

BMJ Open Positron emission tomography (PET) for prediction of glioma histology: protocol for an individual-level data meta-analysis of test performance

Nikolaos A Trikalinos,¹ Takashi Nihashi,^{2,3} Evangelos Evangelou,^{4,5} Teruhiko Terasawa⁶

To cite: Trikalinos NA, Nihashi T, Evangelou E, *et al*. Positron emission tomography (PET) for prediction of glioma histology: protocol for an individual-level data meta-analysis of test performance. *BMJ Open* 2018;**8**:e020187. doi:10.1136/bmjopen-2017-020187

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2017-020187>).

Received 1 November 2017
Revised 27 December 2017
Accepted 18 January 2018



¹Department of Oncology, Washington University in St Louis School of Medicine, St Louis, Missouri, USA

²Department of Radiology, Nagoya University Graduate School of Medicine, Nagoya, Aichi, Japan

³Department of radiology, Komaki City Hospital, Jobushi, Japan

⁴Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece

⁵Department of Epidemiology and Biostatistics, Imperial College London, London, UK

⁶Section of General Internal Medicine, Department of Emergency and General Internal Medicine, Fujita Health University School of Medicine, Toyoake, Aichi, Japan

Correspondence to

Dr Nikolaos A Trikalinos;
ntrikalinos@wustl.edu

ABSTRACT

Introduction Gliomas, the most commonly diagnosed primary brain tumours, are associated with varied survivals based, in part, on their histological subtype. Therefore, accurate pretreatment tumour grading is essential for patient care and clinical trial design.

Methods and analysis We will perform an individual-level data meta-analysis of published studies to evaluate the ability of different types of positron emission tomography (PET) to differentiate high from low-grade gliomas. We will search PubMed and Scopus from inception through 30 July 2017 with no language restriction and full-text evaluation of potentially relevant articles. We will choose studies that assess PET using 18-Fluodeoxyglucose (18F-FDG), L-[Methyl-(¹¹C)]Methionine (11C-MET), 18F-Fluoro-Ethyl-Tyrosine (18F-FET) or (18) F-Fluorothymidine (18F-FLT) for grading, verified with histological confirmation. We will include both prospective and retrospective studies. Bias will be assessed by two reviewers with the Quality Assessment of Diagnostic Accuracy Studies-2 tool and as per method described by Deeks *et al*.

Ethics and dissemination Ethics approval was not applicable, as this is a meta-analytic study. Results of the analysis will be submitted for publication in a peer-reviewed journal.

PROSPERO registration number CRD42017078649.

BACKGROUND

Gliomas are among the most commonly diagnosed primary brain tumours with an estimated annual incidence of over 20 000 in the USA¹. Based on clinical, histological and molecular characteristics, gliomas can be broadly divided into two major clinical subcategories: low grade and high grade.

High-grade gliomas (also called malignant gliomas) are rapidly growing tumours and include glioblastomas (grade IV), anaplastic astrocytomas (grade III), mixed anaplastic oligoastrocytomas (grade III) and anaplastic oligodendrogliomas (grade III). Despite recent advances in multimodality therapies

Strengths and limitations of this study

- This is a first of its kind individual patient data meta-analysis (IPD-MA) aiming to establish the diagnostic accuracy of positron emission tomography with various tracers for grading of glioma. Individual level meta-analysis with pooling of data can provide more statistical power to determine differences between imaging modalities.
- Variability of data obtained from external sources and limited number of patients.
- Inherent limitations of IPD-MA for data interpretation/findings might be speculative.

including temozolomide, high-grade gliomas remain incurable with a median survival of less than 3 years for glioblastomas and less than 5 years for anaplastic gliomas².

In contrast, the low-grade gliomas include astrocytomas (grade II), oligodendrogliomas (grade II) and oligoastrocytomas (grade II) and are indolent, with a median survival over 5 years.³ Treatment of low-grade gliomas is evolving and includes maximal tumour resection with the option of radiation and chemotherapy.⁴

From the above, it is evident that accurate pretherapy histological grading is of paramount importance. In the usual clinical setting, histology is clarified with conventional CT or MRI, image-guided biopsy/subtotal surgical resection or near total resection. Unfortunately, apart from complete resection, other approaches are error prone: imaging does not provide usable tissue for final diagnosis. Biopsies and partial resections can misguide, as gliomas can demonstrate histological heterogeneity due to histopathological progression; a tumour often has both low-grade and high-grade components. Unless total surgical resection is performed,

histological results depend on which part of the tumour is sampled.

As much as it is desirable, extensive tumour resection is not always feasible in frail patients or those with tumours located adjacent to critical structures. Therefore, in order to improve tumour grading either non-invasively or by better, targeted biopsies, one needs to be able to supplement morphological information obtained from conventional imaging with tumour-specific functional information.

Positron emission tomography (PET) is a useful imaging modality that provides additional metabolic information to CT-based or MRI-based morphological characterisation of tumours. Its usefulness has been proven in lung cancer⁵ and aggressive lymphomas.⁶ Several studies have evaluated PET using various tracers such as 18-Fluorodeoxyglucose (18F-FDG), L-[Methyl-(¹¹C)] Methionine (¹¹C-MET), 18F-Fluoro-Ethyl-Tyrosine (18F-FET) or (18)F-Fluorothymidine (18F-FLT) for pretherapy histological prediction to differentiate high-grade from low-grade glioma and reported promising results.⁷⁻⁹ The sample sizes of these studies, however, are typically small, resulting in imprecise estimates of diagnostic accuracy and use heterogeneous designs/protocols for PET assessment, making interpretation of the published data difficult. Often, the positivity cut-off criteria are defined post hoc to calculate the best pairs of sensitivity and specificity for each study and lack generalisability. In this situation, aggregate-level data meta-analysis (ALD-MA) typically calculates overestimated summary statistics.

A useful approach to the above problem to perform a meta-analysis based on individual-level data (individual-level data meta-analysis (ILD-MA)). This technique employs a predefined uniform cut-off value to estimate test performance measures for all included studies and combines their results. ILD-MA can also evaluate the effect of specific factors on test performance at the level of individual, simultaneously accounting for between-study variations. We therefore planned an ILD-MA to provide a comprehensive overview and quantitative synthesis of information on the test performance of PET for this purpose.

METHODS

This meta-analysis is an extension of a part of a series of systematic reviews on PET in clinical management of patients with glioma.^{10 11} Although these reviews share a common literature search until June 2011, each project employs an independent prespecified research protocol with standard systematic review methodologies and assesses mutually exclusive research objectives. Interim results of an earlier related study of this research project have been presented at an international meeting¹² but never published as full text.

Literature search

We will use our literature database of publications on PET assessed for patients with glioma. We established the database based on the searches of PubMed and Scopus from inception through 30 June 2011 with no language restriction and full-text evaluation of potentially relevant articles found through abstract screening. The complete search strategy and full list of the database is reported elsewhere¹¹ and also available as a online Supplementary file. We will update the searches until 30 July 2017 and examine the reference lists of eligible studies and relevant review articles.

One reviewer (TN or NAT) will screen abstracts and at least two of three investigators (TT, NAT and TT) examine full-text articles of potentially eligible citations. Discrepancies will be resolved by consensus.

Inclusion and exclusion criteria

We will select studies that assess PET using 18F-FDG, ¹¹C-MET, 18F-FET or 18F-FLT for predicting glioma grading, verified with histological confirmation by surgery or biopsy to be eligible. We have selected these four particular tracers before conducting this research on the basis of an empirical evaluation of published studies of PET on glioma¹¹ and recently published narrative reviews.^{13 14} We will include both prospective and retrospective studies. We will define pathological confirmation (either by biopsy or surgical resection) as the acceptable reference standard and explicitly exclude studies (or individual patients in a study) in which (or for whom, respectively) pathological confirmation is not performed. We will consider the total surgical resection as the (nearly) perfect reference standard, whereas (stereotactic) biopsy will be deemed to be the imperfect reference standard. We will include any language publications that evaluated at least 10 patients for whom PET scanning and histopathological confirmation is successfully performed. We will exclude editorials, comments, letters to the editor and review articles. When multiple publications with potentially overlapping patient populations are available, we will only include the publication with the largest sample size.

We will contact study authors by email if studies do not report adequate information on PET and histological results in the participant level or if ILD data are not presented in the paper. We will consider our request to be rejected if two email request reminders separately sent 14 days after the initial contact attempt are rejected. Even in this case, we will allow for the inclusion of a study report in which quantitative data is not reported but the digital extraction (ie, data extraction by using a digitizer from its published graphical presentations) is feasible. Otherwise, we will exclude these studies. Data will be kept in a shared secure folder accessible by all co-authors.

Data extraction

One reviewer (TN or NAT) will extract descriptive data from each eligible study. Another one non-overlapping investigator (TN, NAT or TT) will verify all the extracted

data. We will extract the following published descriptive information from eligible studies: first author, year of publication, journal, patient demographics and clinical characteristics, therapeutic interventions in the case of post-therapy or recurrence assessment, technical specifications of PET and interpretation of PET results.

One reviewer (TN or NAT) will extract published quantitative data regarding imaging results (ie, visual assessment and quantitative assessment such as standard uptake values (SUVs) or tumour-to-normal uptake ratio (T/N ratio)) and final diagnoses (ie, histopathological subtype and grading) at the individual level. Another one reviewer (TT) will verify all the data. We will exclude any cases in which PET scanning is unsuccessful (and thus results are arbitrary imputed post hoc). We will also exclude any cases in which alternative reference standards such as clinical follow-up instead of pathological confirmation is used to determine histological grading.

If reported, we will also extract the following individual-level variables as the candidate effect modifiers to be evaluated in meta-regression: age, sex, clinical scenario (ie, primary diagnosis vs post-therapy/recurrence assessment) and tissue sampling methods (ie, how the pathological specimens are obtained, either biopsy, partial or total resection).

Assessment of risk of bias

To assess the risk of bias and applicability of each study, two reviewers (TN and NAT) will independently assess patient selection, index test, reference standard and their flow and timing based on the revised Quality Assessment of Diagnostic Accuracy Studies 2. Methods to detect publication bias are not very reliable when used in diagnostic accuracy data, especially in case of heterogeneity, but we will use the method of Deeks *et al*¹⁵ that has been shown to be the least biased. Discrepant ratings will be resolved by consensus. The complete list of operational definitions to rate each item is available from the authors on request.

Data synthesis

We will use a bivariate model to obtain an estimate of the summary sensitivity and specificity with their corresponding CIs. We will fit a two-level generalised mixed regression model conditional on the sensitivity and specificity of each study and a bivariate normal model for the sensitivity and specificity between studies¹⁶. We will calculate summary positive and negative likelihood ratios (LRs) based on the summary sensitivity and specificity estimates. Positive LR (LR+) is the ratio of sensitivity over (1-specificity), whereas negative LR (LR-) is defined as the ratio of (1-sensitivity) over specificity. The discriminating ability of a diagnostic test is better with higher LR+ and lower LR-. A good diagnostic test typically has LR+ >5.0 and LR- <0.2. We will also construct a hierarchical summary receiver operating characteristic curve (HSROC) based on the parameters of the fitted model. We will assess between-study heterogeneity visually using forest plots and also by plotting sensitivity and specificity

in the ROC space. We will construct 95% credible regions for summary sensitivity and specificity from the estimated parameters as proposed.¹⁷

Subgroup analysis and meta-regression

To explore heterogeneity, we will perform subgroup analyses and if feasible, the statistical differences among mutually exclusive subgroups will be assessed by univariable meta-regression. We will add a candidate modifier (as either 0 (absence) or 1 (presence)) to the test performance parameters (ie, sensitivity and specificity parameters or diagnostic OR and threshold parameters, respectively) jointly in the bivariate random-effects model or binormal random-effects model, respectively.

We will record: publication year, study design (ie, prospective vs retrospective) as the study-level covariates, age, sex, clinical scenario (ie, primary diagnosis vs post-therapy/recurrence assessment) and tissue sampling methods (ie, biopsy vs partial or subtotal resection vs total resection).

Comparisons among different PET methodologies

We will compare the test performance among alternative PET tracers and also those based on different imaging assessment protocols (ie, visual assessment vs quantitative assessment, either T/N ratios or SUVs). Regarding the T/N ratios, we will operationally define three categories based on the specified referent tissues: mean or median uptakes in the grey matter (GM), those in the white matter (WM) and other miscellaneous (MISC) methods including mean or median uptakes in the contralateral corresponding anatomical sites or those in the adjacent normal tissue regardless of GM or WM.

For indirect comparisons, we will visually compare the constructed summary ROC curves and the credible regions of sensitivity and specificity. We will also statistically assess the differences by univariable meta-regression using the tracers or assessment methods as the covariate being incorporated into the meta-analytic model. These results, however, need to be interpreted carefully because findings from the indirect comparisons may be only speculative. To tackle this issue, we will also perform these analyses limiting only to comparative studies that assess multiple tracers or methods for the same participants, from which direct comparisons can be performed.

We will conduct all analyses using STATA V.14/SE and OpenBUGS V.3.2.3 (members of OpenBUGS Project Management Group; see www.openbugs.net). All tests will be two-sided and statistical significance will be defined as a P value <0.05.

DISCUSSION

Differentiating high from low-grade malignancies has significant prognostic information for a variety of tumours; gliomas are not an exception. While an optimally resected biopsy specimen is always preferable to indirect evidence, imaging can provide helpful adjunct

information. Moreover, functional imaging, such as PET with select biotracers, has the potential to become a powerful tool in assessment of patients and influence treatment decisions. By conducting an IPD meta-analysis, this study hopes to clarify the diagnostic accuracy of PET/CT with various tracers in differentiating glioma histology and establish thresholds, on which further studies can be performed.

Contributors NAT and TN were responsible for the abstract screening and assessment of the risk of bias and applicability of each study. NAT, TN and TT were responsible for the full-text article examination and data extraction. EE and TT did the statistical analysis. All authors have equally contributed to the final version of the protocol, reviewed and approved it.

Funding This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

1. Surveillance, Epidemiology and End Results (SEER) Data. 2017 <https://seer.cancer.gov/registries/>.
2. National Comprehensive Cancer Network 2017 <https://www.nccn.org/>
3. Claus EB, Walsh KM, Wiencke JK, *et al.* Survival and low-grade glioma: the emergence of genetic information. *Neurosurg Focus* 2015;38:E6.
4. Buckner JC, Shaw EG, Pugh SL, *et al.* Radiation plus procarbazine, ccnu, and vincristine in low-grade glioma. *N Engl J Med* 2016;374:1344–55.
5. Sheikhabaei S, Mena E, Yanamadala A, *et al.* The value of FDG PET/CT in treatment response assessment, follow-up, and surveillance of lung cancer. *AJR Am J Roentgenol* 2017;208:420–33.
6. Adams HJ, Kwee TC. Pretransplant FDG-PET in aggressive non-Hodgkin lymphoma: systematic review and meta-analysis. *Eur J Haematol* 2017;98:337–47.
7. Miyake K, Shinomiya A, Okada M, *et al.* Usefulness of FDG, MET and FLT-PET studies for the management of human gliomas. *J Biomed Biotechnol* 2012;2012:1–11.
8. Sweeney R, Polat B, Samnick S, *et al.* O-(2-[(18F)fluoroethyl]-L-tyrosine uptake is an independent prognostic determinant in patients with glioma referred for radiation therapy. *Ann Nucl Med* 2014;28:154–62.
9. Thon N, Kunz M, Lemke L, *et al.* Dynamic 18F-FET PET in suspected WHO grade II gliomas defines distinct biological subgroups with different clinical courses. *Int J Cancer* 2015;136:2132–45.
10. Nihashi T, Dahabreh IJ, Terasawa T. Diagnostic accuracy of PET for recurrent glioma diagnosis: a meta-analysis. *AJNR Am J Neuroradiol* 2013;34:944–50.
11. Nihashi T, Dahabreh IJ, Terasawa T. PET in the clinical management of glioma: evidence map. *AJR Am J Roentgenol* 2013;200:W65 4–W660.
12. Nihashi T, Dahabreh IJ, Terasawa T. Positron emission tomography (PET) to differentiate high-grade from low-grade glioma: a meta-analysis of test performance (Abstract). *J Neuroimaging* 2014;23:280.
13. la Fougère C, Suchorska B, Bartenstein P, *et al.* Molecular imaging of gliomas with PET: opportunities and limitations. *Neuro Oncol* 2011;13:806–19.
14. Boellaard R. Need for standardization of 18F-FDG PET/CT for treatment response assessments. *J Nucl Med* 2011;52(Suppl 2):93S–100.
15. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* 2005;58:882–93.
16. Riley RD, Dodd SR, Craig JV, *et al.* Meta-analysis of diagnostic test studies using individual patient data and aggregate data. *Stat Med* 2008;27:6111–36.
17. Harbord RM, Deeks JJ, Egger M, *et al.* A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 2007;8:239–51.