Reliability and Agreement of Radiological and Pathological Tumor Size in Patients with Multiple Endocrine Neoplasia Type 1-Related Pancreatic Neuroendocrine Tumors: Results from a Population-Based Cohort

Dirk-Jan van Beek[a] Helena M. Verkooijen[b] Sjoerd Nell[a] Bert A. Bonsing[c]
Casper H. van Eijck[d] Harry van Goor[e] Frederik J.H. Hoogwater[f]
Elisabeth J.M. Nieveen van Dijkum[g] Geert Kazemier[h] Cornelis H.C. Dejong[i, j]
Lodewijk A.A. Brosens[k] Frank J. Wessels[l] Inne H.M. Borel Rinkes[a]
Gerlof D. Valk[m] Menno R. Vriens[a]

[a]Department of Endocrine Surgical Oncology, University Medical Center Utrecht, Utrecht, The Netherlands;
[b]Imaging Division, University Medical Center Utrecht, Utrecht, The Netherlands; [c]Department of Surgery, Leiden
University Medical Center, Leiden, The Netherlands; [d]Department of Surgery, Erasmus Medical Center, Rotterdam,
The Netherlands; [e]Department of Surgery, Radboud University Medical Center, Nijmegen, The Netherlands;
[f]Department of Surgery, University of Groningen, University Medical Center Groningen, Groningen,
The Netherlands; [g]Department of Surgery, Cancer Center Amsterdam, Amsterdam University Medical Center
location Academic Medical Center, Amsterdam, The Netherlands; [h]Department of Surgery, Cancer Center
Amsterdam, Amsterdam University Medical Center location VUmc University Medical Center, Amsterdam,
The Netherlands; [i]Department of Surgery, Maastricht University Medical Center, NUTRIM School for Nutrition and
Translational Research in Metabolism, Maastricht, The Netherlands; [j]Department of Surgery, Universitätsklinikum
Aachen, Aachen, Germany; [k]Department of Pathology, University Medical Center Utrecht, Utrecht, The Netherlands;
[l]Department of Radiology, University Medical Center Utrecht, Utrecht, The Netherlands; [m]Department of Endocrine
Oncology, University Medical Center, Utrecht, The Netherlands

Abstract
Background: Pancreatic neuroendocrine tumors (pNETs)
have a high prevalence in patients with multiple endocrine
neoplasia type 1 (MEN1) and are the leading cause of death.
Tumor size is still regarded as the main prognostic factor and
therefore used for surgical decision-making. We assessed re-
liability and agreement of radiological and pathological tu-
mor size in a population-based cohort of patients with
MEN1-related pNETs. ***Methods:*** Patients were selected from
the Dutch MEN1 database if they had undergone a resection
for a pNET between 2003 and 2018. Radiological (MRI, CT,
and endoscopic ultrasonography [EUS]) and pathological tu-
mor size were collected from patient records. Measures of
agreement (Bland-Altman plots with limits of agreement
[LoA] and absolute agreement) and reliability (intraclass cor-

Menno R. Vriens
Department of Endocrine Surgical Oncology, University Medical Center Utrecht
Room G04.228, Heidelberglaan 100
NL–3584 CX Utrecht (The Netherlands)
mvriens@umcutrecht.nl

relation coefficients [ICC] and unweighted kappa) were calculated for continuous and categorized (< or ≥2 cm) pNET size. **Results:** In 73 included patients, the median radiological and pathological tumor sizes measured were 22 (3–160) and 21 (4–200) mm, respectively. Mean bias between radiological and pathological tumor size was −0.2 mm and LoA ranged from −12.9 to 12.6 mm. For the subgroups of MRI, CT, and EUS, LoA of radiological and pathological tumor size ranged from −9.6 to 10.9, −15.9 to 15.8, and −13.9 to 11.0, respectively. ICCs for the overall cohort, MRI, CT, and EUS were 0.80, 0.86, 0.75, and 0.76, respectively. Based on the 2 cm criterion, agreement was 81.5%; hence, 12 patients (18.5%) were classified differently between imaging and pathology. Absolute agreement and kappa values of MRI, CT, and EUS were 88.6, 85.7, and 75.0%, and 0.77, 0.71, and 0.50, respectively. **Conclusion:** Within a population-based cohort, MEN1-related pNET size was not systematically over- or underestimated on preoperative imaging. Based on agreement and reliability measures, MRI is the preferred imaging modality.

Introduction

Duodenopancreatic neuroendocrine tumors (dpNETs) affect over 80% of patients with multiple endocrine neoplasia type 1 (MEN1) by the age of 80 years [1, 2]. The trait occurs in 2–3 per 100,000 people and leads to multiple tumors in endocrine and non-endocrine organs [3]. MEN1-related dpNETs represent 2 distinct groups of tumors regarding hormone production: functioning dpNETs, which excessively produce hormones leading to a distinct clinical syndrome, and non-functioning pancreatic neuroendocrine tumors (NF-pNETs), which do not induce any secretory disorder. Duodenal gastrinomas and pancreatic insulinomas are the most frequently encountered functioning dpNETs [4]. Since dpNETs are the leading cause of death in MEN1, guidelines recommend intensive radiological screening programs starting in childhood to diagnose dpNETs at an early stage, enabling timely initiation of treatment [1, 5–7].

Nowadays, World Health Organization tumor grade and tumor size of NF-pNETs and gastrinomas in patients with MEN1 are still the main predictors of metastases and survival, and treatment indications are primarily based on tumor size [2, 7, 8]. Most NF-pNETs <2 cm have an indolent natural disease course and can be safely managed without surgery [9–12]. European Neuroendocrine Tumor Society (ENETS) guidelines recommend operative resection of pancreatic neuroendocrine tumors (pNETs) and gastrinomas >2 cm [8]. Underestimation of tumor size may lead to a prolonged period of watchful waiting, whereas overestimation may hypothetically lead to unnecessary surgery. Since surgery for MEN1-related NF-pNETs is associated with a high risk of severe complications, the decision to operate should be carefully balanced against the risk of complications [13].

Considering that pNET size is used for individual patient risk stratification and subsequent clinical decision-making, accurate tumor size estimation is important. Since surgical decision-making is currently guided by the 2 cm criterion, adequate measurements of tumor size are of utmost importance. Although previous studies have focused on diagnosing pNETs on conventional imaging, guidelines lack evidence-based recommendations regarding a preferred imaging modality [7, 14–19]. In the debate on the optimal modality, the accuracy of tumor size estimation by MRI, CT, or endoscopic ultrasonography (EUS) should be taken into account.

One single-center study from an ENETS center of excellence investigated the correlation between preoperative radiological tumor size and the pathological tumor size in 44 patients with MEN1-related pNETs and concluded that tumor size is frequently overestimated [20]. Since patients with MEN1 are generally treated in tertiary referral centers by dedicated multidisciplinary teams, data from a population-based cohort are necessary to estimate the outcomes in general MEN1 care [7]. In addition, a large multicenter study makes subgroup analyses for imaging modalities more robust. Therefore, the present study assessed the reproducibility, in terms of reliability and agreement, of radiological and pathological tumor size of resected pNETs in patients with MEN1 from a population-based cohort representing daily clinical practice.

Patients and Methods

Reporting of the study was performed according to the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) recommendations [21].

Study Design and Patient Selection
Patients were selected from the DutchMEN Study Group (DMSG) database [22]. In brief, the database includes MEN1 patients aged 16 years and older and under treatment in one of the 8 University Medical Centers (UMCs) in the Netherlands. In each center, patients were identified by reviewing hospital databases of medical conditions and diseases. MEN1 diagnosis was established according to the clinical practice guidelines [7]. Over 90% of the

Dutch MEN1 population is included in the database [23]. Clinical and demographic data were collected longitudinally every quarter from 1990 to 2017 by standardized medical record review, according to a predefined protocol. From 2016 onward, data were captured prospectively. The protocol was approved by the Medical Ethics Committees of all UMCs.

All consecutive patients who underwent resection of a MEN1-related pNET from 2003 up to and including 2017 with information on preoperative radiological tumor size within 9 months before surgery and histopathological tumor size were identified. By including patients after 2003, the population is more representative for current practice, where patients are screened according to clinical practice guidelines [7, 24]. In addition, considering the screening program, reproducibility analyses were performed in patients with pNETs <5 cm on radiology or pathology.

Clinical Definitions

In the absence of excessive hormone production leading to a clinical syndrome, a pNET was considered as an NF-pNET [9, 10]. Insulinomas were diagnosed based on a 72-h fasting test [9, 25, 26]. Gastrinomas were diagnosed based on hypergastrinemia and a gastrin positive (duodenal) neuroendocrine tumor [27]. In patients with both a pNET and (duodenal) gastrinoma, the resection was considered for an NF-pNET and gastrinoma.

Data Collection

Data were collected from routine patient care, that is, no prospective study protocol existed for radiologists, gastroenterologists, and pathologists to determine tumor size, so size reflected the tumor size used in clinical practice for decision-making. In line, preoperative imaging was generally examined once by 1 local senior radiologist (CT or MRI) or gastroenterologist (EUS) and the surgical specimen was assessed by 1 local senior pathologist. Surgical resection specimens were processed in each center's pathology department according to local practice; tumor size was generally measured after formalin fixation. According to the guidelines, outcomes were discussed in the multidisciplinary tumor boards within the individual tertiary centers ensuring reliable outcomes [7]. Over the years, multiple radiologists, gastroenterologists, and pathologists have performed the observations within the UMCs in the Netherlands; the exact number of observers is unknown. Radiologists, gastroenterologists, and pathologists were not blinded to previous imaging or to clinical information, but at the time of examination, they were not aware of their observations being used for scientific research, so the Hawthorne effect (i.e., an alteration in observers' behavior due to awareness of being observed) is unlikely. At the time of radiological assessment, radiologists were by definition unaware of histopathological tumor size.

MRI, CT, and EUS reports up to 9 months before surgery were collected. Imaging and histopathological reports were reviewed for the location and size of the pNETs. For each patient, the diameter of the largest tumor in the pancreatic head and/or body/tail was obtained from the imaging (MRI, CT, and/or EUS) and pathology reports. For patients with multiple resected tumors (i.e., from both the pancreatic head and body/tail), the size of the largest tumor (either from the pancreatic head or body/tail) on imaging was used for the overall analysis. The tumor size measured on the modality closest to the date of surgery was used for the overall analysis. For patients with multiple modalities, these data were used for subgroup analyses of the respective modality.

Statistical Analysis

Baseline characteristics were described as mean (±standard deviation [SD]), median (range), or counts (percentages), as appropriate. Reproducibility of tumor size between imaging and pathology was assessed in terms of agreement and reliability [28]. Agreement indicates the degree to which measurements are identical and is particularly relevant when assessing the absolute closeness of repeated measures [21, 28]. Reliability implies the ability to distinguish patients with different tumor sizes from each other despite biological variability between study objects [21, 28].

Agreement between imaging and histology of continuous tumor size was assessed using the Bland-Altman plots and their limits of agreement (LoA) with 95% confidence intervals (95% CI) [29]. In brief, for every patient, the difference between the radiological and pathological tumor size (*y* axis) was plotted against the mean of the 2 measurements (*x* axis). The mean and SD of these differences were used to calculate the LoA (mean difference ± 1.96 × SD). The LoA represent the maximum range by which repeated measurements would be expected to differ in 95% of repetitions and indicate the range of observer variation; differences beyond the LoA are not accounted for by observer variation alone [29]. Reliability of tumor size was assessed by the intraclass correlation coefficient (ICC), which was calculated by a 2-way model with absolute agreement and single measures (ICC(2,1)) [30]. The ICC is interpreted as the percentage of variability between the measurements, which is not caused by measurement error. ICC values should generally be at least 0.90 to guide important clinical decisions [21].

Contingency tables were tabulated for pNET size categorized to <2 or ≥2 cm on imaging and pathology, according to current clinical insights [8, 10, 11]. These were additionally performed for <3 and ≥3 cm [10]. Agreement was assessed using percentages of absolute agreement and specific agreement. Reliability was assessed using unweighted Cohen's kappa. Kappa's are rated as "fair" for values between 0.21 and 0.40, "moderate" for 0.41–0.60, "substantial" for 0.61–0.80, and "almost perfect" for values above 0.80 [31]. 95% CIs for reliability measures (kappa and ICC) were generated by drawing 10,000 bias-corrected and accelerated bootstrap replications [32–34].

All analyses were performed for all imaging modalities (MRI/CT/EUS) together and separately. Furthermore, subgroup analyses were performed including patients with an NF-pNET, with or without concurrent gastrinoma, as surgical indication, since tumor size guides clinical management in these patients [8, 10, 11]. To investigate the overall effect of time and the effect of different modalities over time, subgroup analyses were additionally performed for patients undergoing surgery before 2011 (2003–2010) and in 2011 or later (2011–2017). Statistical analyses were conducted using R version 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria) using the "IRR," "blandr," and "boot" packages.

Results

Of the 445 patients in the DMSG database, 275 (61.8%) had a pNET on imaging (Fig. 1). Seventy-three patients underwent surgery for a pNET between 2003 and 2018

Table 1. Baseline table

Variable	N = 73 (%)
Age at surgery, years, mean (±SD)	44.6±14.5
Sex, n (%)	
Male	34 (46.6)
Female	39 (53.4)
Surgery, n (%)	
Primary surgery	64 (87.7)
Reoperation	9 (12.3)
Surgical indication, n (%)	
NF-pNET	47 (64.4)
Insulinoma	14 (19.2)
Gastrinoma	4 (5.5)
NF-pNET and gastrinoma	7 (9.6)
Other functioning pNET	1 (1.4)
Type of resection, n (%)	
Enucleation	12 (16.4)
Enucleation of head	4 (5.5)
Enucleation of body/tail	7 (9.6)
Enucleation of head and body/tail	1 (1.4)
Distal pancreatectomy	35 (47.9)
Whipple/PPPD	7 (9.6)
Distal pancreatectomy and enucleation	3 (4.1)
Whipple/PPPD and distal pancreatectomy	5 (6.8)
Pancreatic body resection	1 (1.4)
Total/completion pancreatectomy	10 (13.7)
Type of imaging available for size analysis, n (%)	
MRI	36 (49.3)
CT	43 (58.9)
EUS	29 (39.7)
Multiple imaging strategies available for size analysis	33 (45.2)
MRI and CT	11 (15.1)
MRI and EUS	7 (9.6)
CT and EUS	13 (17.8)
MRI, CT, and EUS	2 (2.7)
Time from imaging to surgery, n (%)	
0 quarters	18 (24.7)
1 quarter	41 (56.2)
2 quarters	14 (19.2)
Period of surgery, n (%)	
2003–2010	39 (53.4)
2011–2017	34 (46.6)

CT, computed tomography; EUS, endoscopic ultrasonography; MRI, magnetic resonance imaging; N, number of; NF-pNET, non-functioning pancreatic neuroendocrine tumor; pNET, pancreatic neuroendocrine tumor; PPPD, pylorus-preserving pancreatoduodenectomy; SD, standard deviation.
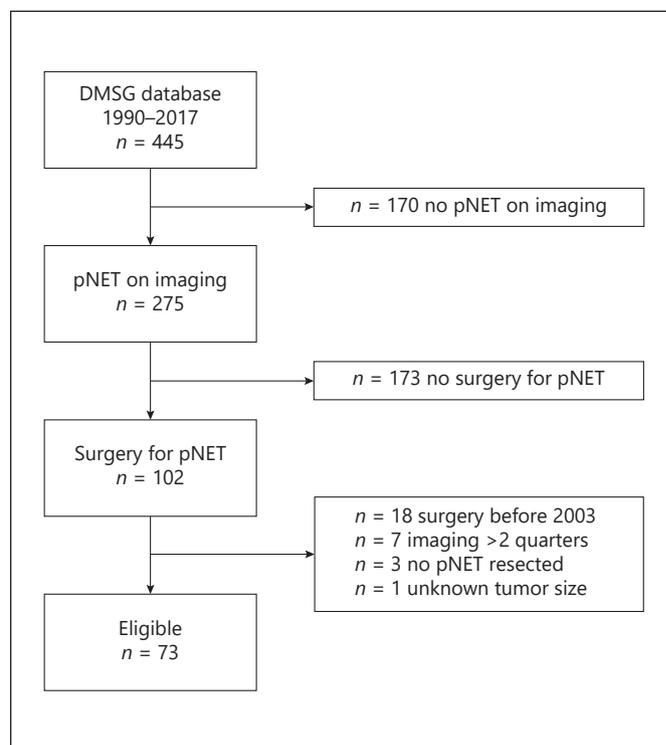

Fig. 1. Flowchart of patient selection. Sixty-five of the 73 patients had a pNET <5 cm. DMSG, DutchMEN Study Group, pNET, pancreatic neuroendocrine tumor.

and were eligible for inclusion; reasons for exclusion are listed in Figure 1. Patients had undergone surgery at a mean age of 44.6 years (±14.5), and 39 (53.4%) were females (Table 1). Median age at diagnosis was 36.6 years (range 18.8–81.8) in patients with an insulinoma and 42.5 years (range 14.5–73.3) in the other patients. An NF-pNET was the most frequent surgical indication in 47 patients (64.4%), and 15 (20.5%) patients underwent surgery for a functioning pNET. Twenty-five patients had a resected pNET of the head, and 57 patients had a resected pNET of the body/tail. Almost half of the patients (47.9%, n = 35) underwent a distal pancreatectomy; 12 (16.4%) underwent 1 or more enucleation(s); 10 (13.7%), a total/completion pancreatectomy; 7 (9.6%), a Whipple/PPPD; and 1 (1.4%), a central pancreatectomy, and in 8 patients (11.0%), a combined resection was performed. Most recent imaging before surgery was MRI in 23 (31.5%), CT in 28 (38.4%), and EUS in 22 (30.1%). Overall, MRI, CT, and EUS were available for analysis in 36 (49.3%), 43 (58.9%), and 29 (39.7%) patients, respectively. Thirty-three patients (45.2%) had multiple imaging modalities for analysis. Thirty-nine patients (53.4%) underwent surgery before 2011 and 34 (46.6%) in 2011 or later. Before 2011, MRI, CT, and EUS were available for 11 (28.2%), 28 (71.8%), and 18 (46.2%) patients, and after 2011, these were available for 25 (73.5%), 15 (44.1%), and 11 (32.4%) patients, respectively.
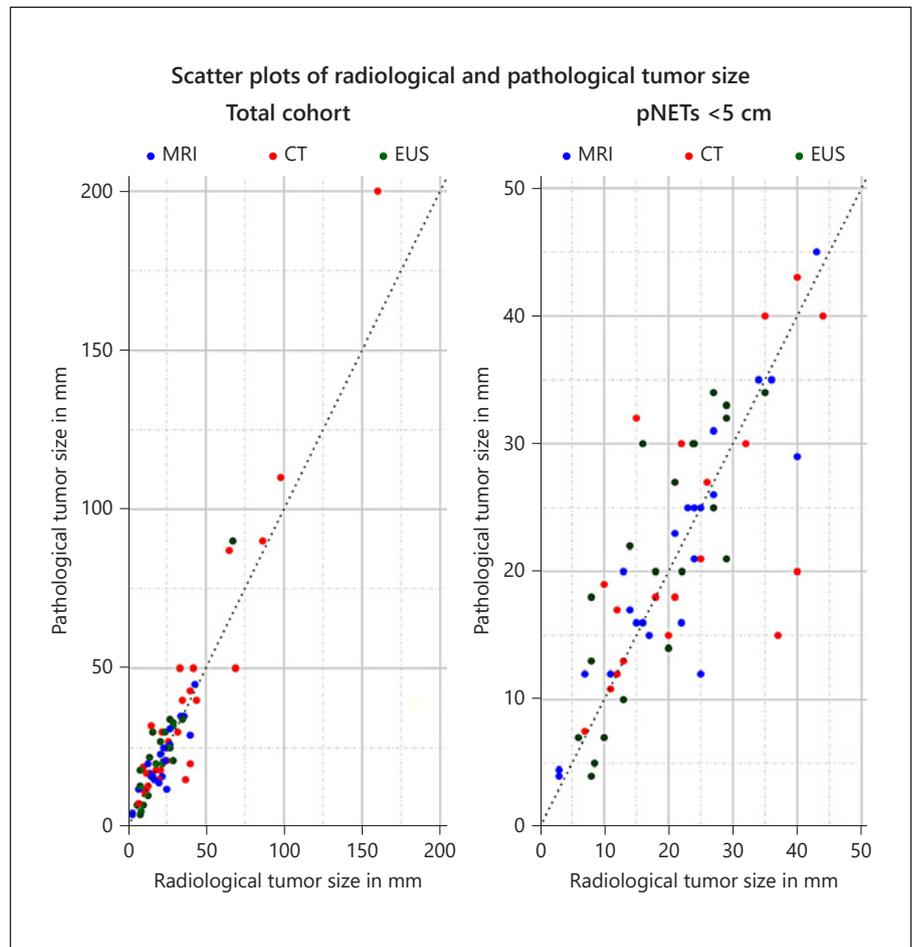
van Beek et al.

Fig. 2. Scatterplots of radiological and pathological tumor size of the total cohort and for pNETs <5 cm. pNET, pancreatic neuroendocrine tumor; EUS, endoscopic ultrasonography.

Radiological and Pathological Tumor Size

Radiological and pathological tumor sizes of the largest tumor of individual patients are presented in Figure 2. The median radiological and pathological tumor sizes were 22 mm (range 3–160) and 21 mm (range 4–200), respectively (Table 2). In 28 patients (38.4%), radiological size was larger than pathological size; in 40 patients (54.8%), radiological size was smaller than pathological size; and in 5 (6.8%), both were exactly similar. Twelve patients (16.4%) had a difference of >10 mm between radiological and pathological size, of whom 5 had a pNET of ≥5 cm. Nine of these 12 patients underwent surgical resection before 2011; all 8 patients with a difference of >15 mm were operated on before 2011.

For the subgroups of MRI, CT, and EUS, radiological size was larger than pathological size in 38.9% (14/36), 37.2% (16/43), and 48.3% (14/29) patients, and smaller than pathological size in 55.6% (20/36), 48.8% (21/43), and 51.7% (15/29), respectively. Radiological and patho-

logical sizes were exactly similar in 2/36 patients (5.6%) with an MRI and 6/14 (14.0%) with a CT.

Median radiological and pathological size of pNETs of the head were 19 mm (range 3–43) and 15 mm (range 4–50), and 22 mm (range 5–160), and 21 mm (range 4–200) of pNETs of the body/tail. Median radiological and pathological sizes of the NF-pNET subgroup (*n* = 54, 74.0%) were 25 mm (range 3–160) and 25 mm (range 4–200), respectively. Seven patients (9.6%) had a pNET > 5 cm on imaging or pathology, and therefore, reproducibility analyses were performed using 65 patients.

Tumor Size

Agreement between radiological and pathological tumor size is shown in Figure 3 and Table 3. Overall, imaging underestimated pathological tumor size with −0.2 mm (95% CI: −1.8 to 1.4). LoA ranged from −12.9 mm (95% CI: −15.6 to −10.1) to 12.6 mm (95% CI: 9.8–15.3), respectively. This indicates that differences in size of up to 13 mm

Table 2. Radiological and pathological tumor size of the total cohort and stratified by imaging modality and pNET localization

Overall					pNETs head					pNETs body/tail				
imaging modality	N (%)	radiological size, mm	pathological size, mm	difference, mm	N (%)	radiological size, mm	pathological size, mm	difference, mm	N (%)	radiological size, mm	pathological size, mm	difference, mm
All	73 (100)	22 [3–160]	21 [4–200]	−1 [−40 to 22]	26 (35.6)	19 [3–43]	15 [4–50]	−1 [−17 to 22]	58 (79.5)	22 [5–160]	21 [4–200]	0 [−40 to 20]
MRI	36 (49.3)	23 [3–43]	21 [4–50]	−1 [−9 to 17]	14 (19.2)	16 [3–43]	14 [4–45]	−1 [−6 to 15]	29 (39.7)	21 [3–41]	20 [4–50]	0 [−15 to 17]
CT	43 (58.9)	25 [8–160]	25 [4–200]	0 [−40 to 22]	14 (19.2)	20 [7–42]	17 [4–50]	0.3 [−23 to 22]	33 (45.2)	25 [9–160]	25 [4–200]	0 [−40 to 20]
EUS	29 (39.7)	20 [6–67]	21 [4–90]	−1 [−23 to 8]	10 (13.7)	19 [5–29]	16 [5–33]	0.8 [−5 to 6]	22 (30.1)	17 [5–67]	22 [4–90]	−1 [−23 to 8]

All values are given as median [range]. The differences were calculated by subtracting the pathological size from the radiological size. CT, computed tomography; EUS, endoscopic ultrasonography; MRI, magnetic resonance imaging; N, number of; pNET, pancreatic neuroendocrine tumor.

Table 3. Bland-Altman analysis showing mean bias and LoA

Imaging modality	Overall				NF-pNETs			
	N	mean bias (95% CI)	LoA (95% CI) lower	upper	N	mean bias (95% CI)	LoA (95% CI) lower	upper
All	65	−0.2 (−1.8 to 1.4)	−12.9 (−15.6 to −10.1)	12.6 (9.8–15.3)	49	0.6 (−1.3 to 2.5)	−12.6 (−16.0 to 9.3)	13.8 (10.5–17.2)
MRI	35	0.6 (−1.2 to 2.4)	−9.6 (−12.7 to −6.5)	10.9 (7.8–14.0)	27	1.5 (−0.7 to 3.7)	−9.4 (−13.2 to −5.6)	12.5 (8.7–16.3)
CT	35	0.0 (−2.7 to 2.8)	−15.9 (−20.7 to 11.1)	15.8 (11.1–20.6)	26	0.1 (−3.6 to 3.8)	−17.8 (−24.2 to 11.4)	18.1 (11.7–24.5)
EUS	28	−1.8 (−3.9 to 1.0)	−13.9 (−18.2 to 9.7)	11.0 (6.7–15.3)	19	−1.1 (−4.2 to 1.9)	−13.5 (−18.8 to −8.2)	11.2 (5.9–16.5)

Imaging modality	2003–2010				2011–2017			
	N	mean bias (95% CI)	LoA (95% CI) lower	upper	N	mean bias (95% CI)	LoA (95% CI) lower	upper
Overall	32	−0.2 (−3.0 to 2.6)	−15.5 (−20.3 to −10.6)	15.1 (10.2–19.9)	33	−0.1 (−1.9 to 1.7)	−10.0 (−13.0 to −6.9)	9.7 (6.6–12.8)
MRI	10	−1.9 (−5.6 to 1.8)	−12.2 (−18.8 to −5.5)	8.4 (1.7–15.0)	25	0.7 (−1.4 to 2.9)	−9.6 (−13.3 to −5.8)	11.0 (7.3–14.8)
CT	21	1.0 (−2.8 to 4.9)	−15.4 (−22.1 to −8.8)	17.5 (10.9–24.1)	14	−1.0 (−3.8 to 1.9)	−10.5 (−15.4 to −5.6)	8.6 (3.7–13.5)
EUS	17	−2.7 (−6.2 to 0.9)	−16.2 (−22.4 to −10.0)	10.9 (4.7–17.1)	11	0.4 (−3.2 to 4.0)	−10.1 (−16.5 to −3.8)	10.9 (4.6–17.3)

CI, confidence interval; CT, computed tomography; EUS, endoscopic ultrasonography; LoA, limits of agreement; MRI, magnetic resonance imaging; N, number of; NF-pNET, non-functioning pancreatic neuroendocrine tumor.
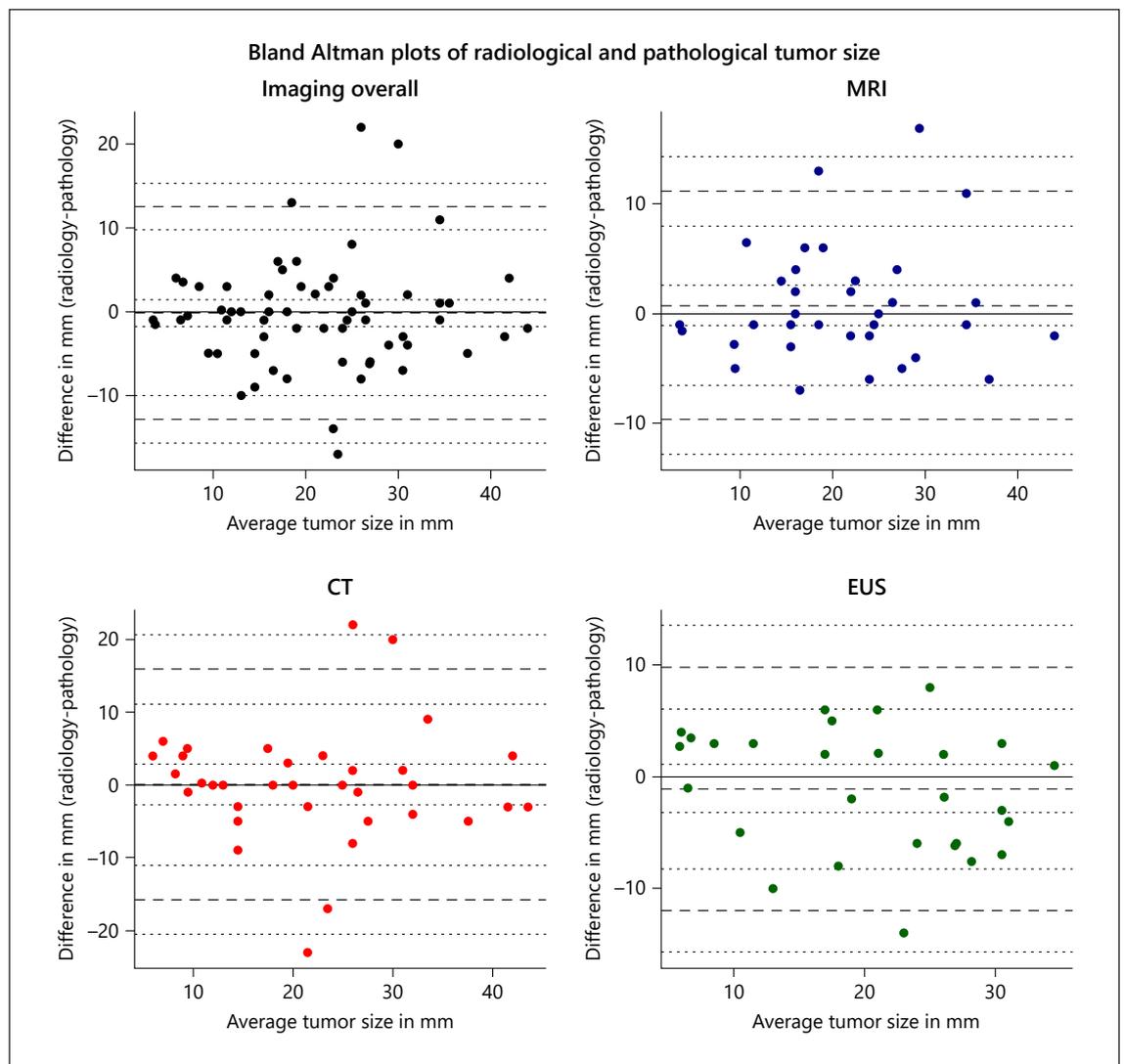
Fig. 3. Bland-Altman plots of radiological and pathological tumor size of MEN1-related pNETs. For every patient, the difference between the radiological and pathological tumor size (*y* axis) is plotted against the mean of the 2 measurements (*x* axis). The mean and standard deviation (SD) of these differences are used to calculate the limits of agreement (LoA) (mean difference ±1.96 × SD). The LoA indicate the range of observer variation; differences beyond the LoA are not accounted for by observer variation alone [29].

The solid line shows a mean bias of 0. The middle dashed line represents the mean difference between radiological and pathological tumor size, with dotted lines indicating 95% confidence intervals (95% CI) of the mean bias. The upper and lower dashed lines represent the upper and lower LoA with subsequent upper and lower bounds of the 95% CI of the LoA shown in dotted lines. EUS, endoscopic ultrasonography.

between imaging and pathology can be contributed to observer or measurement error, so a tumor measured on imaging or pathology can be up to 13 mm smaller or larger. No systematic over- or underestimation of radiological tumor size was observed for any of the modalities. LoA of MRI and EUS were generally smaller than those of CT (Fig. 3; Table 3). For MRI, LoA ranged from −9.6 mm (95% CI: −12.7 to −6.5) to 10.9 mm (95% CI: 7.8–14.0), and for

EUS, from −13.9 mm (95% CI: −18.2 to −9.7) to 11.0 mm (95% CI: 6.7–15.3). LoA did not increase with increasing tumor size (Fig. 3). Within the NF-pNET subgroup, similar results were observed (Table 3). For pNETs in the pancreatic head, mean bias was 0.9 mm (95% CI: −1.4 to −3.1) and LoA ranged from −9.6 mm (95% CI: −13.5 to −5.8) to 11.4 mm (95% CI 7.5–15.2). For pancreatic body/tail tumors, mean bias, lower LoA, and upper LoA were −0.3 mm

Table 4. Reliability of tumor size and pNETs categorized as < or ≥2 cm

Type	Overall cohort		NF-pNET subgroup	
	pNET ≥2 cm kappa (95% CI)*	tumor size ICC (95% CI)*	pNET ≥2 cm kappa (95% CI)*	tumor size ICC (95% CI)*
All	0.63 (0.40–0.78)	0.80 (0.65–0.89)	0.55 (0.27–0.77)	0.79 (0.58–0.90)
MRI	0.77 (0.46–0.94)	0.86 (0.69–0.94)	0.68 (0.29–0.92)	0.84 (0.62–0.94)
CT	0.71 (0.40–0.89)	0.75 (0.48–0.89)	0.67 (0.28–0.92)	0.70 (0.35–0.89)
EUS	0.50 (0.08–0.77)	0.76 (0.50–0.88)	0.36 (−0.16 to 0.77)	0.76 (0.28–0.91)
	2003–2010		**2011–2017**	
Type	pNET ≥2 cm kappa (95% CI)*	tumor size ICC (95% CI)*	pNET ≥2 cm kappa (95% CI)*	tumor size ICC (95% CI)*
All	0.55 (0.19–0.80)	0.72 (0.43–0.89)	0.70 (0.38–0.88)	0.89 (0.78–0.95)
MRI	1.00 (NA)	0.93 (0.71–0.99)	0.68 (0.33–0.92)	0.84 (0.64–0.94)
CT	0.62 (0.17–0.90)	0.69 (0.28–0.92)	0.86 (0.34–1.00)	0.82 (0.25–0.96)
EUS	0.29 (−0.21 to 0.66)	0.67 (0.25–0.87)	0.81 (NA)	0.91 (0.74–0.96)

CI, confidence interval; CT, computed tomography; EUS, endoscopic ultrasonography; ICC, intraclass correlation coefficient; MRI, magnetic resonance imaging; NF-pNET, non-functioning pancreatic neuroendocrine tumor; pNET, pancreatic neuroendocrine tumor. * 95% CIs were generated by drawing 10,000 bias-corrected and accelerated bootstrap replications.

Table 5. Agreement of pNETs categorized as < or ≥2 cm on radiology and pathology

Type	Overall cohort				NF-pNET subgroup			
	N	absolute agreement, %	specific agreement, %		N	absolute agreement, %	specific agreement, %	
			negative	positive			negative	positive
All	65	81.5	79.3	83.3	49	79.6	70.6	84.4
MRI	35	88.6	86.7	90.0	27	85.2	80.0	88.2
CT	35	85.7	83.9	87.2	26	84.6	80.0	87.5
EUS	28	75.0	72.0	77.4	19	73.7	54.5	81.5
Type	**2003–2010**				**2011–2017**			
	N	absolute agreement, %	specific agreement, %		N	absolute agreement, %	specific agreement, %	
			negative	positive			negative	positive
All	32	78.1	74.1	81.0	33	84.8	83.9	85.7
MRI	10	100.0	100.0	100.0	25	84.0	83.3	84.6
CT	21	81.0	77.8	83.3	14	92.9	92.3	93.3
EUS	17	64.7	62.5	66.7	11	90.9	88.9	92.3

CT, computed tomography; EUS, endoscopic ultrasonography; MRI, magnetic resonance imaging; NF-pNET, non-functioning pancreatic neuroendocrine tumor; pNET, pancreatic neuroendocrine tumor.

(95% CI: −2.1 to 1.5), −13.0 mm (95% CI: −16.2 to −9.9), and 12.4 mm (95% CI: 9.3–15.6), respectively. LoA of MRI and EUS seemed narrower for pancreatic head tumors compared to those in the body/tail, whereas patients with a CT had narrower LoA for pancreatic body/tail tumors, but these subgroups were small.

LoA of patients who underwent surgery before 2011 and after 2011 are shown in Table 3. Before 2011, LoA

Table 6. Two by two tables of imaging and pathology ≥2 cm

Imaging overall versus pathology	PA <2 cm	PA ≥2 cm	MRI versus pathology	PA <2 cm	PA ≥2 cm
Imaging <2 cm	23 (35.4)	5 (7.7)	MRI <2 cm	13 (37.1)	1 (2.9)
Imaging ≥2 cm	7 (10.8)	30 (46.1)	MRI ≥2 cm	3 (8.6)	18 (51.4)
CT versus pathology	PA <2 cm	PA ≥2 cm	EUS versus pathology	PA <2 cm	PA ≥2 cm
CT <2 cm	13 (37.1)	2 (5.7)	EUS <2 cm	9 (32.1)	4 (14.3)
CT ≥2 cm	3 (8.6)	17 (48.6)	EUS ≥2 cm	3 (10.7)	12 (42.9)

Contingency tables of pNETs categorized as < or ≥2 cm. CT, computed tomography; EUS, endoscopic ultrasonography; MRI, magnetic resonance imaging; PA, pathology; pNET, pancreatic neuroendocrine tumor.

ranged from −15.5 mm (95% CI: −20.3 to −10.6) to 15.1 mm (95% CI 10.2–19.9), and after 2011, ranged from −10.0 mm (95% CI: −13.0 to −6.9) to 9.7 mm (95% CI 6.6–12.8).

The overall ICC for tumor size was 0.80 (95% CI: 0.65–0.89), indicating moderate reliability of continuous tumor size measurement between imaging and pathology (Table 4). The ICCs of MRI, CT, and EUS were 0.86 (95% CI: 0.69–0.94), 0.75 (95% CI: 0.48–0.89), and 0.76 (95% CI: 0.50–0.88), respectively. These ICCs are probably due to the relatively wide range of observed radiological and pathological tumor sizes and therefore can adequately distinguish between patients.

Clinical Implications

Forty-two patients (64.6%) were classified as having a pNET ≥2 cm by imaging or pathology. For pNETs categorized as < and ≥2 cm, agreement was 81.5%, indicating that imaging and pathology classified patients differently in 18.5% (n = 12/65) of the patients (Table 5). Of these patients, 7 (10.8%) were operated on with a radiological pNET ≥2 cm but had a pNET <2 cm according to pathology (Table 6). On the contrary, 5 patients (7.7%) underwent surgery for a histopathological pNET ≥2 cm, while they would have been refrained from surgery based on preoperative imaging. Agreement between MRI, CT, and EUS, and pathology was 88.6, 85.7, and 75.0%, respectively. In terms of percentage, MRI and CT would lead least often to under- or overtreatment. Agreement was similar for the patients with NF-pNETs, except for the negative agreement of EUS. Before and after 2011, agreement was 78.1 and 84.8%, respectively. In the subgroup of patients with a preoperative MRI and EUS (n = 8), MRI and pathology were concordant in 8/8 (100%), whereas EUS and pathology were concordant in 6/8 (75.0%). In the subgroup of patients with an NF-pNET, 12 (24.5%) were classified as radiology and pathology <2 cm, 4 (8.2%) as radiology <2 cm and pathology ≥2 cm, 6 (12.2%) as radiology ≥2 cm and pathology <2 cm, and 27 (55.1%) as radiology and pathology ≥2 cm.

For pNETs categorized as < and ≥3 cm, agreement was 81.5%. Radiology was considered as ≥3 cm and pathology as <3 cm in 3 patients (6.3%) and radiology <3 cm and pathology ≥3 cm in 9 patients (16.7%).

Reliability was substantial, with a kappa of 0.63 (95% CI: 0.40–0.78), indicating that 63% extra agreement was observed beyond the chance agreement (Table 4). Stratified for each modality, the kappa values ranged from moderate to substantial (EUS 0.50 [95% CI: 0.11–0.78], CT 0.71 [95% CI: 0.48–0.89], and MRI 0.77 [95% CI: 0.47–0.94]). Within NF-pNETs, reliability was slightly lower, but still substantial for MRI and CT, whereas it was considered fair for EUS (Table 4).

Discussion

The present study shows that preoperative tumor size of MEN1-related pNETs was not systematically over- or underestimated in a population-based cohort, reflecting daily clinical practice. Mean bias was −0.2 mm and LoA ranged from −12.9 to 12.6 mm, indicating that 95% of the repetitions; the difference between radiological and pathological size will be less than 13 mm. Radiology and pathology were in agreement for a pNET of 2 cm or larger in 81.5%. Seven patients underwent surgery, while the tumor was <2 cm, whereas 5 would be refrained from an operation while they would be considered potential surgical candidates according to current insights. MRI appeared to be superior to other modalities, but this was not

formally tested. Furthermore, the data indicate that agreement and reliability have increased in more recent years.

An overestimation of MEN1-related pNET size on imaging was previously reported [20]. In that study, pathological size was 21 versus 13 mm in the present study, while radiological size was similar [20]. Multiple differences exist between both studies. First, Polenta et al. [20] measured histopathological size from fresh specimens, whereas the histopathological report, including formalin fixation, was leading in the current study. Although no data are available regarding effect of formalin fixation for pNETs, others have assessed this in breast, lung, and renal tumors [35–37]. A decrease in tumor size after formalin fixation was reported in 4% (breast cancer) and 46.8% (lung cancer), respectively [35, 36]. Park et al. [36] observed a decrease of 4.06% and 0.66 mm between fresh and formalin-fixed specimens, which is similar to the 4.6% reported in renal tumors by Tran et al. [37]. Therefore, it is unlikely that formalin fixation has contributed to the observed differences, especially considering that major shrinkage will occur during the processing after formalin fixation [35–38]. Second, data were collected from daily clinical practice, indicating that no measurement criteria existed. In addition, the transverse diameter, which was used to determine radiological and pathological size by Polenta et al. [20], is not necessarily the largest diameter used for clinical decision-making, and determining the transverse diameter could be challenging in histopathological examination. Third, patients were included from population-based cohort, thereby increasing the sample size (73 vs. 44) and making subgroup analyses more robust. MRI and CT data were available in a larger proportion of the present cohort, 36 (49.3%) and 41 (58.9%) patients, respectively, compared to 18 (40.9%) and 16 (36.4%) patients in the study by Polenta et al. [20], especially EUS overestimated tumor size – which tended to be similar in this cohort – while the differences between MRI and pathology were smallest. Last, data were analyzed according to current standards for reproducibility research, which are endorsed by guidelines for reproducibility research [29].

Another single-center study has investigated MRI and CT measurements with pathology in 292 patients with sporadic pNETs [39]. Tumors were larger on imaging and pathology (26 and 25 mm), reflecting differences between sporadic and MEN1-related pNETs. The mean bias of MRI and CT was 0.2 and 1 mm, which is similar to that of the present analysis. Nevertheless, compared to our study, LoA were wider −15.4 to 15.8 and −15.7 to 17.7 [39]. Agreement coefficients are population specific, and differences could be partially caused by the sporadic

pNETs, since size estimation might be more difficult in very large (symptomatic) tumors. In line, we also observed large differences in some patients with pNETs larger than 5 cm. These results underscore that preoperative imaging in MEN1 gives reliable estimates of tumor size. EUS was not studied in the sporadic pNETs, but EUS data are demanded considering its frequent use in patients with MEN1.

The relation between tumor size and risk of metastases was described in 2006 and has been translated into the 2 cm cutoff for (NF-)pNET surgical decision-making [2, 10, 11]. In 12 patients (18.5%), imaging and pathology were discordant regarding the 2 cm criterion. Similar results were observed for a potential 3 cm cutoff. Radiology >2 cm and pathology <2 cm were observed in 5/27 patients (19%) with an NF-pNET by Polenta et al. [20], compared to 6/49 patients (12.2%) with an NF-pNET in the present cohort. Other studies investigated the concordance between MRI and EUS for pNETs ≥2 cm. Daskalakis et al. [40] observed a concordance of 97% in 31 patients with a kappa of 0.912. Barbe et al. [14] identified 20 patients with a pNET ≥2 cm on EUS or MRI, but EUS classified 7 (35%) as <2 cm and MRI only 3 (15%). In the current study, 8 patients had an MRI and EUS, showing concordance between MRI and pathology in all cases and between EUS and pathology in 6 out of 8. Hence, EUS seems to classify pNETs as <2 cm more frequently than MRI. Given the variability in tumor size in some individual patients, relying solely on tumor size as indicator for operative resection will be outdated in the near future. JunD and CHES1, interacting domains of Menin, have been reported as risk factors in MEN1-related pNETs, but those have not been validated [9, 12, 41, 42]. Recent insight in the existence of alpha and beta subtypes of MEN1-related pNETs, which determine prognosis, might cause a paradigm shift in patient selection for surgery [43, 44]. Subtype-specific cutoffs could be established, which would underscore the need for accurate tumor size estimation by using the most reliable modality.

In the debate on the most suitable imaging for MEN1-related pNETs, tumor size estimation should be considered. Although CT is widely available and frequently used for anatomical planning preoperatively, repeated exposure to ionizing radiation makes this modality less suitable for the lifelong screening for (NF)-pNETs in MEN1 [45, 46]. LoA were relatively wide, showing that CT is less accurate for tumor size estimation. For adequate size measurements on CT, a contrast difference is demanded to adequately distinguish the tumor from surrounding (unaffected) pancreatic parenchyma, which depends on

the (hyper)vascularity of a pNET. This is even more challenging for iso- and hypovascular pNETs and MRI might overcome these limitations. After 2011, LoA of CT were more similar to those of MRI, and moreover, agreement and reliability of pNETs categorized as 2 cm or larger between CT and pathology were comparable to MRI. CT and MRI visualize surrounding organs including peripancreatic lymph nodes and the liver, which is important for pNET staging, but MRI is superior in the detection of (p)NET liver metastases, and diffusion-weighted MRI increases the sensitivity and specificity over conventional MRI [47–50]. MRI and EUS are generally recommended for the detection of pNETs [14, 46]. Both MRI and pathology classified pNETs more often as 2 cm or larger than EUS and pathology. In addition, EUS is operator dependent and detects many very small NF-pNETs that are not of clinical significance but often misses clinically relevant pNETs in the pancreatic tail and has a complication risk [9, 14, 46]. On the other hand, EUS offers the possibility of concurrent biopsies of suspected pNETs or lymph nodes with subsequent grading of tumors and growth can be assessed [51, 52]. Nevertheless, the need for diagnostic biopsies is uncertain in these patients, considering the very high a priori chance [53]. Notwithstanding, a cost reduction of 0–67% for the diagnosis of MEN1-related pNETs by MRI over EUS was recently reported [40]. The latter should be investigated in a population-based cohort of MEN1 patients with pNETs, taking all costs into account.

The major strengths of the present study are the patient inclusion from a population-based cohort including multiple MEN1 centers and the data collection from routine care. Thus, reliability and agreement were investigated in daily clinical practice [21]. The study also has several limitations. Only patients undergoing surgery were included. Since the decision to perform surgery has been center dependent, small(er) tumors were included, thereby increasing the generalizability of the results [10]. Size was only measured once, and the workup was center- and physician-dependent, so imaging might have been chosen selectively [7]. No standardized protocol for measurements of size was available; in clinical practice, size measurements on MRI or CT are generally performed in the axial plane and acquisition of MRI sequences is often performed in the axial plane. A potential predominance of transverse measurements – instead of 3D size measurements – is therefore a limitation of the study. However, the latter probably has had no influence on the observed outcomes, since no systematic over- or underestimation of radiological size was noted. Nevertheless, future studies should aim to identify the optimal measurement methodology – by using radiological acquisitions, which are suitable for 3D measurements, and taking intra- and interobserver variation into account – to prospectively establish a protocol for size measurements of pNETs which will improve patient care. Numbers of patients were too small to investigate the effect of center on agreement and reliability. The time from imaging until surgery differed between patients, but still the growth rate of NF-pNETs in MEN1 is generally low [9, 46]. Due to the rarity of MEN1, patients were included over a relatively long time, so the quality, experience, and techniques of CT, MRI and EUS might have increased. Subgroup analyses, based on the period of surgery, showed that the LoA were narrower in patients operated on in 2011 and later compared to those operated on before 2011. In addition, all 8 patients with a difference of 15 mm or more between radiological and pathological size were operated on before 2011.

In conclusion, preoperative tumor size is not systematically over- or underestimated. Nevertheless, in 18.5% of patients, radiology and pathology would classify patients differently regarding the 2 cm cutoff, having direct clinical implications. Agreement and reliability of MRI seemed to be superior to CT and EUS for estimating tumor size. In the debate on the preferred imaging modality for radiological pancreas screening programs in MEN1, combining the diagnostic performance, several other advantages, and reproducibility of radiological size, we would suggest performing MRI.

Acknowledgements

Statement of Ethics

The protocol was approved by the Medical Ethical Committees of all University Medical Centers in the Netherlands. Patients have provided written informed consent for inclusion in the database.

Conflict of Interest Statement

There is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding Sources

Author Contributions

D.-J.B.: study design, acquisition of data, statistical analysis, and interpretation of data, and drafting and final approval of the manuscript. H.M.V.: study design, interpretation of data, critical revision, and final approval of the manuscript. S.N.: study design, acquisition of data, interpretation of data, and critical revision and final approval of the manuscript. B.A.B: study design and critical revision and final approval of the manuscript. C.H.E.: study design and critical revision and final approval of the manuscript. H.G.: study design and critical revision and final approval of the manuscript. F.J.H.H.: study design and critical revision and final approval of the manuscript. E.J.M.N.D.: study design and critical revision and final approval of the manuscript. G.K.: study design and critical revision and final approval of the manuscript. C.H.C.D.: study design and critical revision and final approval of the manuscript. L.A.A.B.: study design and critical revision and final approval of the manuscript. F.J.W.: study design and critical revision and final approval of the manuscript. I.H.M.B.R.: study design and critical revision and final approval of the manuscript. G.D.V.: study design, interpretation of data, critical revision and final approval of the manuscript, and study supervision. M.R.V.: study design, interpretation of data, critical revision and final approval of the manuscript, and study supervision.

References

1 de Laat JM, van der Luijt RB, Pieterman CR, Oostveen MP, Hermus AR, Dekkers OM, et al. MEN1 redefined, a clinical comparison of mutation-positive and mutation-negative patients. BMC Med. 2016 Nov;14(1):182.

2 Triponez F, Dosseh D, Goudet P, Cougard P, Bauters C, Murat A, et al. Epidemiology data on 108 MEN 1 patients from the GTE with isolated nonfunctioning tumors of the pancreas. Ann Surg. 2006;243(2):265–72.

3 Chandrasekharappa SC, Guru SC, Manickam P, Olufemi SE, Collins FS, Emmert-Buck MR, et al. Positional cloning of the gene for multiple endocrine neoplasia-type 1. Science. 1997 Apr [cited 2016 Apr 11];276(5311): 404–7.

4 Pieterman CR, Conemans EB, Dreijerink KM, De Laat JM, Timmers HT, Vriens MR, et al. Thoracic and duodenopancreatic neuroendocrine tumors in multiple endocrine neoplasia type 1: natural history and function of menin in tumorigenesis. Endocr Relat Cancer. 2014;21(3):R121–42.

5 Ito T, Igarashi H, Uehara H, Berna MJ, Jensen RT. Causes of death and prognostic factors in multiple endocrine neoplasia type 1: a prospective study: comparison of 106 MEN1/Zollinger-Ellison syndrome patients with 1613 literature MEN1 patients with or without pancreatic endocrine tumors. Medicine. 2013 May;92(3):135–81.

6 Goudet P, Murat A, Binquet C, Cardot-Bauters C, Costa A, Ruszniewski P, et al. Risk factors and causes of death in MEN1 disease. A GTE (Groupe d'etude des tumeurs endocrines) cohort study among 758 patients. World J Surg. 2010 Feb;34(2):249–55.

7 Thakker RV, Newey PJ, Walls GV, Bilezikian J, Dralle H, Ebeling PR, et al. Clinical practice guidelines for multiple endocrine neoplasia type 1 (MEN1). J Clin Endocrinol Metab. 2012;97(9):2990–3011.

8 Falconi M, Eriksson B, Kaltsas G, Bartsch DK, Capdevila J, Caplin M, et al. ENETS consensus guidelines update for the management of patients with functional pancreatic neuroendocrine tumors and non-functional pancreatic neuroendocrine tumors. Neuroendocrinology. 2016;103(2):153–71.

9 Pieterman CRC, de Laat JM, Twisk JWR, van Leeuwaarde RS, de Herder WW, Dreijerink KMA, et al. Long-term natural course of small nonfunctional pancreatic neuroendocrine tumors in MEN1-results from the Dutch MEN1 study group. J Clin Endocrinol Metab. 2017; 102(10):3795–805.

10 Nell S, Verkooijen HM, Pieterman CRC, De Herder WW, Hermus AR, Dekkers OM, et al. Management of MEN1 related nonfunctioning pancreatic NETs: a shifting paradigm: results from the DutchMEN1 study group. Ann Surg. 2018;267(6):1155–60.

11 Triponez F, Sadowski SM, Pattou F, Cardot-Bauters C, Mirallié E, Le Bras M, et al. Long-term follow-up of MEN1 patients who do not have initial surgery for small ≤2 cm nonfunctioning pancreatic neuroendocrine tumors, an AFCE and GTE study. Ann Surg. 2018; 268(1):158–64.

12 Vinault S, Mariet AS, Le Bras M, Mirallié E, Cardot-Bauters C, Pattou F, et al. Metastatic potential and survival of duodenal and pancreatic tumors in multiple endocrine neoplasia type 1: a GTE and AFCE cohort study (Groupe d'étude des tumeurs endocrines and association francophone de chirurgie endocrinienne). Ann Surg. 2018 Dec:1.

13 Nell S, Borel Rinkes IHM, Verkooijen HM, Bonsing BA, van Eijck CH, van Goor H, et al. Early and late complications after surgery for MEN1-related nonfunctioning pancreatic neuroendocrine tumors. Ann Surg. 2018 Nov; 267(2):352–6.

14 Barbe C, Murat A, Dupas B, Ruszniewski P, Tabarin A, Vullierme MP, et al. Magnetic resonance imaging versus endoscopic ultrasonography for the detection of pancreatic tumours in multiple endocrine neoplasia type 1. Dig Liver Dis. 2012;44(3):228–34.

15 Lewis MA, Thompson GB, Young WF. Preoperative assessment of the pancreas in multiple endocrine neoplasia type 1. World J Surg. 2012;36(6):1375–81.

16 van Asselt SJ, Brouwers AH, van Dullemen HM, van der Jagt EJ, Bongaerts AH, Kema IP, et al. EUS is superior for detection of pancreatic lesions compared with standard imaging in patients with multiple endocrine neoplasia type 1. Gastrointest Endosc. 2015 Jan;81(1):159–e2.

17 Waldmann J, Fendrich V, Habbe N, Bartsch DK, Slater EP, Kann PH, et al. Screening of patients with multiple endocrine neoplasia type 1 (MEN-1): a critical analysis of its value. World J Surg. 2009;33(6):1208–18.

18 Gauger PG, Scheiman JM, Wamsteker EJ, Richards ML, Doherty GM, Thompson NW. Role of endoscopic ultrasonography in screening and treatment of pancreatic endocrine tumours in asymptomatic patients with multiple endocrine neoplasia type 1. Br J Surg. 2003;90(6):748–54.

19 Albers MB, Librizzi D, Lopez CL, Manoharan J, Apitzsch JC, Slater EP, et al. Limited value of Ga-68-DOTATOC-PET-CT in routine screening of patients with multiple endocrine neoplasia type 1. World J Surg. 2017;41(6):1521–7.

20 Polenta V, Slater EP, Kann PH, Albers MB, Manoharan J, Ramaswamy a, et al. Preoperative imaging overestimates the tumor size in pancreatic neuroendocrine neoplasms associated with multiple Endocrine neoplasia type 1. World J Surg. 2018 May;42(5):1440–47.

21 Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. Int J Nurs Stud. 2011;48(6):661–71.

22 van Beek D-J, van Leeuwaarde RS, Pieterman CR, Vriens MR, Valk GD. "Quality in, quality out", a stepwise approach to EBM for rare diseases promoted by MEN1. Endocr Connect. 2018;7(11):260–74.

23 de Laat JM, Tham E, Pieterman CR, Vriens MR, Dorresteijn JA, Bots ML, et al. Predicting the risk of multiple endocrine neoplasia type 1 for patients with commonly occurring endocrine tumors. Eur J Endocrinol. 2012 Aug;167(2):181–7.

24 Brandi ML, Gagel RF, Angeli a, Bilezikian JP, Beck-Peccoz P, Bordi C, et al. Guidelines for diagnosis and therapy of MEN type 1 and type 2. J Clin Endocrinol Metab. 2001 Dec;86(12):5658–71.

25 van Beek DJ, Nell S, Verkooijen HM, Borel Rinkes IHM, Valk GD, Vriens MR (on behalf of the DutchMEN study group); International MEN1 Insulinoma Study Group. Surgery for multiple endocrine neoplasia type 1-related insulinoma: long-term outcomes in a large international cohort. Br J Surg. 2020;107(11):1489–99. 10.1002/bjs.11632. 32352164.

26 Cryer PE, Axelrod L, Grossman AB, Heller SR, Montori VM, Seaquist ER, et al., Endocrine Society. Evaluation and management of adult hypoglycemic disorders: an Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab. 2009;94(3):709–28.

27 van Beek DJ, Nell S, Pieterman CRC, de Herder WW, van de Ven AC, Dekkers OM, et al. Prognostic factors and survival in MEN1 patients with gastrinomas: results from the DutchMEN study group (DMSG). J Surg Oncol. 2019;120(6):966.

28 de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol. 2006;59(10):1033–9.

29 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;327:307–10.

30 Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979;86(2):420–8.

31 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159.

32 Lee J, Fung KP. Confidence interval of the kappa coefficient by bootstrap resampling. Psychiatry Res. 1993;49(1):97–8.

33 Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Acce Stat Sci. 1986;1(1):54–75.

34 Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. Stat Med. 2000;19(9):1141–64.

35 Pritt B, Tessitore JJ, Weaver DL, Blaszyk H. The effect of tissue fixation and processing on breast cancer size. Hum Pathol. 2005;36(7):756–60.

36 Park HS, Lee S, Haam S, Lee GD. Effect of formalin fixation and tumour size in small-sized non-small-cell lung cancer: a prospective, single-centre study. Histopathology. 2017;71(3):437–45.

37 Tran T, Sundaram CP, Bahler CD, Eble JN, Grignon DJ, Monn MF, et al. Correcting the shrinkage effects of formalin fixation and tissue processing for renal tumors: toward standardization of pathological reporting of tumor size. J Cancer. 2015;6(8):759–66.

38 Fox CH, Johnson FB, Whiting J, Roller PP. Formaldehyde fixation. J Histochem Cytochem. 1985;33(8):845–53.

39 Paiella S, Impellizzeri H, Zanolin E, Marchegiani G, Miotto M, Malpaga A, et al. Comparison of imaging-based and pathological dimensions in pancreatic neuroendocrine tumors. World J Gastroenterol. 2017;23(17):3092–8.

40 Daskalakis K, Tsoli M, Alexandraki KI, Angelousi A, Chatzellis E, Tsolakis AV, et al. Magnetic resonance imaging or endoscopic ultrasonography for detection and surveillance of pancreatic neuroendocrine neoplasms in patients with multiple endocrine neoplasia type 1? Horm Metab Res. 2019;51(9):580–5.

41 Thevenon J, Bourredjem A, Faivre L, Cardot-bauters C, Calender A, Murat A, et al. Higher risk of death among MEN1 patients with mutations in the junD interacting domain: a groupe d'étude des tumeurs endocrines (GTE) cohort study. Hum Mol Genet. 2013;22(10):1940–8.

42 Bartsch DK, Langer P, Wild A, Schilling T, Celik I, Rothmund M, et al. Pancreaticoduodenal endocrine tumors in multiple endocrine neoplasia type 1: surgery or surveillance? Surgery. 2000;128(6):958–66.

43 Cejas P, Drier Y, Dreijerink KMA, Brosens LAA, Deshpande V, Epstein CB, et al. Enhancer signatures stratify and predict outcomes of non-functional pancreatic neuroendocrine tumors. Nat Med. 2019;25(8):1260–5.

44 Chan CS, Laddha SV, Lewis PW, Koletsky MS, Robzyk K, Da Silva E, et al. ATRX, DAXX or MEN1 mutant pancreatic neuroendocrine tumors are a distinct alpha-cell signature subgroup. Nat Commun. 2018;9(1):4158–10.

45 Casey RT, Saunders D, Challis BG, Pitfield D, Cheow H, Shaw A, et al. Radiological surveillance in multiple endocrine neoplasia type 1: a double-edged sword? Endocr Connect. 2017;6(3):151–8.

46 van Treijen MJC, van Beek DJ, van Leeuwaarde RS, Vriens MR, Valk GD. Diagnosing nonfunctional pancreatic NETs in MEN1: the evidence base. J Endocr Soc. 2018;2(9):1067–88.

47 Lloyd RV, Osamura YR, Kloppel G, Rosai J. WHO classification of tumours of endocrine organs; 2017.

48 Dromain C, de Baere T, Lumbroso J, Caillet H, Laplanche A, Boige V, et al. Detection of liver metastases from endocrine tumors: a prospective comparison of somatostatin receptor scintigraphy, computed tomography, and magnetic resonance imaging. J Clin Oncol. 2005;23(1):70–8.

49 D'Assignies G, Fina P, Bruno O, Vullierme M-P, Tubach F, Paradis V, et al. High sensitivity of diffusion-weighted MR imaging for the detection of liver metastases from neuroendocrine tumors: comparison with T2-weighted and dynamic gadolinium-enhanced MR imaging. Radiology. 2013;268(2):390–9.

50 De Robertis R, D'Onofrio M, Zamboni G, Tinazzi Martini P, Gobbo S, Capelli P, et al. Pancreatic neuroendocrine neoplasms: clinical value of diffusion-weighted imaging. Neuroendocrinology. 2016;103(6):758–70.

51 Larghi A, Capurso G, Carnuccio A, Ricci R, Alfieri S, Galasso D, et al. Ki-67 grading of nonfunctioning pancreatic neuroendocrine tumors on histologic samples obtained by EUS-guided fine-needle tissue acquisition: a prospective study. Gastrointest Endosc. 2012;76(3):570–7.

52 Kann PH, Kann B, Fassbender WJ, Forst T, Bartsch DK, Langer P. Small neuroendocrine pancreatic tumors in multiple endocrine neoplasia type 1 (MEN1): least significant change of tumor diameter as determined by endoscopic ultrasound (EUS) imaging. Exp Clin Endocrinol Diabetes. 2006 Jul;114(7):361–5.

53 Conemans EB, Brosens LAA, Raicu-Ionita GM, Pieterman CRC, De Herder WW, Dekkers OM, et al. Prognostic value of WHO grade in pancreatic neuro-endocrine tumors in multiple endocrine neoplasia type 1: results from the DutchMEN1 study group. Pancreatology. 2017;17(5):766–72.