DOTA-SSA peptides (DOTA-TOC, DOTA-NOC and DOTA-TATE) were individually as effective in their diagnostic accuracy. All of them were considered to have clinical utility in determining clinical management.

Overall, imaging was considered more sensitive than existing biomarkers for detecting the disease. The group concurred that more effective circulating biomarkers would be useful adjuncts for assessing treatment. It was agreed that current biomarkers such as CgA do not correlate with imaging, particularly \textsuperscript{68}Ga-DOTA-SSA and \textsuperscript{18}F-FDG imaging. No consensus could be reached for the relationship between CT or MRI and CgA. Overall, the panel agreed that integration of a clinically relevant, biologically effective biomarker strategy into response criteria was required to improve NEN therapy monitoring.

**Molecular status of NETs**

Consensus was achieved in majority of questions (95%). Metabolic pathways were agreed to be poorly characterized. The PI\textsubscript{3}K/mTOR pathway was not considered to be the principal growth regulatory pathway in NENs. It is as yet unclear what constitutes the precise mechanistic basis of the critical growth regulatory pathways of neuroendocrine tumor cells. Despite the proposal of numerous putative targetable pathways, current agents are not generally accepted as being of robust clinical utility (67). Alternative pathways remain to be defined. Mutations in the mTOR pathway were noted to occur in <15\% of pancreatic NENs, and the objective response rate for everolimus (mTOR pathway inhibitor) is ~10\% with disease stabilization in ~75\% (68). The discrepancy between mutation rate and therapeutic efficacy is currently difficult to reconcile. Selective PI\textsubscript{3}K inhibitors were considered useful for overcoming everolimus resistance although the mechanisms of resistance remain to be defined. Mutations in the ATRX/DAXX pathways were not considered major indicators of clinical outcome, and it was agreed that they should not be routinely assessed in pancreatic NENs. In patients with multiple endocrine neoplasia type I (MEN1 syndrome) (germline MEN-1 mutation), the type of menin mutation was not considered to be of prognostic significance. Likewise, alterations in methylation patterns were not considered clinically useful, whereas O6-methylguanine DNA transferase deficiency was regarded as not significant in influencing the choice of therapy. Irrespective of the

---

**Figure 1**

Clinical utility of imaging overview (Section B). Imaging for diagnosis (left) was considered effective (71\% positive); \textsuperscript{68}Ga-DOTA-SSA PET/CT was considered more useful than either \textsuperscript{111}In-pentetreotide scintigraphy (100\%) or \textsuperscript{18}F-DOPA-PET/CT (89\%) for diagnosis of well-differentiated NENs. \textsuperscript{18}F-DOPA-PET/CT was agreed to accurately differentiate (88\%) low- from high-grade tumors. Imaging in therapeutic assessment (right) was overall considered suboptimal (36\%). No consensus (gray) could be reached regarding the utility of either CT/MRI (40\%) or PET-CT (46\%) in the assessment of therapy. A combination of CT/MRI and functional imaging were considered useful (84\%). There was a negative assessment of current methodologies including RECIST criteria (82\%) and Hounsfield Units (Choi criteria) (76\%). \textsuperscript{68}Ga=\textsuperscript{68}Ga-DOTA-SSA PET/CT; \textsuperscript{111}In=\textsuperscript{111}In-pentetreotide scintigraphy; \textsuperscript{18}F=\textsuperscript{18}F-DOPA-PET/CT; HU=Hounsfield Units.
individual molecular abnormality described, cell line models were considered unreliable for identifying and confirming the utility of any targeted agent.

No consensus could be reached regarding the role of VEGF expression and tumor aggressiveness. It was agreed that immunohistochemistry for SSR was not needed to define a treatment strategy, but immunohistochemistry (IHC) e.g., CDX2 and PAX6 was recommended when a primary site was unknown (CUP). Gene profiling, in this setting (CUP), however, was not clinically recommended. Overall, it remained unclear how molecular alterations, particularly at a DNA level, could potentially improve clinical management strategies. It was concluded that molecular alterations as currently defined did not have a current role in NEN treatment, but the panel did support continued investigation in these areas to further define the molecular basis of NEN disease.

**Biomarkers**

A consensus was reached in 89% of questions (Fig. 2). It was agreed that despite the paucity of DNA-related clinically actionable biomarkers, genomics technology had significant potential for identifying novel tissue biomarkers. The conclusion, however, was that at present, insufficient specific mutations and treatment-targetable mutations had been identified. As such, circulating DNA was therefore not considered a viable option for the development of a biomarker.

In general, circulating tumor cells (CTCs) were agreed not to be reliable, sensitive or specific for the detection (88% no) and diagnosis (92% no) of NENs. Furthermore, once tumors were diagnosed, CTCs were considered not to correlate with grade (77% no) or to have clinical utility as either a prognostic (85% no) or predictive biomarker (77% no). No consensus was achieved relating the utility of CTCs as an indicator of tumor burden. Although miRNA was considered interesting and potentially useful as a circulating biomarker, the group agreed that current technology was not adequately robust to support its clinical usage. Metabolomics was also considered of positive interest (83% yes) as was the identification of novel blood GEP-NEN biomarkers. The consideration of metabolomic assessment in urine was not supported (83% no). Tumor transcriptomes and mRNA studies were agreed to be useful for identifying tissue biomarkers and are more sensitive than standard biomarkers. Circulating mRNA assays were agreed to be worthy of further investigation given their potential clinical utility.

**Discussion**

The Delphi method, originally developed by the RAND Corporation (62), has been used extensively to develop consensus in health care. We have previously assessed its utility in similar clinical decision-making settings (61, 69). In this meeting, a substantial overall consensus (~80%) was achieved with 31 questions (~20%) ultimately unresolved (no consensus achieved). A consensus level of 75% was used as clear evidence of a majority opinion. Voting was anonymized (electronic) and followed by discussion when there was no consensus. The actual numbers of participants who completed all three rounds (n = 30, 91% inclusion) is similar to other Delphi-based studies for NENs and met the acceptability criteria for validity (69, 70).

Therapeutic management and imaging achieved the lowest consensus (72%) compared with molecular biology and biomarkers (88–95%). This likely reflects two issues. First, individual approaches to management (despite a focus on multidisciplinary methods) and secondly, differential access to imaging ($^{68}$Ga-DOTA-SSA PET/CT is currently not generally available in the United States). There was a full consensus that surgery was potentially curative. Similarly, there was broad consensus of the utility of $^{68}$Ga-DOTA-SSA PET/CT both in establishing
a diagnosis and having a role in staging, predicting response to PRRT and determining prognosis. There are a number of different national and societal neuroendocrine guidelines that variably evaluate the usage of biomarkers and imaging (North American – NENETs, National Comprehensive Cancer Network – NCCN, Canadian NETs and the European Neuroendocrine Tumor Society – ENETs (14, 71, 72, 73, 74, 75)). Each broadly supports the points defined in this Delphi consensus, but neither specifically addresses the interface between imaging and biomarkers nor the best strategy to integrate anatomical and functional imaging with circulating molecular information. In particular, the current consensus meeting evaluated not only the utility of the different strategies (imaging and biomarkers) but also how such modalities could be interfaced to provide a real-time assessment of the biological evolution of a neuroendocrine neoplasm. It was widely agreed that current approaches (RECIST) for assessing therapeutic responses were inadequate. In particular, clinical knowledge was considered insufficient for early and accurate predictions of progressive or stable disease. Moreover, it was agreed that a clinically actionable, biologically relevant biomarker should be included in treatment response assessments. This is consistent with the agreement reached in the previous Delphi consensus meeting (2014) that was designed to specifically address biomarker metrics and clinical utility (61).

Although biomarkers such as CgA are currently used in conjunction with imaging as adjuncts for clinical decision making (Fig. 3), significant refinements are required (61). In particular, implementations of more informative molecular tools such as multianalyte biomarkers are needed. Dynamic characterization of tumor behavior based on blood-derived genomic information is likely to be of considerable clinical utility, especially if used as an adjunct to both spatial and functional imaging. This is underscored by the lack of utility and clinical effectiveness of solely secretory biomarkers. For example, CgA does not correlate with imaging, particularly 68Ga-DOTA-SSA and 18F-FDG imaging, whereas CgA biochemical ‘responses’ to therapy are also typically non-concordant with imaging (61). Indeed, a number of national and societal guidelines adjudge CgA to be ‘controversial’ in clinical decision making (14, 71).

Imaging alone, however, also has its limitations. The panel agreed that current strategies, although useful in diagnosis, were unlikely to be improved in NENs in the near future. For example, measurements of changes

**PROPOSED STRATEGY FOR ASSESSING THERAPEUTIC EFFICACY**

<table>
<thead>
<tr>
<th>MODALITY</th>
<th>FOLLOW-UP</th>
<th>RATIONALE</th>
<th>ASSESSMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morphological</td>
<td>CT/MRI: 4-6 monthly</td>
<td>Extent and Characterization</td>
<td>Accurate Assessment</td>
</tr>
<tr>
<td>Functional</td>
<td>68Ga-SSA-PET: 4-6 monthly</td>
<td>Quantify SSR</td>
<td>Metabolic Activity</td>
</tr>
<tr>
<td>IMAGING</td>
<td>18F-FDG-PET: 6 monthly</td>
<td>Metabolism</td>
<td>Disease Topography/Extent</td>
</tr>
<tr>
<td>BIOMARKERS</td>
<td>Circulating mRNA</td>
<td>Tumor Stabilization or Progression</td>
<td>Accurate Assessment</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Treatment Efficacy</td>
</tr>
</tbody>
</table>

*Figure 3*

Proposed strategy for assessing the therapeutic efficacy. An integration of functional imaging and biomarker measurement including circulating tumor mRNA will provide combinatorial information on a real-time basis of disease status. The combination of individual imaging strategies will quantify tumor location/extent and in addition delineate somatostatin receptor expression (SRI – typically 68Ga-DOTA-SSA PET/CT) and tumor metabolism (18F-FDG-PET/CT). Circulating mRNA will measure tumor biological activity and identify treatment response.
in Hounsfield Units, proposed in the Choi criteria for measuring GIST treatment responses (15), may not be useful in GEP-NENs. Although suitable for a rough estimate, SUV\(_{\text{max}}\) determined by \(^{68}\text{Ga-SSA-PET/CT}\), was also not considered to be ideal because SSR heterogeneity in individual tumors is a problematic factor for sensitive assessment of treatment response. Moreover, the differences in intrinsic variabilities in SUV\(_{\text{max}}\) in separate PET/CT scanners at different institutions were a limitation for image-based assessment and patient follow-up (54). Changes in tumor SUV\(_{\text{max}}\) during PRRT do not always correlate with the outcome (25, 76) and in tumors with SUV\(_{\text{max}}\) >20–25, SUV does not linearly correlate with SSR expression (77). Other imaging biomarkers, such as activated glucose metabolisms (\(^{18}\text{F-FDG-PET}\)) are now being re-evaluated, and optimism exists regarding their future prognostic role in NEN management although prospective validation is required (17). Although guidelines have, in general, supported serial comparisons between images to evaluate changes in tumors (14, 71), a RECIST approach has not been recommended in neuroendocrine tumor disease. This is consistent with the opinions of the experts at this Delphi consensus who opined that the current configuration of RECIST criteria was suboptimal for application to NET disease assessment. Additional parameters that potentially could be included to improve imaging, however, remained unresolved. The overall consensus was that adjunct biomarker tools should be developed to provide synergistic information with imaging as a means to facilitate the assessment of therapy. It was agreed that a better understanding of tumor biology would unquestionably expedite the development of an appropriate therapeutic biomarker(s). The determination of therapeutic strategy by identification of a biomarker is limited to the assessment of SSR expression before the use of PRRT. The use of current pharmacological therapy is critically limited by the absence of pretreatment biomarker identification and the lack of tools to accurately define efficacy.

Molecular strategies have thus far typically focused on DNA alterations but are clinically non-informative. Mutations in MEN-1, the predominant sporadic NEN mutation (pancreatic NENs), are not associated with differences in SSR expression and detection by SRI (78, 79). Moreover, the clinical usefulness of alterations in ATRX, DAXX, mTOR signaling (40) and YY1 (80) (all principally identified as sporadic mutations in pancreatic NENs) remain to be proven. Furthermore, the prognostic and predictive utility of the recently identified IMPK mutation in a single small bowel carcinoid family (81) remains to be defined. In addition, the clinical usefulness of chemical-based DNA modifications e.g., methylation, requires elucidation. Alternatives to DNA-based molecular strategies included assessment of CTCs, miRNA, metabolomics and transcriptome-based approaches. The panel considered miRNA to have potential utility. Data indicated that tissue-derived microRNAs are detectable in patient serum samples and may be altered by somatostatin analogs (82). Similarly, metabolomics investigations were considered of interest because functional and non-functional tumors are readily separated (\(R^2=0.98\)) (83). More clinical data are necessary to further assess the clinical utility. With respect to CTCs, the consensus was that this parameter remained problematic at the present time. Although there is some literature to support CTCs (84, 85), all represent a single-center study and hence enthusiasm was diminished. Concerns were also raised with regard to technological aspects of the measurement. Analysis of the results demonstrates that the clinical sensitivity (number of patients with detectable CTCs) is low, 33% in the first study and 49% in the second study. Such low numbers may reflect variable EpCAM expression used for tumor cell capture. Irrespective of technical issues, it remains difficult to reconcile the utility of a test that is based on the absence or presence of 1 circulating tumor cell. This opinion directly recapitulated the opinion expressed at the biomarker-focused Delphi consensus meeting (2014) where a separate group of international experts expressed a similar lack of enthusiasm for the clinical utility of circulating tumor cell technology (61). None of these parameters (CTC, miRNA and metabolomics) are currently clinically recommended in guidelines. Overall, blood-based multianalyte transcript analysis (44, 45), with a clinical sensitivity >95% was considered, by the group, to be more sensitive than standard biomarkers and of potential clinical utility. This is concordant with the consensus from the previous Delphi panel (2014) which evaluated the efficacy, metrics and clinical utility of current NET biomarkers (61). Its precise application to guiding therapy was considered to require further evaluation. Current preliminary data (6, 46) were, however, noted to have specifically addressed clinical utility in sporadic, well-differentiated GEP-NETs. A role in familial NETs (including germine MEN-1 and VHL mutations) is currently under evaluation. The efficacy of a molecular tool capable of detecting germine disease evolution over time is of particular clinical relevance given the low accuracy of current biomarkers and the limitations of imagery (sensitivity and radiation exposure) as a life-long monitoring tool (86). The areas of efficacy were identified
as assessment of the effectiveness of curative surgery, assessment of the efficacy of SSA therapy, prediction of disease stability/progression and identification of response to PRRT. The signature was decreased by surgery and values corresponded to the completeness of tumor removal (49). In addition, elevated levels after R0 resection predicted subsequent disease recurrence. In a different study, elevated transcript levels were prognostic of SSA failure/disease progression (51). Of note was the observation that alterations in transcript levels occurred significantly earlier than RECIST- or SRI-based measures of disease progression (51). Finally, levels were prognostic for PRRT efficacy and could be used to evaluate therapy, correlating with image-based assessments (53). The observation that NEN gene blood levels correlated with $^{68}$Ga-DOTAPET/CT imaging and could define disease status was considered worthy of further clinical study (52). In the latter study, a quotient including specific genes and the $SUV_{\text{max}}$ accurately predicted clinical status. Thus, stable disease could be differentiated from progression using a time point amalgam of a single image/blood sample. The group considered that the combination of imaging and circulating blood biomarkers offered a potential for fusing these two functional modalities of treatment assessment into a clinical index of disease status. This novel consideration had not been previously evaluated at the initial Delphi analysis (2014), which developed a biomarker-centric analysis of disease management. The larger and more diverse international cohort of experts that comprised the current Delphi group was designed to assess the effectiveness and facility of the integration of validated imaging strategies as a combinatorial clinical assessment tool with biomarkers.

In conclusion, there was consensus among a large ($n=33$) group of NEN disease experts from diverse medical and scientific disciplines and countries that current

Figure 4

Conceptual proposal for the evaluation of therapeutic efficacy. This provides an integration of functional imaging and tumor molecular biology using circulating multianalyte assays with algorithm analyses (MAAAs), mRNA or miRNA. Disease progress can be delineated using a combination of functional imaging modalities quantifying somatostatin receptor expression (SSR) by $^{68}$Ga-DOTA-SSA PET/CT and tumor metabolism using either $^{18}$F-DOPA PET/CT (in well-differentiated tumors) or $^{18}$F-FDG (mainly in undifferentiated forms or to assess tumor aggressiveness). The MAAA e.g., circulating mRNA, provides an accurate reflection of tumor activity. Overall, the combination of functional imaging ($^{68}$Ga-SSA and $^{18}$F-FDG-PET/CT) and circulating mRNA could, in the future, help to delineate treatment efficacy.
imaging and circulating biomarkers for NEN disease have substantial limitations for predicting disease activity and for measuring therapeutic efficacy. In addition, RECIST remains suboptimal as a metric of disease status, and better tools for assessment and improved techniques for imaging require development. These views broadly recapitulate published guidelines for GEP-NETs (14, 71, 72, 73, 74, 75) while providing a more in depth and detailed evaluation of the strengths and weaknesses of the different strategies and how best they might be integrated to provide synergistic information of clinical utility. It was concluded that a critical requirement was the development of a multianalyte molecular tool that can better identify disease status and define treatment response. In this respect, the use of circulating RNA as a biomarker was confirmed to supersede the effectiveness of standard monoanalyte biomarkers and have potential clinical applicability. This assessment corroborated the outcome of the previous biomarker-centric Delphi consensus meeting (61). Current data suggest added value for the transcript analysis in the monitoring of diverse therapeutic modalities, particularly in conjunction with other parameters to monitor disease progression (Fig. 4). The NEN experts concluded that combinations of imaging and blood-based molecular information provided by transcriptome analysis could offer the most promising future strategy for refining and improving the evaluation of therapy.

Acknowledgements
Clifton Life Sciences provided funding but was not involved in the selection of topics, the choice of experts, the discussion and analysis of the data or the manuscript compilation.

References
Endocrine Connections

K Oberg et al.

NET biomarkers and imaging: a Delphic assessment

185–187

5:185


NET biomarkers and imaging: a Delphic assessment

Oberg K et al. NET biomarkers and imaging: a Delphic assessment
Endocrine Connections

Received in final form 18 August 2016
Accepted 31 August 2016
Accepted Preprint published online 31 August 2016