# Multiple Testing, Cut-Point Optimization, and Signs of Publication Bias in Prognostic FDG–PET Imaging Studies of Head and Neck and Lung Cancer: A Review and Meta-Analysis

**Malene M. Clausen** [1,2,*] **, Ivan R. Vogelius** [2,3] **, Andreas Kjær** [1] **and Søren M. Bentzen** [3]

1   Department of Clinical Physiology, Nuclear Medicine & PET and Cluster for Molecular Imaging, Copenhagen University Hospital, Rigshospitalet and University of Copenhagen, 2100 Copenhagen, Denmark; akjaer@sund.ku.dk
2   Department of Oncology, Copenhagen University Hospital, Rigshospitalet, 2100 Copenhagen, Denmark; ivan.richter.vogelius@regionh.dk
3   Division of Biostatistics and Bioinformatics, University of Maryland School of Medicine, Baltimore, MD 21201, USA; sbentzen@som.umaryland.edu
*   Correspondence: malene.martini.clausen@regionh.dk

check for updates

**Abstract:** Positron emission tomography (PET) imaging with 2-deoxy-2-[$^{18}$F]-fluorodeoxyglucose (FDG) was proposed as prognostic marker in radiotherapy. Various uptake metrics and cut points were used, potentially leading to inflated effect estimates. Here, we performed a meta-analysis and systematic review of the prognostic value of pretreatment FDG–PET in head and neck squamous cell carcinoma (HNSCC) and non-small cell lung cancer (NSCLC), with tests for publication bias. Hazard ratio (HR) for overall survival (OS), disease free survival (DFS), and local control was extracted or derived from the 57 studies included. Test for publication bias was performed, and the number of statistical tests and cut-point optimizations were registered. Eggers regression related to correlation of SUVmax with OS/DFS yielded $p = 0.08$/$p = 0.02$ for HNSCC and $p < 0.001$/$p = 0.014$ for NSCLC. No outcomes showed significant correlation with SUVmax, when adjusting for publication bias effect, whereas all four showed a correlation in the conventional meta-analysis. The number of statistical tests and cut points were high with no indication of improvement over time. Our analysis showed significant evidence of publication bias leading to inflated estimates of the prognostic value of SUVmax. We suggest that improved management of these complexities, including predefined statistical analysis plans, are critical for a reliable assessment of FDG–PET.

**Keywords:** FDG–PET/CT; HNSCC; NSCLC; prognostication; publication bias

## 1. Introduction

Positron emission tomography (PET) offers a non-invasive method to assess functional biological characteristics of a tumor in an individual patient with cancer. A number of positron emitting tracers were developed to study various aspects of tumor biology [1–7]. However, clinical practice in cancer PET imaging is still dominated by very few tracers, with 2-deoxy-2-[$^{18}$F]-fluorodeoxyglucose (FDG) being the clinical workhorse for most tumor sites. FDG–PET imaging is primarily used for staging purposes, as a supplement to anatomical images, but advances in the availability of PET imaging led to an increased interest in the feasibility of PET guided radiotherapy planning [8]. Several studies investigated the prognostic value of FDG–PET, and dose escalation to PET-positive areas within the tumor is one of the potential strategies for increasing effect of radiotherapy [9–12].

Despite the widespread use of FDG, only few prospective studies exist for this tracer—the value of FDG as prognostic factor was mainly tested in retrospective cohort studies, for example [13]. There are major limitations with this approach of evidence-generation in medicine. For example, the lack of prospective clinical studies with study registration, formalized sample size estimation and a predefined statistical analysis plans. All of this leads to a risk of "fishing expeditions" with a high risk of false positive findings, exaggerated effect sizes, and subsequent publication bias. Multiple testing is a well-described contributor to false positive findings; when comparisons are made for several subgroups or multiple variables, without adjustment of the type I error rate (false positives).

Searching for a standardized uptake value (SUV) cut-point for dichotomization of the patient group into a poor vs. good prognosis group, which minimizes the *p*-value when comparing outcome in the resulting groups, is an approach frequently used in imaging studies. This "optimization" approach invalidates a simple interpretation of the resulting *p*-value and is associated with a substantial increase in the rate of type I errors [14,15]. Deciding before the start of analyses to use the median SUV as a cut-point, is unbiased. However, in many cases this might not be a biologically meaningful way to classify patients into prognostic subgroups.

In the current study, we reviewed the methodology of published studies of the prognostic value of FDG in head and neck squamous cell carcinoma (HNSCC) and non-small cell lung cancer (NSCLC), and assessed the evidence of publication bias. We included studies where patients received radiotherapy (RT), possibly in combination with other treatment modalities. The Hazard Ratio (HR) for overall survival (OS), disease free survival (DFS), or local control (LC) were considered as outcomes, and analyzed as a function of FDG uptake metrics.

## 2. Materials and Methods

### 2.1. Search Strategy and Eligibility Criteria for Studies

Published reports on the prognostic value of pretreatment FDG–PET in two common tumors, HNSCC and NSCLC, were included.

We searched PubMed for published reports using the following search strings with 'human' filter:
*HNSCC: (FDG OR "18-F") AND ("Head and neck" OR "HNSCC" OR "SCCHN") AND (radiotherapy OR chemoradi\* OR radio\* OR chemo-radi\*) NOT review*
*NSCLC: (FDG OR "18-F") AND ("non-small cell lung cancer" OR "NSCLC") AND (radiotherapy OR chemoradi\* OR radio\* OR chemo-radi\*) NOT review*

In addition, manual screening of the selected articles, reviews, and meta-analyses were used to complement the search. The final date of the search was 1 November 2018, and we did not restrict publication date prior to this date. In this analysis, only articles in English were included. Studies reporting the HR for OS, DFS or LC, versus SUVmax or SUVpeak, were included. The primary treatment was required to include RT, but studies with mixed cohorts including RT, chemoradiotherapy, and surgery were also allowed. However, studies with only few patients receiving RT were excluded [16]. No restrictions on the study design were used. Where multiple reports with overlapping patient cohorts were available, only data from the largest study were included.

### 2.2. Data Extraction

Data were extracted for each study by MMC and entered into the meta-analysis software Review Manager (RevMan) software version 5.3 [17]. There was no attempt to obtain unpublished data. Data were analyzed by MMC and IRV. The HR for OS, DFS, and LC for each trial was extracted or derived from the available data. An HR above 1 implied a survival benefit for lower SUVmax. Both the HR and its confidence interval (CI) were required from each study for inclusion in the meta-analysis. Data from multivariate (MVA) were prioritized over univariate (UVA) analysis, when both were reported.

If the HR with CI was not stated in a report, one of two methods were used for its estimation. If the HR was given with a *p*-value, but without the CI, we assumed a normal distribution of the logarithm of the HR and estimated the CI by first finding the z-parameter of the normal distribution pertaining to the reported *p*-value. We then calculated the standard error of the ln(HR) estimate as $S_{\ln(HR)} = \frac{\ln(HR)}{z}$.

A number of studies did not report an HR, but only a *p*-value, together with the outcome at one or more specified points in time, in most cases in the form of a plot of Kaplan–Meier curves. In these cases, we estimated the HR from the relationship $R(t) = \frac{\ln(p_2(t))}{\ln(p_1(t))}$, where $p_1$ is the Kaplan–Meier estimate for low SUVmax at time t and $p_2$ is the estimate for high SUVmax. When possible, we sampled the ratio at multiple time-points, ranging from the first time an event occurred in both groups to the end of follow-up, and averaged the resulting HR(t) estimates. The CI was then calculated from the *p*-value as explained above. The methodology was previously described in more detail [18]. Variances were then calculated and used as study weights in meta-analysis, using RevMan [17].

Assessment of publication bias was made visually by ordering the studies in forest plots according to variance and by funnel plots. We did not systematically assess the risk of bias in the individual studies. The risk of bias across studies was assessed using the so-called *Egger–Var* method as a formal test for publication bias [19]. We performed quantitative assessment using the Egger's method as follows. For each endpoint and comparison, a linear fit of log HR versus standard error of the study estimate was performed to assess if the included studies effect size depended on study precision. Formally, the regression equation was

$$\ln(HR_i) = \alpha + \beta * SE_i + \varepsilon_i, \tag{1}$$

weighted by the inverse variance of each study. Here, $\alpha$ and $\beta$ were the fitting parameters, $SE_i$ was the standard error of lnHR of the ith study, and $\varepsilon$ was the residual error assumed to have a normal distribution. If $\beta$ was different from zero at the 95% confidence level, it was concluded that the effect size estimates depended on the study precision—a clear indication of publication bias. $\alpha$ in Equation (1) was the extrapolation to zero SE and was used to estimate a publication bias corrected value of ln(HR).

Turning to the assessment of the number of statistical tests and the use of cut-point optimization, the assessment was made independently by MMC and IRV and disagreements were resolved by a consensus meeting. Only statistical tests related to the association of an imaging metric on one side and oncological outcome or baseline characteristics on the other, were counted. When the same question was addressed in univariate and multivariate analysis, the corresponding *p*-value was only counted once. Similarly, it was only counted once if it was part of a rational model building procedure, including forward or backwards elimination. In cases where a large number of multivariate models with a functional image metric were examined, outside of a model building procedure, the multivariate tests were included in the assessment of a number of statistical tests that were investigated, e.g., Schwartz et al. [20].

## 3. Results

The study selection process for this analysis is presented in Figure 1. Of the 930 studies identified by the initial search, 57 were analyzable. A total of 178 full text articles were screened, and 133 of these were excluded, as data were not assessable (not reporting HR, univariate log-rank test without *p*-value, no cut-point for SUV). In other words, 25.3% of the screened full text reports were included. Twelve studies were added from manual cross-referencing of articles, reviews, and browsing for a total of 57 included studies; 27 studies in patients with HNSCC [20–46], and 30 studies in NSCLC [47–76].

Study characteristics are summarized in Tables 1 and 2 for HNSCC and NSCLC, respectively. The vast majority of studies were retrospective analyses—20 studies of HNSCC (74%) and 26 studies of NSCLC (87%). The included studies comprised a total of 5102 patients—1704 patients in the HNSCC group and 3398 in the NSCLC group. The median study size was 74. The NSCLC studies were generally larger, with a median study sample size of 95 patients compared to 58 in the HNSCC group. Twenty-seven studies did not perform MVA for DFS or OS, which were the primary endpoints of

this analysis. A few studies reported no events until long follow-up, which gave rise to additional uncertainty in the HR estimate [54,68]. One study reported no events in the low-uptake group, giving rise to an infinite HR estimate, and the study had to be excluded [77]. A single study was excluded due to problems with interpretation of the KM plots [78].

The patient cohorts in both the HNSCC and NSCLC group were quite heterogeneous with respect to stage, treatment, and follow-up time.

Figures 2 and 3 display the forest plots for SUVmax as a predictor of DFS and OS for HNSCC and NSCLC, respectively, with the studies ordered according to inverse variance from top to bottom. There was a trend for HR to decrease with decreasing variance, and publication bias is therefore suspected. The HR estimate from the pooled analysis is shown by the diamond-shaped mark, and it favored low SUV. However, this should be interpreted with caution due to the suspicion of publication bias. The same trend was observed for LC (Supplemental Figure S1). The data entered in RevMan and HR for UVA and MVA for all studies are listed in the Supplemental Table S1.
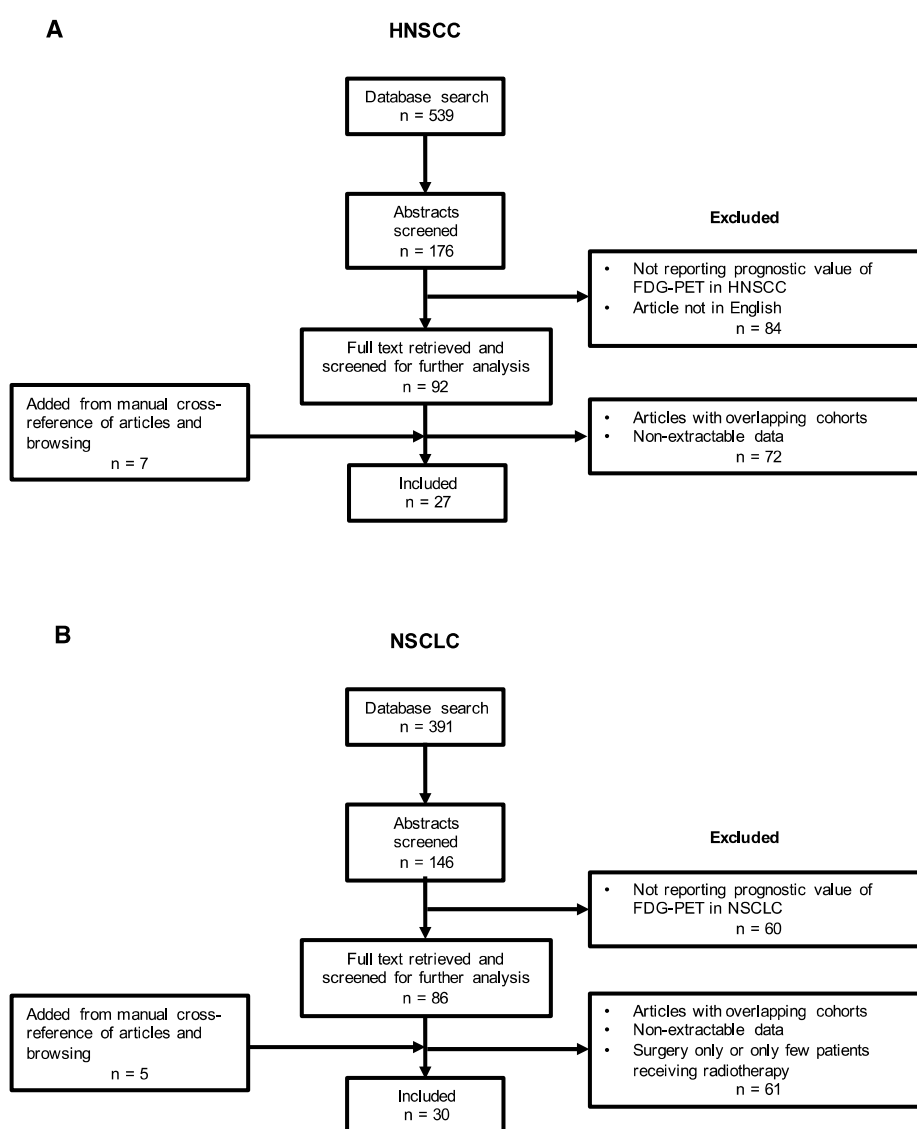


**Figure 1.** Flow diagram of the study selection process for the meta-analysis (**A**: HNSCC, **B**: NSCLC).

**Table 1.** Characteristics of HNSCC studies included in the meta-analysis.

| Author | Year | Tumor Type | Patients | Endpoints | MVA * | Uptake Metric | Cut-Off Value | SUV Threshold | Reconstruction Algorithm | Treatment | Stage | Median Follow Up Time | Data Extraction |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Akagunduz et al. | 2015 | HN | 62 | LRFS, DFS, OS | No | SUVmax, SULmax, MTV | 10.15 (SUL) | Fitted | - | RT/CRT | | 18 months | KM |
| *Allal* et al. | 2004 | HNSCC | 120 | LC, DFS, OS | Yes | SUVmax | 4.76 3.5 | Median Fitted | - | RT +/− CT, surgery +/− RT | I-IV | 48 months | HR |
| Baschnagel et al. | 2015 | HNSCC | 74 | LC, LRC, DFS | Yes | SUVmax | 13.8 | Median | - | CRT | T1-T4, N0-N3 | 35 months | HR |
| *Brun* et al. | 2002 | HNSCC | 47 | CR, LRC, OS | No | SUV, MR | 9.0 | Median | Iterative ML | RT, CRT | II-IV | 3.3 years | HR |
| *Cacicedo* et al. | 2017 | HNSCC | 58 | DFS, LRC, DMFS, OS | Yes | SUVmax | 11.85 (SUV-T), 5.4 (SUV-N) | Median | - | Surgery + RT, RT +/− CT | III-IVB | 31 months | KM |
| Chan et al. | 2017 | OHSCC | 124 | OS, RFS | Yes | SUVmax, SUVmean, MTV, TLG, entropy, contrast, busyness, complexity | 14.22 | | OSEM | CRT | III-IV | 28.7 months | HR |
| Chung et al. | 2009 | SCC | 82 | CR, DFS | Yes | MTV, SUV > 2.5 | 10.0 | Median | OSEM | RT, CRT | I-IV | 34.8 months | HR |
| Halfpenny et al. | 2002 | HNSCC | 58 | Survival | Yes | SUVpeak | 10.0 | Fitted | FBP | Surgery, +/−RT | I-IV | 39 months | HR |
| Higgins et al. | 2012 | HNSCC | 8 | DFS, LRC, DMFS, OS | No | SUVmax, SUV mean, TLG | 15.4 | Median | OSEM | RT, CRT | III-IV (97%) | 15 months | KM |
| Kim et al. | 2007 | OSCC | 52 | LC, DFS, OS | Yes | SUVmax | 6.0 | Median | - | Surgery +/− RT/CRT | I-IV | 36 months | HR |
| Kitajima et al. | 2014 | Laryngeal | 51 | PFS, LC, NPFS, DMFS | Yes | SUVmax | 4.6 | Fitted | RAMLA | RT +/− CT, surgery +/− CRT | | 48.6 months | KM |
| Komar et al. | 2014 | HNSCC | 22 | OS | No | SUVmax, MATV | 11.74 | Median | - | Surgery +/− CRT, RT | I-IV | 41 months | KM |
| Koyasu et al. | 2014 | SCC | 108 | DFS | Yes | SUVmax, MTV, TLG | 10.0 | Fitted | 3D iterative | RT +/− CT, surgery +/− RT | I-IV | 36.4 months | HR |
| Kunkel et al. | 2003 | OSCC | 44 | OS | Yes | SUVpeak | 5.6 | Median | - | RT (preop.) + surgery | I-IV | 38 months | HR |

Table 1. *Cont.*

| Author | Year | Tumor Type | Patients | Endpoints | MVA * | Uptake Metric | Cut-Off Value | SUV Threshold | Reconstruction Algorithm | Treatment | Stage | Median Follow Up Time | Data Extraction |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Liao et al. | 2009 | OSCC | 109 | LC, DFS, DSS, OS | No | SUVmax | 19.3 | Median | ML, OSEM | Surgery + RT/CRT | III-IV | 39 months | HR |
| Machtay et al. | 2009 | HNSCC | 60 | DFS, OS | Yes | SUVmax | 9.0 | Literature | - | RT, CRT, surgery + CRT/RT | I-IV | - | HR |
| *Minn* et al. | 1997 | HNSCC | 37 | OS | No | SUVlean, MR | 9.0 | Median | - | RT +/− surgery | II-IV | 43 months | HR |
| Moon et al. | 2015 | NPC | 44 | DFS | Yes | SUVmax, SUVmean, TLG, MTV | 7.8 | Fitted | OSEM | CRT | II-IVB | 40 months | HR |
| *Ng* et al. | 2016 | OHSCC | 86 | PFS, OS | Yes | SUVmax, SUVmean, TLG, MTV | 19.44 | Fitted | - | CRT | III-IVB | 28 months | HR |
| Preda et al. | 2016 | HNSCC | 57 | DFS | Yes | SUVmax | 5.75 | Fitted | OSEM | Surgery + RT +/− CT, RT + CT | T1-T4, N0-N2 | 21.3 months | HR |
| Roh et al. | 2007 | SCC | 79 | DFS, LC, OS | No | SUVmax | 8.0 | Fitted | - | Surgery +/− RT or RT +/− CT | III-IV | 36 months | KM |
| *Schwartz* et al. | 2004 | HNSCC | 54 | LRFS, DFS, OS | No | SUVmax | 9.0 | Median | FBP | RT +/− CT, surgery +/− RT | I-IV | 17.5 months | KM |
| *Schwartz* et al. | 2015 | HNSCC | 74 | PFS, OS | No | SUVmax, MTV | 15.07 | Median | - | CRT | III-IV | 4.2 years | HR |
| Suzuki et al. | 2014 | OPSCC + HPSCC | 49 | OS | Yes | SUVmax | 8.0 | Fitted | OSEM | Surgery + RT +/− CT, RT + CT | I-IV | 33 months | HR |
| Suzuki-Shibata et al. | 2017 | OTSCC | 33 | PFS, OS | Yes | SUVmax, MTV | 15.7 | Fitted | FORE-OSEM | CRT | II-IVA | 36 months | HR |
| Torizuka et al. | 2009 | HNSCC | 50 | LC, DFS | No | SUVpeak, SUV cont. variable | 7.0 | Fitted | OSEM | RT, CRT, surgery +/− CRT | I-IV | 15 months | KM |
| Xie et al. | 2010 | NPC | 62 | OS, DFS | No | SUVmax | 8.0 | Fitted | - | CRT | III-IVB | 61 months | KM |

* Performed multivariate analysis (MVA) in regards to the endpoints analyzed in this study: DFS and OS. Studies listed with author in *italic* are performed as prospective studies. Other studies are retrospective studies. HNSCC: head and neck squamous cell carcinoma, HN: head and neck cancer, LRFS: local recurrence free survival, DFS: disease free survival, OS: overall survival, SUV: standardized uptake value, SUL: lean body mass corrected standardized uptake value, MTV: metabolic tumor volume, RT: radiotherapy, CRT: chemoradiotherapy, KM: Kaplan-Meier, LC: local control, HR: hazard ratio, LRC: locoregional control, CR: complete response, ML: maximum likelihood, DMFS: distant metastasis free survival, CT: chemotherapy, OHSCC: oropharyngeal or hypopharyngeal squamous cell carcinoma, RFS: recurrence free survival, TLG: total lesion glycolysis, OSEM: ordered-subset expectation maximum, SCC: squamous cell carcinoma, FBP: filtered back projection, OSCC: oral squamous cell carcinoma, PFS: progression free survival, NPFS: nodal progression free survival, RAMLA: row-action maximum-likelihood algorithm, MATV: metabolically active tumor volume, DSS: disease-specific survival, MR: metabolic rate, NPC: nasopharyngeal carcinoma, OPSCC: oropharyngeal squamous cell carcinoma, HPSCC: hypopharyngeal squamous cell carcinoma, OTSCC: oral tongue squamous cell carcinoma, and FORE-OSEM: Fourier rebinning-ordered-subset expectation maximum.

**Table 2.** Characteristics of NSCLC studies included in the meta-analysis.

| Author | Year | Tumor Type | Patients | Endpoints | MVA * | Uptake Metric | Cut-Off Value | SUV Threshold | Reconstruction Algorithm | Treatment | Stage | Median Follow Up Time | Data Extraction |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ahuja et al. | 1998 | NSCLC | 155 | OS | Yes | SUVpeak (SUR) 80% of max | 10.0 | Fitted | - | Surgery, RT, CRT | I-IV | 20.9 months | HR |
| Aoki et al. | 2016 | NSCLC | 74 | LC | Yes | SUVmax, AID | 4.0 | Literature | - | SBRT | I | 24.5 months | HR |
| Borst et al. | 2005 | NSCLC | 51 | DSS, OS | No | SUVmax, SUV cont. variable | 15.0 | Median | OSEM | CRT | I-III | 17 months | KM |
| *Carvalho* et al. | 2013 | NSCLC | 220 | OS | No | MTV, SUVmax, SUV | 10.12 | Median | OSEM | RT, CRT | I-IIIB | 1.47 years | KM |
| Cerfolio et al. | 2005 | NSCLC | 315 | OS, DFS | Yes | SUVmax | 10.0 | Median | Iterative | Surgery +/− CRT | I-IV | 26 months | HR |
| Chen et al. | 2012 | NSCLC | 105 | PFS, OS | Yes | TLG, MTV, SUVmax | 15.0 | Fitted | OSEM | Surgery, CT, RT or CRT | I-IV | 3.1 years | HR |
| Clarke et al. | 2012 | NSCLC | 82 | OS, RFS, DFS, CSS, RR, LR, DM | No | SUVmax | 4.75 | Median | - | SBRT | I | 2 years | KM |
| Hamamoto et al. | 2011 | NSCLC | 26 | LFF | No | SUVmax | 5.0 | Fitted | - | SBRT | I | 21 months | KM |
| Hofheinz et al. | 2016 | NSCLC | 31 | PFS, OS | No | SUV, SURtc, Kslope | 7.6 | Fitted | PSF + TOF | CRT and/or surgery | T1-4N0-3M0 | - | HR |
| Horne et al. | 2014 | NSCLC | 95 | LC, PFS, OS | Yes | SUVmax | 5.0 | Literature | - | SBRT | IA-IB | 16.33 months | HR |
| Huyn et al. | 2015 | NSCLC | 161 | DFS, OS | Yes | SUVmax, MTV | 14.0 | Fitted | OSEM | Surgery +/− CT and/or RT | IIIA-N2 | 20 months | HR |
| Imamura et al. | 2011 | NSCLC | 62 | OS, PFS | No | SUVmax | 6.0 | Median | 3D-RAMLA | CT or CRT | IIB-IV | 464 days | KM |
| Jiang et al. | 2018 | NSCLC | 151 | OS | No | SUVmax | 13.8 | Median | - | CRT, RT, CT | I-IV | 10 years | HR |
| Kohutek et al. | 2015 | NSCLC | 211 | OS | Yes | SUVmax, GTV | 3.0 | Fitted | OSEM | SBRT | T1-2N0M0 | 25.2 months | HR |
| Lee et al. | 2012 | NSCLC | 205 | OS | Yes | SUVmax | 13.0 | Fitted | Iterative | Neoadj. CRT, + surgery | IIIA | 1.6 years | HR |
| Nair et al. | 2014 | NSCLC | 163 | PFS, OS, LRFS, DMFS | Yes | SUVmax | 7.0 | Median | - | RT, SBRT | T1-2N0M0 | 16 months | KM |
| *Nawara* et al. | 2012 | NSCLC | 91 | OS | No | SUVmax, SUVmean | 7.0 | Median | Iterative | RT +/− induction CT | I-IIIB | - | KM |
| Pyka et al. | 2015 | NSCLC | 45 | DSS, OS | Yes | SUVmax, SUVmean, MTV, COV, entropy, coarseness, contrast, correlation | 11.2 (OS), 12.3 (DSS) | Fitted | OSEM | SBRT | T1-2N0M0 | 21.4 months | KM |
| Sasaki et al. | 2005 | NSCLC | 162 | OS, DFS | Yes | SUVmax | 5.0 | Fitted | Iterative | Surgery +/− RT or RT/CRT | I-IIIB | 17 months | HR |

<div align="center">**Table 2.** *Cont.*</div>

| Author | Year | Tumor Type | Patients | Endpoints | MVA * | Uptake Metric | Cut-Off Value | SUV Threshold | Reconstruction Algorithm | Treatment | Stage | Median Follow Up Time | Data Extraction |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Shirai et al. | 2017 | NSCLC | 45 | LC, PFS, OS | No | SUVmax | 5.5 | Median | - | C-ion RT | I | 28.9 months | KM |
| Sugawara et al. | 1999 | NSCLC | 38 | OS | No | SUVlean | 8.72 | Median | Hanning filter | Surgery, CRT | I-IV | 26.5 months | KM |
| Takeda et al. | 2011 | NSCLC | 95 | LC | No | SUVmax | 6.0 | Fitted | DRAMA | SBRT | IA-IIIB | 16 months | KM |
| Takeda et al. | 2014 | NSCLC | 152 | OS, DFS. LC, RC, DMC, CSS | Yes | SUVmax | 3.35 (LC), 3.64 (RC), 2.47 (DMC, DFS), 2.55 (CSS, OS) | Fitted | RAMLA | SBRT | T1-2N0M0 | 25.3 months | HR |
| Takeda et al. | 2017 | NSCLC | 26 | LC, PFS, OS | No | SUV, MTV, TLG, entropy, dissimilarity, HILAE, zone percentage | 8.18 | Median | OSEM | SBRT | T1-2N0M0 | 36 months | KM |
| Ulger et al. | 2014 | NSCLC | 103 | OS, RFS, DFS | Yes | SUVmax | 10.7 | Median | - | 3D-CRT | IIIA-IIIB | 22.63 months | HR |
| Vansteenkiste et al. | 1999 | NSCLC | 125 | OS | Yes | SUVmax | 7.0 | Fitted | - | Surgery +/− induction CT, RT +/− induction CT | I-IIIB | 19 months (mean) | HR |
| *Vesselle* et al. | 2007 | NSCLC | 208 | OS, DFS | Yes | SUVmax | 7.0 | Fitted | Hanning filter | Surgery +/− neoadj. Or adjuvant therapy | I-IV | 37 months | HR |
| Vu et al. | 2013 | NSCLC | 50 | OS, RFS | No | SUVmax, TLG, MTV | 6.43 | Median | - | SBRT | I | 25.1 months | HR |
| *Xiang* et al. | 2012 | NSCLC | 84 | LRFS, DMFS, PFS, OS | Yes | SUVmax | 14.2 | Median | - | High dose proton + CT | III | 19.2 months | HR |
| Yilmaz et al. | 2018 | NSCLC | 67 | PFS, OS | Yes | SUVmax | 15.0 | Median | - | CRT | III | 20.7 months | HR |

* Performed multivariate analysis (MVA) in regards to the endpoints analyzed in this study: DFS and OS. Studies listed with author in *italic* are performed as prospective studies. Other studies are retrospective studies. NSCLC: non-small cell lung cancer, OS: overall survival, SUV: standardized uptake value, SUR: standardized uptake ratio, RT: radiotherapy, CRT: chemoradiotherapy, HR: hazard ratio, LC: local control, AID: average iodine density, SBRT: stereotactic body radiation therapy, DSS: disease-specific survival, OSEM: ordered-subset expectation maximum, KM: Kaplan-Meier, MTV: metabolic tumor volume, DFS: disease free survival, PFS: progression free survival, TLG: total lesion glycolysis, CT: chemotherapy, RFS: recurrence-free survival, CSS: cause-specific survival, RR: regional relapse, LR: local relapse, DM: distant metastasis, LFF: local failure free rate, PSF: point spread function, TOF: time of flight, RAMLA: row-action maximum-likelihood algorithm, GTV: gross tumor volume, LRFS: local recurrence-free survival, DMFS: distant metastasis free survival, COV: coefficient of variation, DRAMA: dynamic row-action expectation maximization algorithm, RC: regional control, DMC: distant metastasis control, HILAE: high-intensity large-area emphasis, and 3D-CRT: three-dimensional conformal radiotherapy.
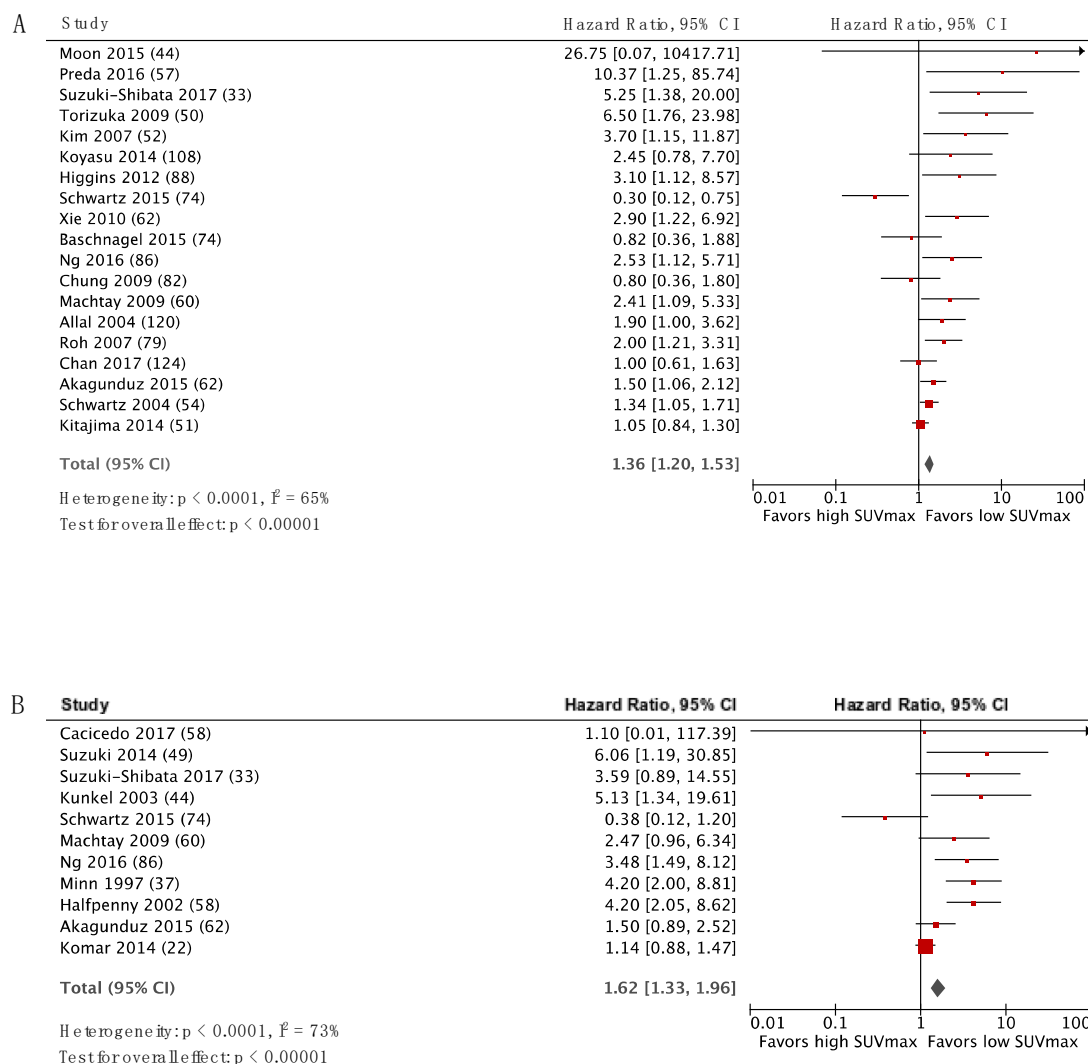
**Figure 2.** Forest plots for SUVmax as a predictor of DFS (**A**) and OS (**B**) in HNSCC. The number of patients included in each study is specified in parenthesis.

When Eggers regression was applied, neither OS nor DFS appeared to be significantly associated with FDG uptake in neither HNSCC or NSCLC. The regression slopes were significantly greater than zero in three of the four cases: DFS for HNSCC ($p = 0.02$) and both OS ($p < 0.001$) and DFS ($p = 0.014$) for NSCLC. See the supplement for details and the associated plots (Supplemental Figure S2).

Figure 4 shows a plot of the number of patients, number of statistical tests, and number of cut-point optimizations against the year of publication. Unfortunately, there is little sign of a consistent improvement in study characteristics, i.e., number of statistical tests, number of cut-point analyses, and size of study population over time. For HNSCC studies, there was a statistically significant increase in the number of cut-point optimizations versus time (Spearman rho = 0.5, $p = 0.02$), while there was no significant increase in the number of patients for later studies and data were also in accordance with no change in the number of tests performed (Figure 4A). Data for the NSCLC studies are shown in Figure 4B, where the Spearman rank correlation coefficients were statistically in agreement with no change over time, in either the number of patients, number of cut-points, or number of tests performed ($p > 0.27$ for all coefficients).

Power calculations were not performed in any of the included studies, and only three studies mention or adjust for multiple testing [27,33,60].
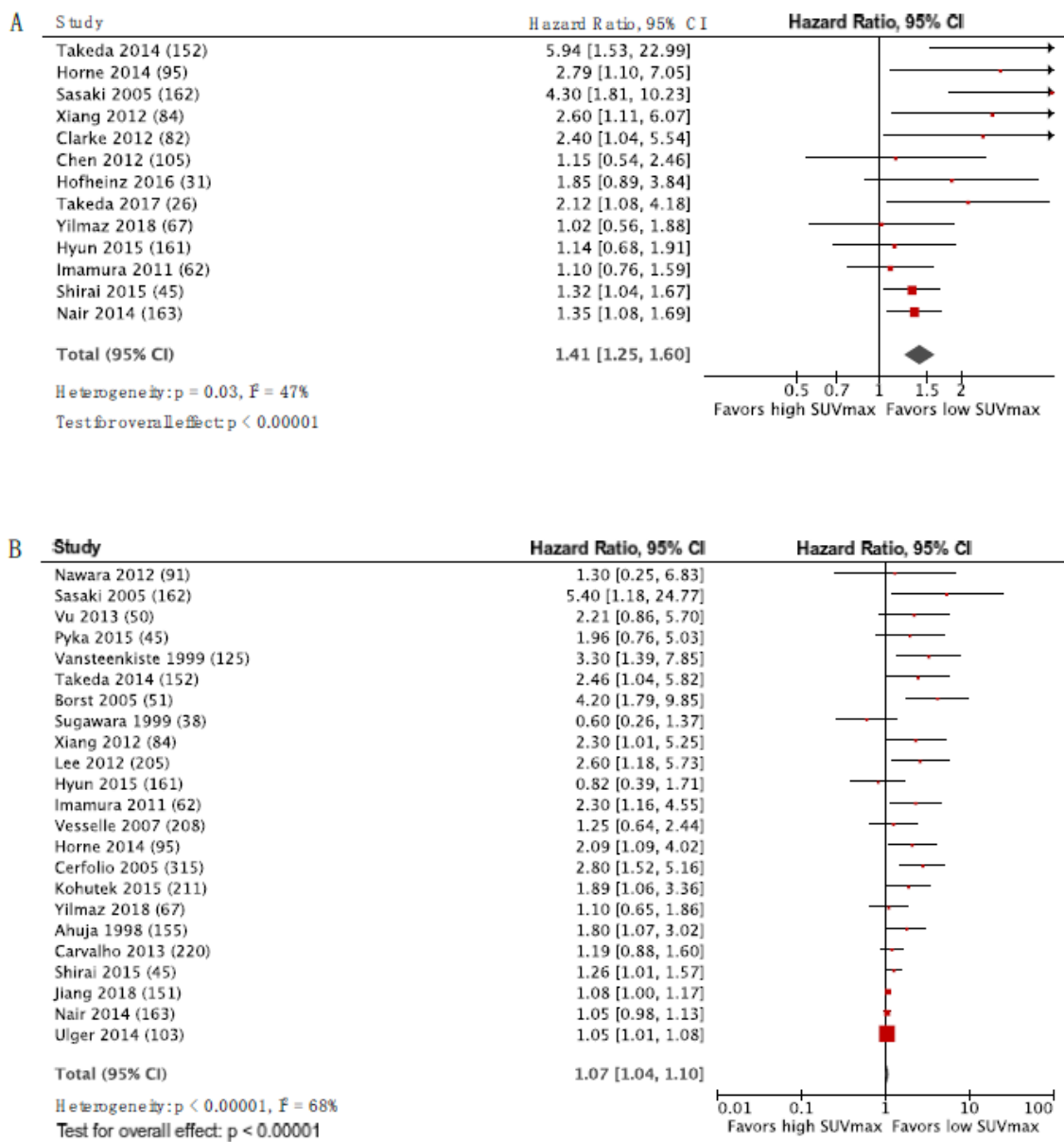
A

| Study | Hazard Ratio, 95% CI |
|---|---|
| Takeda 2014 (152) | 5.94 [1.53, 22.99] |
| Horne 2014 (95) | 2.79 [1.10, 7.05] |
| Sasaki 2005 (162) | 4.30 [1.81, 10.23] |
| Xiang 2012 (84) | 2.60 [1.11, 6.07] |
| Clarke 2012 (82) | 2.40 [1.04, 5.54] |
| Chen 2012 (105) | 1.15 [0.54, 2.46] |
| Hofheinz 2016 (31) | 1.85 [0.89, 3.84] |
| Takeda 2017 (26) | 2.12 [1.08, 4.18] |
| Yilmaz 2018 (67) | 1.02 [0.56, 1.88] |
| Hyun 2015 (161) | 1.14 [0.68, 1.91] |
| Imamura 2011 (62) | 1.10 [0.76, 1.59] |
| Shirai 2015 (45) | 1.32 [1.04, 1.67] |
| Nair 2014 (163) | 1.35 [1.08, 1.69] |
| Total (95% CI) | 1.41 [1.25, 1.60] |

Heterogeneity: $p = 0.03$, $I^2 = 47\%$
Test for overall effect: $p < 0.00001$

B

| Study | Hazard Ratio, 95% CI |
|---|---|
| Nawara 2012 (91) | 1.30 [0.25, 6.83] |
| Sasaki 2005 (162) | 5.40 [1.18, 24.77] |
| Vu 2013 (50) | 2.21 [0.86, 5.70] |
| Pyka 2015 (45) | 1.96 [0.76, 5.03] |
| Vansteenkiste 1999 (125) | 3.30 [1.39, 7.85] |
| Takeda 2014 (152) | 2.46 [1.04, 5.82] |
| Borst 2005 (51) | 4.20 [1.79, 9.85] |
| Sugawara 1999 (38) | 0.60 [0.26, 1.37] |
| Xiang 2012 (84) | 2.30 [1.01, 5.25] |
| Lee 2012 (205) | 2.60 [1.18, 5.73] |
| Hyun 2015 (161) | 0.82 [0.39, 1.71] |
| Imamura 2011 (62) | 2.30 [1.16, 4.55] |
| Vesselle 2007 (208) | 1.25 [0.64, 2.44] |
| Horne 2014 (95) | 2.09 [1.09, 4.02] |
| Cerfolio 2005 (315) | 2.80 [1.52, 5.16] |
| Kohutek 2015 (211) | 1.89 [1.06, 3.36] |
| Yilmaz 2018 (67) | 1.10 [0.65, 1.86] |
| Ahuja 1998 (155) | 1.80 [1.07, 3.02] |
| Carvalho 2013 (220) | 1.19 [0.88, 1.60] |
| Shirai 2015 (45) | 1.26 [1.01, 1.57] |
| Jiang 2018 (151) | 1.08 [1.00, 1.17] |
| Nair 2014 (163) | 1.05 [0.98, 1.13] |
| Ulger 2014 (103) | 1.05 [1.01, 1.08] |
| Total (95% CI) | 1.07 [1.04, 1.10] |

Heterogeneity: $p < 0.00001$, $I^2 = 68\%$
Test for overall effect: $p < 0.00001$

**Figure 3.** Forest plots for SUVmax as a predictor of DFS (**A**) and OS (**B**) in NSCLC. The number of patients included in each study is specified in parenthesis.
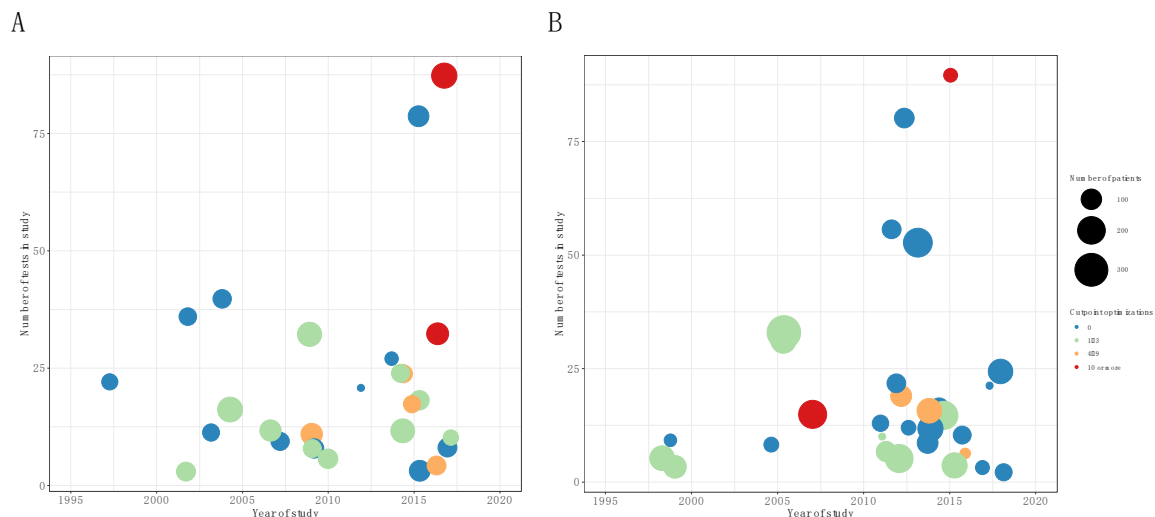
**Figure 4.** Bubble plot of number of patients included in study versus year of publication for HNSCC (**A**) and NSCLC (**B**). Bubble size is proportional to number of statistical tests counted in paper (large is problematic) and color denotes the number of cut-point optimization performed (the fewer the better).

## 4. Discussion

Publication bias is a well-known problem, enriching the literature with false/true positive studies that will not be balanced by other studies with negative findings that are more likely to remain unpublished. This in turn will inflate the effect size estimated from an intervention or the discriminatory power of a diagnostic test. The inflated effect sizes from individual studies will carry over to a meta-analysis [79], thus reducing the value of the meta-analysis in evidence-based medicine. Indeed, our systematic analysis of prognostic studies of FDG uptake found statistically significant evidence of publication bias. Small studies are at particular risk of inflated effect-size bias [80] and it is thus a concern that the median study size was only 58 and 95 patients in published HNSCC and NSCLC studies, respectively. The TRIPOD reporting guidelines [81], attempts to address the problem by requiring a sample size justification in reporting, but this is not provided in any of the studies included here. Additionally, it might be argued that the general reporting guidelines of TRIPOD, albeit relevant, are not sufficiently specific for adequate reporting of image-based prognostic studies. In particular, it is an important concern that a large number of possible predictors can be extracted from a PET scan—SUVmax, SUVpeak, SUVmean, MTV, and TLG, just to name a few. Multiple comparisons, post-hoc search for positive associations and scanning for 'optimal' cut-off values separating the low and the high uptake groups, increase the risk of false positive findings [82], as also discussed by Vesselle et al., in the context of FDG prognostication [73].

A limitation of our study was that we did not have access to individual patient data, which led to the exclusion of some reports. Most of the included studies were conducted as retrospective studies (80.7%), without a predefined data analysis plan. While this might be defendable in the explorative setting, it increases the risk of overestimating the effect size if the cohort studies are not followed by controlled trials or studies with pre-specified protocols. In particular, with FDG, we would argue that we are beyond the exploratory phase and should perform larger studies with predefined protocols, to unequivocally reveal the prognostic or predictive role of FDG uptake in cancers that are common in the two sites studied in the present work. With the high number of correlations that are testable in image-based prognostication, it appears prudent to require predefined research protocols and, perhaps, publication of raw data to allow independent validation of findings, regardless of the chosen cut-point or predictor. It is possible that a functional imaging specific extension to the TRIPOD or REMARK guidelines could be of use. When published studies perform tens of comparisons and multiple cut-point optimizations in datasets of less than 100 patients, and without correction for multiple comparisons, the field is bound to be dominated by false or exaggerated correlations, which will

ultimately harm patients if applied in clinical decision making and harm a promising field of research by misusing resources.

It is a substantial challenge to accommodate cross-study synthesis of data in meta-analysis at the same time as allowing the individual authors to appropriately handle the coding of image metrics in their study. Decisions to use continuous coding of SUV, logarithmic transformation, or a limited number of cut-points are all fair (if performed correctly), but hampers the ability to perform a meta-analysis. It appears to us that the complexity of these analyses implies that publication of the raw modeling data is a necessity for meaningful synthesis of data. We believe that the observations of the current study imply that such a synthesis is necessary for real progress.

## 5. Conclusions

Functional imaging with FDG or other tracers remains a promising tool for prognostication, prediction, and treatment selection for cancer patients. However, the current study points to issues limiting the interpretation, including inadequate sample sizes, lack of predefined analysis plans, lack of correction for multiple testing, and post-hoc cut-point optimizations. These issues result in a high risk of inflated effect sizes or false positive correlations that must be addressed to avoid leading the field astray.

## References

1.  Young, H.; Baum, R.; Cremerius, U.; Herholz, K.; Hoekstra, O.; Lammertsma, A.A.; Pruim, J.; Price, P. Measurement of Clinical and Subclinical Tumour Response Using [$^{18}$F]-Fluorodeoxyglucose and Positron Emission Tomography: Review and 1999 Eortc Recommendations. European Organization for Research and Treatment of Cancer (Eortc) Pet Study Group. *Eur. J. Cancer* **1999**, *35*, 1773–1782. [CrossRef]

2.  Yang, D.J.; Wallace, S.; Cherif, A.; Li, C.; Gretzer, M.B.; Kim, E.E.; Podoloff, D.A. Development of F-18-Labeled Fluoroerythronitroimidazole as a Pet Agent for Imaging Tumor Hypoxia. *Radiology* **1995**, *194*, 795–800. [CrossRef] [PubMed]

3.  Shields, A.F.; Grierson, J.R.; Dohmen, B.M.; Machulla, H.J.; Stayanoff, J.C.; Lawhorn-Crews, J.M.; Obradovich, J.E.; Muzik, O.; Mangner, T.J. Imaging Proliferation in Vivo with [F-18]Flt and Positron Emission Tomography. *Nat. Med.* **1998**, *4*, 1334–1346. [CrossRef] [PubMed]

4.  Fujibayashi, Y.; Taniuchi, H.; Yonekura, Y.; Ohtani, H.; Konishi, J.; Yokoyama, A. Copper-62-Atsm: A New Hypoxia Imaging Agent with High Permeability and Low Redox Potential. *J. Nucl. Med.* **1997**, *38*, 1155–1160.

5.  Weber, W.A.; Wester, H.J.; Grosu, A.L.; Herz, M.; Dzewas, B.; Feldmann, H.J.; Molls, M.; Stocklin, G.; Schwaiger, M. O-(2-[18F]Fluoroethyl)-L-Tyrosine and L-[Methyl-11c]Methionine Uptake in Brain Tumours: Initial Results of a Comparative Study. *Eur. J. Nucl. Med.* **2000**, *27*, 542–549. [CrossRef]

6.  Panebianco, V.; Sciarra, A.; Lisi, D.; Galati, F.; Buonocore, V.; Catalano, C.; Gentile, V.; Laghi, A.; Passariello, R. Prostate Cancer: 1hmrs-Dcemr at 3t Versus [(18)F]Choline Pet/Ct in the Detection of Local Prostate Cancer Recurrence in Men with Biochemical Progression after Radical Retropubic Prostatectomy (Rrp). *Eur. J. Radiol.* **2012**, *81*, 700–708. [CrossRef]

7.  Kurdziel, K.A.; Shih, J.H.; Apolo, A.B.; Lindenberg, L.; Mena, E.; McKinney, Y.Y.; Adler, S.S.; Turkbey, B.; Dahut, W.; Gulley, J.L.; et al. The Kinetics and Reproducibility of 18F-Sodium Fluoride for Oncology Using Current Pet Camera Technology. *J. Nucl. Med.* **2012**, *53*, 1175–1184. [CrossRef]

8.  Das, S.K.; Ten Haken, R.K. Functional and molecular image guidance in radiotherapy treatment planning optimization. *Semin. Radiat. Oncol.* **2011**, *21*, 111–118. [CrossRef]

9.  Wanet, M.; Delor, A.; Hanin, F.X.; Ghaye, B.; Van Maanen, A.; Remouchamps, V.; Clermont, C.; Goossens, S.; Lee, J.A.; Janssens, G.; et al. An individualized radiation dose escalation trial in non-small cell lung cancer based on FDG-PET imaging. *Strahlenther. Onkol.* **2017**, *193*, 812–822. [CrossRef]

10. van Diessen, J.; De Ruysscher, D.; Sonke, J.J.; Damen, E.; Sikorska, K.; Reymen, B.; van Elmpt, W.; Westman, G.; Fredberg Persson, G.; Dieleman, E.; et al. The acute and late toxicity results of a randomized phase II dose-escalation trial in non-small cell lung cancer (PET-boost trial). *Radiother. Oncol.* **2019**, *131*, 166–173. [CrossRef]

11. Differding, S.; Sterpin, E.; Janssens, G.; Hanin, F.X.; Lee, J.A.; Gregoire, V. Methodology for adaptive and robust FDG-PET escalated dose painting by numbers in head and neck tumors. *Acta Oncol.* **2016**, *55*, 217–225. [CrossRef] [PubMed]

12. Rasmussen, J.H.; Hakansson, K.; Vogelius, I.R.; Aznar, M.C.; Fischer, B.M.; Friborg, J.; Loft, A.; Kristensen, C.A.; Bentzen, S.M.; Specht, L. Phase I trial of 18F-Fludeoxyglucose based radiation dose painting with concomitant cisplatin in head and neck cancer. *Radiother. Oncol.* **2016**, *120*, 76–80. [CrossRef] [PubMed]

13. Xie, P.; Li, M.; Zhao, H.; Sun, X.; Fu, Z.; Yu, J. 18F-FDG PET or PET-CT to evaluate prognosis for head and neck cancer: A meta-analysis. *J. Cancer Res. Clin. Oncol.* **2011**, *137*, 1085–1093. [CrossRef] [PubMed]

14. Faraggi, D.; Simon, R. A simulation study of cross-validation for selecting an optimal cutpoint in univariate survival analysis. *Stat. Med.* **1996**, *15*, 2203–2213. [CrossRef]

15. Hilsenbeck, S.G.; Clark, G.M.; McGuire, W.L. Why do so many prognostic factors fail to pan out? *Breast Cancer Res. Treat.* **1992**, *22*, 197–206. [CrossRef]

16. Nguyen, X.C.; Lee, W.W.; Chung, J.H.; Park, S.Y.; Sung, S.W.; Kim, Y.K.; So, Y.; Lee, D.S.; Chung, J.K.; Lee, M.C.; et al. FDG uptake, glucose transporter type 1, and Ki-67 expressions in non-small-cell lung cancer: Correlations and prognostic values. *Eur. J. Radiol.* **2007**, *62*, 214–219. [CrossRef]

17. *Review Manager (RevMan) [Computer Program]*; The Nordic Cochrane Centre, The Cochrane Collaboration: Copenhagen, Switzerland, 2012.

18. Diez, P.; Vogelius, I.S.; Bentzen, S.M. A new method for synthesizing radiation dose-response data from multiple trials applied to prostate cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **2010**, *77*, 1066–1071. [CrossRef]

19. Moreno, S.G.; Sutton, A.J.; Ades, A.E.; Stanley, T.D.; Abrams, K.R.; Peters, J.L.; Cooper, N.J. Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. *BMC Med. Res. Methodol.* **2009**, *9*, 2. [CrossRef]

20. Schwartz, D.; Rajendran, J.; Yueh, B.; Coltrera, M.; LeBlanc, M.; Eary, J.; Krohn, K. FDG-PET prediction of head and neck squamous cell cancer outcomes. *Arch. Otolaryngol. Head Neck Surg.* **2004**, *130*, 1361–1367. [CrossRef]

21. Akagunduz, O.O.; Savas, R.; Yalman, D.; Kocacelebi, K.; Esassolak, M. Can adaptive threshold-based metabolic tumor volume (MTV) and lean body mass corrected standard uptake value (SUL) predict prognosis in head and neck cancer patients treated with definitive radiotherapy/chemoradiotherapy? *Nucl. Med. Biol.* **2015**, *42*, 899–904. [CrossRef]

22. Allal, A.S.; Slosman, D.O.; Kebdani, T.; Allaoua, M.; Lehmann, W.; Dulguerov, P. Prediction of outcome in head-and-neck cancer patients using the standardized uptake value of 2-[18F]fluoro-2-deoxy-D-glucose. *Int. J. Radiat. Oncol. Biol. Phys.* **2004**, *59*, 1295–1300. [CrossRef]

23. Baschnagel, A.M.; Wobb, J.L.; Dilworth, J.T.; Williams, L.; Eskandari, M.; Wu, D.; Pruetz, B.L.; Wilson, G.D. The association of (18)F-FDG PET and glucose metabolism biomarkers GLUT1 and HK2 in p16 positive and negative head and neck squamous cell carcinomas. *Radiother. Oncol.* **2015**, *117*, 118–124. [CrossRef] [PubMed]

24. Brun, E.; Kjellén, E.; Tennvall, J.; Ohlsson, T.; Sandell, A.; Perfekt, R.; Wennerberg, J.; Strand, S. FDG PET studies during treatment: Prediction of therapy outcome in head and neck squamous cell carcinoma. *Head Neck* **2002**, *24*, 127–135. [CrossRef] [PubMed]

25. Cacicedo, J.; Fernandez, I.; Del Hoyo, O.; Navarro, A.; Gomez-Iturriaga, A.; Pijoan, J.I.; Martinez-Indart, L.; Escudero, J.; Gomez-Suarez, J.; de Zarate, R.O.; et al. Prognostic value of maximum standardized uptake value measured by pretreatment 18F-FDG PET/CT in locally advanced head and neck squamous cell carcinoma. *Clin. Transl. Oncol.* **2017**, *19*, 1337–1349. [CrossRef] [PubMed]

26. Chan, S.C.; Cheng, N.M.; Hsieh, C.H.; Ng, S.H.; Lin, C.Y.; Yen, T.C.; Hsu, C.L.; Wan, H.M.; Liao, C.T.; Chang, K.P.; et al. Multiparametric imaging using (18)F-FDG PET/CT heterogeneity parameters and functional MRI techniques: Prognostic significance in patients with primary advanced oropharyngeal or hypopharyngeal squamous cell carcinoma treated with chemoradiotherapy. *Oncotarget* **2017**, *8*, 62606–62621. [CrossRef] [PubMed]

27. Chung, M.K.; Jeong, H.S.; Park, S.G.; Jang, J.Y.; Son, Y.I.; Choi, J.Y.; Hyun, S.H.; Park, K.; Ahn, M.J.; Ahn, Y.C.; et al. Metabolic tumor volume of [18F]-fluorodeoxyglucose positron emission tomography/computed tomography predicts short-term outcome to radiotherapy with or without chemotherapy in pharyngeal cancer. *Clin. Cancer Res.* **2009**, *15*, 5861–5868. [CrossRef] [PubMed]

28. Halfpenny, W.; Hain, S.; Biassoni, L.; Maisey, M.; Sherman, J.; McGurk, M. FDG-PET. A possible prognostic factor in head and neck cancer. *Br. J. Cancer* **2002**, *86*, 512–516. [CrossRef]

29. Higgins, K.A.; Hoang, J.K.; Roach, M.C.; Chino, J.; Yoo, D.S.; Turkington, T.G.; Brizel, D.M. Analysis of pretreatment FDG-PET SUV parameters in head-and-neck cancer: Tumor SUVmean has superior prognostic value. *Int. J. Radiat. Oncol. Biol. Phys.* **2012**, *82*, 548–553. [CrossRef]

30. Kim, S.Y.; Roh, J.L.; Kim, M.R.; Kim, J.S.; Choi, S.H.; Nam, S.Y.; Lee, S.W.; Kim, S.B. Use of 18F-FDG PET for primary treatment strategy in patients with squamous cell carcinoma of the oropharynx. *J. Nucl. Med.* **2007**, *48*, 752–757. [CrossRef]

31. Kitajima, K.; Suenaga, Y.; Kanda, T.; Miyawaki, D.; Yoshida, K.; Ejima, Y.; Sasaki, R.; Komatsu, H.; Saito, M.; Otsuki, N.; et al. Prognostic value of FDG PET imaging in patients with laryngeal cancer. *PLoS ONE* **2014**, *9*, e96999. [CrossRef]

32. Komar, G.; Lehtio, K.; Seppanen, M.; Eskola, O.; Levola, H.; Lindholm, P.; Sipila, H.; Seppala, J.; Grenman, R.; Solin, O.; et al. Prognostic value of tumour blood flow, [$^{18}$F]EF5 and [$^{18}$F]FDG PET/CT imaging in patients with head and neck cancer treated with radiochemotherapy. *Eur. J. Nucl. Med. Mol. Imaging* **2014**, *41*, 2042–2050. [CrossRef] [PubMed]

33. Koyasu, S.; Nakamoto, Y.; Kikuchi, M.; Suzuki, K.; Hayashida, K.; Itoh, K.; Togashi, K. Prognostic value of pretreatment 18F-FDG PET/CT parameters including visual evaluation in patients with head and neck squamous cell carcinoma. *AJR Am. J. Roentgenol.* **2014**, *202*, 851–858. [CrossRef] [PubMed]

34. Kunkel, M.; Reichert, T.E.; Benz, P.; Lehr, H.A.; Jeong, J.H.; Wieand, S.; Bartenstein, P.; Wagner, W.; Whiteside, T.L. Overexpression of Glut-1 and increased glucose metabolism in tumors are associated with a poor prognosis in patients with oral squamous cell carcinoma. *Cancer* **2003**, *97*, 1015–1024. [CrossRef] [PubMed]

35. Liao, C.T.; Chang, J.T.; Wang, H.M.; Ng, S.H.; Hsueh, C.; Lee, L.Y.; Lin, C.H.; Chen, I.H.; Huang, S.F.; Cheng, A.J.; et al. Pretreatment primary tumor SUVmax measured by FDG-PET and pathologic tumor depth predict for poor outcomes in patients with oral cavity squamous cell carcinoma and pathologically positive lymph nodes. *Int. J. Radiat. Oncol. Biol. Phys.* **2009**, *73*, 764–771. [CrossRef] [PubMed]

36. Machtay, M.; Natwa, M.; Andrel, J.; Hyslop, T.; Anne, P.R.; Lavarino, J.; Intenzo, C.M.; Keane, W. Pretreatment FDG-PET standardized uptake value as a prognostic factor for outcome in head and neck cancer. *Head Neck* **2009**, *31*, 195–201. [CrossRef]

37. Minn, H.; Lapela, M.; Klemi, P.; Grénman, R.; Leskinen, S.; Lindholm, P.; Bergman, J.; Eronen, E.; Haaparanta, M.; Joensuu, H. Prediction of survival with fluorine-18-fluorodeoxyglucose and PET in head and neck cancer. *J. Nucl. Med.* **1997**, *38*, 1907–1911.

38. Moon, S.H.; Choi, J.Y.; Lee, H.J.; Son, Y.I.; Baek, C.H.; Ahn, Y.C.; Ahn, M.J.; Park, K.; Kim, B.T. Prognostic value of volume-based positron emission tomography/computed tomography in patients with nasopharyngeal carcinoma treated with concurrent chemoradiotherapy. *Clin. Exp. Otorhinolaryngol.* **2015**, *8*, 142–148. [CrossRef]

39. Ng, S.H.; Liao, C.T.; Lin, C.Y.; Chan, S.C.; Lin, Y.C.; Yen, T.C.; Chang, J.T.; Ko, S.F.; Fan, K.H.; Wang, H.M.; et al. Dynamic contrast-enhanced MRI, diffusion-weighted MRI and (18)F-FDG PET/CT for the prediction of survival in oropharyngeal or hypopharyngeal squamous cell carcinoma treated with chemoradiation. *Eur. Radiol.* **2016**, *26*, 4162–4172. [CrossRef] [PubMed]

40. Preda, L.; Conte, G.; Bonello, L.; Giannitto, C.; Travaini, L.L.; Raimondi, S.; Summers, P.E.; Mohssen, A.; Alterio, D.; Cossu Rocca, M.; et al. Combining standardized uptake value of FDG-PET and apparent diffusion coefficient of DW-MRI improves risk stratification in head and neck squamous cell carcinoma. *Eur. Radiol.* **2016**, *26*, 4432–4441. [CrossRef]

41.	Roh, J.L.; Pae, K.H.; Choi, S.H.; Kim, J.S.; Lee, S.; Kim, S.B.; Nam, S.Y.; Kim, S.Y. 2-[18F]-Fluoro-2-deoxy-D-glucose positron emission tomography as guidance for primary treatment in patients with advanced-stage resectable squamous cell carcinoma of the larynx and hypopharynx. *Eur. J. Surg. Oncol.* **2007**, *33*, 790–795. [CrossRef]

42.	Schwartz, D.L.; Harris, J.; Yao, M.; Rosenthal, D.I.; Opanowski, A.; Levering, A.; Ang, K.K.; Trotti, A.M.; Garden, A.S.; Jones, C.U.; et al. Metabolic tumor volume as a prognostic imaging-based biomarker for head-and-neck cancer: Pilot results from Radiation Therapy Oncology Group protocol 0522. *Int. J. Radiat. Oncol. Biol. Phys.* **2015**, *91*, 721–729. [CrossRef] [PubMed]

43.	Suzuki, H.; Kato, K.; Fujimoto, Y.; Itoh, Y.; Hiramatsu, M.; Naganawa, S.; Hasegawa, Y.; Nakashima, T. Prognostic value of (18)F-fluorodeoxyglucose uptake before treatment for pharyngeal cancer. *Ann. Nucl. Med.* **2014**, *28*, 356–362. [CrossRef] [PubMed]

44.	Suzuki-Shibata, S.; Yamamoto, Y.; Yoshida, T.; Mizoguchi, N.; Nonaka, T.; Kubota, A.; Narimatsu, H.; Miyagi, Y.; Kobayashi, T.; Kaneta, T.; et al. Prognostic value of volumetric FDG PET/CT parameters in patients with oral tongue squamous cell carcinoma who were treated by superselective intra-arterial chemoradiotherapy. *Jpn. J. Radiol.* **2017**, *35*, 740–747. [CrossRef] [PubMed]

45.	Torizuka, T.; Tanizaki, Y.; Kanno, T.; Futatsubashi, M.; Naitou, K.; Ueda, Y.; Ouchi, Y. Prognostic value of 18F-FDG PET in patients with head and neck squamous cell cancer. *AJR Am. J. Roentgenol.* **2009**, *192*, W156–W160. [CrossRef] [PubMed]

46.	Xie, P.; Yue, J.B.; Fu, Z.; Feng, R.; Yu, J.M. Prognostic value of 18F-FDG PET/CT before and after radiotherapy for locally advanced nasopharyngeal carcinoma. *Ann. Oncol.* **2010**, *21*, 1078–1082. [CrossRef]

47.	Ahuja, V.; Coleman, R.; Herndon, J.; Patz, E. The prognostic significance of fluorodeoxyglucose positron emission tomography imaging for patients with non-small cell lung carcinoma. *Cancer* **1998**, *83*, 918–924. [CrossRef]

48.	Aoki, M.; Akimoto, H.; Sato, M.; Hirose, K.; Kawaguchi, H.; Hatayama, Y.; Seino, H.; Kakehata, S.; Tsushima, F.; Fujita, H.; et al. Impact of pretreatment whole-tumor perfusion computed tomography and 18F-fluorodeoxyglucose positron emission tomography/computed tomography measurements on local control of non-small cell lung cancer treated with stereotactic body radiotherapy. *J. Radiat. Res.* **2016**, *57*, 533–540. [CrossRef]

49.	Borst, G.R.; Belderbos, J.S.; Boellaard, R.; Comans, E.F.; De Jaeger, K.; Lammertsma, A.A.; Lebesque, J.V. Standardised FDG uptake: A prognostic factor for inoperable non-small cell lung cancer. *Eur. J. Cancer* **2005**, *41*, 1533–1541. [CrossRef]

50.	Carvalho, S.; Leijenaar, R.T.; Velazquez, E.R.; Oberije, C.; Parmar, C.; van Elmpt, W.; Reymen, B.; Troost, E.G.; Oellers, M.; Dekker, A.; et al. Prognostic value of metabolic metrics extracted from baseline positron emission tomography images in non-small cell lung cancer. *Acta Oncol.* **2013**, *52*, 1398–1404. [CrossRef]

51.	Cerfolio, R.J.; Bryant, A.S.; Ohja, B.; Bartolucci, A.A. The maximum standardized uptake values on positron emission tomography of a non-small cell lung cancer predict stage, recurrence, and survival. *J. Thorac. Cardiovasc. Surg.* **2005**, *130*, 151–159. [CrossRef]

52.	Chen, H.; Chiu, N.; Su, W.; Guo, H.; Lee, B. Prognostic value of whole-body total lesion glycolysis at pretreatment FDG PET/CT in non-small cell lung cancer. *Radiology* **2012**, *264*, 559–566. [CrossRef] [PubMed]

53.	Clarke, K.; Taremi, M.; Dahele, M.; Freeman, M.; Fung, S.; Franks, K.; Bezjak, A.; Brade, A.; Cho, J.; Hope, A.; et al. Stereotactic body radiotherapy (SBRT) for non-small cell lung cancer (NSCLC): Is FDG-PET a predictor of outcome? *Radiother. Oncol.* **2012**, *104*, 62–66. [CrossRef] [PubMed]

54.	Hamamoto, Y.; Sugawara, Y.; Inoue, T.; Kataoka, M.; Ochi, T.; Takahashi, T.; Sakai, S. Relationship between pretreatment FDG uptake and local control after stereotactic body radiotherapy in stage I non-small-cell lung cancer: The preliminary results. *Jpn. J. Clin. Oncol.* **2011**, *41*, 543–547. [CrossRef] [PubMed]

55.	Hofheinz, F.; Hoff, J.; Steffen, I.G.; Lougovski, A.; Ego, K.; Amthauer, H.; Apostolova, I. Comparative evaluation of SUV, tumor-to-blood standard uptake ratio (SUR), and dual time point measurements for assessment of the metabolic uptake rate in FDG PET. *EJNMMI Res.* **2016**, *6*, 53. [CrossRef]

56.	Horne, Z.D.; Clump, D.A.; Vargo, J.A.; Shah, S.; Beriwal, S.; Burton, S.A.; Quinn, A.E.; Schuchert, M.J.; Landreneau, R.J.; Christie, N.A.; et al. Pretreatment SUVmax predicts progression-free survival in early-stage non-small cell lung cancer treated with stereotactic body radiation therapy. *Radiat. Oncol.* **2014**, *9*, 41. [CrossRef]

57.　Hyun, S.H.; Ahn, H.K.; Ahn, M.J.; Ahn, Y.C.; Kim, J.; Shim, Y.M.; Choi, J.Y. Volume-Based Assessment With 18F-FDG PET/CT Improves Outcome Prediction for Patients With Stage IIIA-N2 Non-Small Cell Lung Cancer. *AJR Am. J. Roentgenol.* **2015**, *205*, 623–628. [CrossRef]

58.　Imamura, Y.; Azuma, K.; Kurata, S.; Hattori, S.; Sasada, T.; Kinoshita, T.; Okamoto, M.; Kawayama, T.; Kaida, H.; Ishibashi, M.; et al. Prognostic value of SUVmax measurements obtained by FDG-PET in patients with non-small cell lung cancer receiving chemotherapy. *Lung Cancer* **2011**, *71*, 49–54. [CrossRef]

59.　Jiang, X.E.; Xu, T.; Wei, Q.; Li, P.; Gomez, D.R.; Court, L.E.; Liao, Z. DNA repair capacity correlates with standardized uptake values from (18)F-fluorodeoxyglucose positron emission tomography/CT in patients with advanced non-small-cell lung cancer. *Chronic Dis. Transl. Med.* **2018**, *4*, 109–116. [CrossRef]

60.　Kohutek, Z.A.; Wu, A.J.; Zhang, Z.; Foster, A.; Din, S.U.; Yorke, E.D.; Downey, R.; Rosenzweig, K.E.; Weber, W.A.; Rimner, A. FDG-PET maximum standardized uptake value is prognostic for recurrence and survival after stereotactic body radiotherapy for non-small cell lung cancer. *Lung Cancer* **2015**, *89*, 115–120. [CrossRef]

61.　Lee, H.Y.; Lee, K.S.; Park, J.; Han, J.; Kim, B.T.; Kwon, O.J.; Ahn, Y.C.; Ahn, M.J.; Park, K.; Kim, J.; et al. Baseline SUVmax at PET-CT in stage IIIA non-small-cell lung cancer patients undergoing surgery after neoadjuvant therapy: Prognostic implication focused on histopathologic subtypes. *Acad. Radiol.* **2012**, *19*, 440–445. [CrossRef]

62.　Nair, V.J.; MacRae, R.; Sirisegaram, A.; Pantarotto, J.R. Pretreatment [18F]-fluoro-2-deoxy-glucose positron emission tomography maximum standardized uptake value as predictor of distant metastasis in early-stage non-small cell lung cancer treated with definitive radiation therapy: Rethinking the role of positron emission tomography in personalizing treatment based on risk status. *Int. J. Radiat. Oncol. Biol. Phys.* **2014**, *88*, 312–318. [CrossRef] [PubMed]

63.　Nawara, C.; Rendl, G.; Wurstbauer, K.; Lackner, B.; Rettenbacher, L.; Datz, L.; Studnicka, M.; Sedlmayer, F.; Pirich, C. The impact of PET and PET/CT on treatment planning and prognosis of patients with NSCLC treated with radiation therapy. *Q. J. Nucl. Med. Mol. Imaging* **2012**, *56*, 191–201. [PubMed]

64.　Pyka, T.; Bundschuh, R.A.; Andratschke, N.; Mayer, B.; Specht, H.M.; Papp, L.; Zsoter, N.; Essler, M. Textural features in pre-treatment [F18]-FDG-PET/CT are correlated with risk of local recurrence and disease-specific survival in early stage NSCLC patients receiving primary stereotactic radiation therapy. *Radiat. Oncol.* **2015**, *10*, 100. [CrossRef]

65.　Sasaki, R.; Komaki, R.; Macapinlac, H.; Erasmus, J.; Allen, P.; Forster, K.; Putnam, J.B.; Herbst, R.S.; Moran, C.A.; Podoloff, D.A.; et al. [18F]fluorodeoxyglucose uptake by positron emission tomography predicts outcome of non-small-cell lung cancer. *J. Clin. Oncol.* **2005**, *23*, 1136–1143. [CrossRef] [PubMed]

66.　Shirai, K.; Abe, T.; Saitoh, J.I.; Mizukami, T.; Irie, D.; Takakusagi, Y.; Shiba, S.; Okano, N.; Ebara, T.; Ohno, T.; et al. Maximum standardized uptake value on FDG-PET predicts survival in stage I non-small cell lung cancer following carbon ion radiotherapy. *Oncol. Lett.* **2017**, *13*, 4420–4426. [CrossRef]

67.　Sugawara, Y.; Quint, L.; Iannettoni, M.; Orringer, M.; Russo, J.; Recker, B.; Saran, P.; Wahl, R. Does the FDG uptake of primary non-small cell lung cancer predict prognosis?: A work in progress. *Clin. Positron Imaging* **1999**, *2*, 111–118. [CrossRef]

68.　Takeda, A.; Yokosuka, N.; Ohashi, T.; Kunieda, E.; Fujii, H.; Aoki, Y.; Sanuki, N.; Koike, N.; Ozawa, Y. The maximum standardized uptake value (SUVmax) on FDG-PET is a strong predictor of local recurrence for localized non-small-cell lung cancer after stereotactic body radiotherapy (SBRT). *Radiother. Oncol.* **2011**, *101*, 291–297. [CrossRef]

69.　Takeda, A.; Sanuki, N.; Fujii, H.; Yokosuka, N.; Nishimura, S.; Aoki, Y.; Oku, Y.; Ozawa, Y.; Kunieda, E. Maximum standardized uptake value on FDG-PET is a strong predictor of overall and disease-free survival for non-small-cell lung cancer patients after stereotactic body radiotherapy. *J. Thorac. Oncol.* **2014**, *9*, 65–73. [CrossRef]

70.　Takeda, K.; Takanami, K.; Shirata, Y.; Yamamoto, T.; Takahashi, N.; Ito, K.; Takase, K.; Jingu, K. Clinical utility of texture analysis of 18F-FDG PET/CT in patients with Stage I lung cancer treated with stereotactic body radiotherapy. *J. Radiat. Res.* **2017**, *58*, 862–869. [CrossRef]

71.　Ulger, S.; Demirci, N.Y.; Eroglu, F.N.; Cengiz, H.H.; Tunc, M.; Tatci, E.; Yilmaz, U.; Cetin, E.; Avci, E.; Cengiz, M. High FDG uptake predicts poorer survival in locally advanced nonsmall cell lung cancer patients undergoing curative radiotherapy, independently of tumor size. *J. Cancer Res. Clin. Oncol.* **2014**, *140*, 495–502. [CrossRef]

72.    Vansteenkiste, J.; Stroobants, S.; Dupont, P.; De Leyn, P.; Verbeken, E.; Deneffe, G.; Mortelmans, L.; Demedts, M. Prognostic importance of the standardized uptake value on 18F-Fluoro-2-deoxy-glucose-positron emission tomography scan in non-small-cell lung cancer: An analysis of 125 cases. *J. Clin. Oncol.* **1999**, *17*, 3201–3206. [CrossRef] [PubMed]

73.    Vesselle, H.; Freeman, J.D.; Wiens, L.; Stern, J.; Nguyen, H.Q.; Hawes, S.E.; Bastian, P.; Salskov, A.; Vallieres, E.; Wood, D.E. Fluorodeoxyglucose uptake of primary non-small cell lung cancer at positron emission tomography: New contrary data on prognostic role. *Clin. Cancer Res.* **2007**, *13*, 3255–3263. [CrossRef] [PubMed]

74.    Vu, C.C.; Matthews, R.; Kim, B.; Franceschi, D.; Bilfinger, T.V.; Moore, W.H. Prognostic value of metabolic tumor volume and total lesion glycolysis from 18F-FDG PET/CT in patients undergoing stereotactic body radiation therapy for stage I non-small-cell lung cancer. *Nucl. Med. Commun.* **2013**, *34*, 959–963. [CrossRef] [PubMed]

75.    Xiang, Z.L.; Erasmus, J.; Komaki, R.; Cox, J.D.; Chang, J.Y. FDG uptake correlates with recurrence and survival after treatment of unresectable stage III non-small cell lung cancer with high-dose proton therapy and chemotherapy. *Radiat. Oncol.* **2012**, *7*, 144. [CrossRef]

76.    Yilmaz, U.; Batum, O.; Koparal, H.; Ozbilek, E.; Kirakli, E. Prognostic value of primary tumor SUVmax on pre-treatment (18)F-FDG PET/CT imaging in patients with stage iii non-small cell lung cancer. *Rev. Esp. Med. Nucl. Imagen Mol.* **2018**. [CrossRef]

77.    Port, J.L.; Andrade, R.S.; Levin, M.A.; Korst, R.J.; Lee, P.C.; Becker, D.E.; Altorki, N.K. Positron emission tomographic scanning in the diagnosis and staging of non-small cell lung cancer 2 cm in size or less. *J. Thorac. Cardiovasc. Surg.* **2005**, *130*, 1611–1615. [CrossRef]

78.    Kim, S.; Oh, S.; Kim, J.S.; Kim, Y.K.; Kim, K.H.; Oh, D.H.; Lee, D.H.; Jeong, W.J.; Jung, Y.H. Prognostic value of FDG PET/CT during radiotherapy in head and neck cancer patients. *Radiat. Oncol. J.* **2018**, *36*, 95–102. [CrossRef]

79.    Turner, E.H.; Matthews, A.M.; Linardatos, E.; Tell, R.A.; Rosenthal, R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N. Engl. J. Med.* **2008**, *358*, 252–260. [CrossRef]

80.    Button, K.S.; Ioannidis, J.P.; Mokrysz, C.; Nosek, B.A.; Flint, J.; Robinson, E.S.; Munafo, M.R. Power failure: Why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* **2013**, *14*, 365–376. [CrossRef]

81.    Collins, G.S.; Reitsma, J.B.; Altman, D.G.; Moons, K.G. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *BMC Med.* **2015**, *13*, 1. [CrossRef]

82.    Royston, P.; Altman, D.G.; Sauerbrei, W. Dichotomizing continuous predictors in multiple regression: A bad idea. *Stat. Med.* **2006**, *25*, 127–141. [CrossRef] [PubMed]