Comparing the accuracy of quantitative versus qualitative analyses of interim PET to prognosticate Hodgkin lymphoma: a systematic review protocol of diagnostic test accuracy

Vít Procházka,1 Miloslav Klugar2,3 Veronika Bachanova,4 Jitka Klugarová,1,3 Dagmar Tučková,2,3 Tomáš Papajík1

ABSTRACT

Introduction: Hodgkin lymphoma is an effectively treated malignancy, yet 20% of patients relapse or are refractory to front-line treatments with potentially fatal outcomes. Early detection of poor treatment responders is crucial for appropriate application of tailored treatment strategies. Tumour metabolic imaging of Hodgkin lymphoma using visual (qualitative) 18-fluorodeoxyglucose positron emission tomography (FDG-PET) is a gold standard for staging and final outcome assessment, but results gathered during the interim period are less accurate. Analysis of continuous metabolic-morphological data (quantitative) FDG-PET may enhance the robustness of interim disease monitoring, and help to improve treatment decision-making processes. The objective of this review is to compare diagnostic test accuracy of quantitative versus qualitative interim FDG-PET in the prognostication of patients with Hodgkin lymphoma.

Methods: The literature on this topic will be reviewed in a 3-step strategy that follows methods described by the Joanna Briggs Institute (JBI). First, MEDLINE and EMBASE databases will be searched. Second, listed databases for published literature (MEDLINE, Tripdatabase, Pedro, EMBASE, the Cochrane Central Register of Controlled Trials and WoS) and unpublished literature (Open Grey, Current Controlled Trials, MedNar, ClinicalTrials.gov, Cos Conference Papers Index and International Clinical Trials Registry Platform of the WHO) will be queried. Third, 2 independent reviewers will analyse titles, abstracts and full texts, and perform hand search of relevant studies, and then perform critical appraisal and data extraction from selected studies using the DATARI tool (JBI). If possible, a statistical meta-analysis will be performed on pooled sensitivity and specificity data gathered from the selected studies. Statistical heterogeneity will be assessed. Funnel plots, Begg’s rank correlations and Egger’s regression tests will be used to detect and/or correct publication bias.

Ethics and dissemination: The results will be disseminated by publishing in a peer-reviewed journal. Ethical assessment will not be needed; only existing sources of literature will be searched.

Trial registration number: CRD42016027953.

INTRODUCTION

Background

Classical Hodgkin lymphoma (cHL) is the most common lymphoid malignancy affecting patients below the age of 30. Incidence rates of cHL in the USA and Central Europe are comparable, with 2.7 new cases per 100 000 men and women per year, and rates trending upwards.1 2 Despite high cure rates and effective treatments for cHL, 20% of patients relapse or are refractory to front-line therapies. About 15% of these patients die within 5 years of diagnosis.3 Overall outcomes are unsatisfactory for patients with relapsed/refractory Hodgkin lymphoma (HL) who proceed to high-dose therapies and autologous stem cell transplants (SCTs). About 40–50% of SCT recipients relapse and require additional treatments.4 Given our entry into the era of novel ‘targeted’ drugs and immune modulators, identification of poor front-line treatment responders is a growing concern.5

Implementations of modern imaging methods such as positron emission tomography (PET)/CT have provided the capability to precisely assess tumour metabolic activity concurrent with an exact measurement of tumour burden. HL has been described as ubiquitously 18-fluorodeoxyglucose (18FDG)-avid. Revised response criteria for malignant lymphoma have therefore included tumour metabolic activity as a key parameter for determining remission status. Historically, complete metabolic responses have been assessed visually using either binary (positive/negative) or semiquantitative (Deauville) scales.6 7 FGG-PET is an inherently quantitative method that generates large amounts of metabolic and morphological data. Visual binary and semiquantitative PET analyses do not
include quantitative and volumetric parameters (eg, total metabolic volume (TMV), total lesion glycolysis (TLG) or maximal standardised uptake volume (SUVmax)), and may be observer-biased. Recent studies have encouraged quantitative FDG-PET (QT PET) analyses to serve as novel biomarkers for staging and assessment of early (referred to as interim) and final malignant lymphoma tumour responses to treatments. FDG-PET-based tumour metabolic activities at diagnoses were demonstrated to predict survival in HL and non-HL cases. Quantitative metabolic parameters have shown superiority when compared with semiquantitative assessments in untreated HL and primary diffuse large B-cell lymphoma cases. Given that personalised medicine has strongly emphasised individualised treatment approaches for all patients, evaluation of chemosensitivity is needed during oncology treatment. For cHL, those at risk of treatment failure may be identified by QT PET after a few cycles of therapy (referred to as ‘interim PET’).

Early (interim) visual assessment of cHL tumour metabolism has shown superiority when compared with standard prognostic scoring methods. Meta-analysis of these studies showed that interim FDG-PET had high prognostic value for identifying treatment failure. Unfortunately, interim PET has not been implemented in routine clinical practice due to the moderate quality of previous evidence and interstudy heterogeneity. Moreover, interim FDG-PET could not be used as a tool for tailored therapy as shown by results of two systematic reviews published by Sickinger et al. One way to circumvent these barriers is to analyse QT PET results as a method of improving interim PET diagnostic accuracy and reproducibility. Several previous studies have investigated QT PET parameters during the interim period. For example, Rossi and colleagues demonstrated that interim PET after two cycles of anthracycline-based chemotherapy captured SUVmax (ΔSUVmax) reductions as large as 71% below baseline. This technique identified positive responders with greater precision than did visual assessment alone. Quantitative ΔSUVmax achieved 85% diagnostic accuracy compared with just 76% from the visual method. Furthermore, positive predictive value increased by 24% (from 46% to 70%) when the ΔSUVmax method was used in lieu of visual inspection. Additionally, Tseng and colleagues analysed 30 patients with cHL who were scanned at diagnosis and again during treatment. In this study, TMV, SUVmax and TLG were calculated together to determine cumulative changes during treatment regimens. Quantitative interim PET predicted both progression-free and overall survival rates.

To the best of our knowledge, a systematic review of the role of quantitative interim PET in patients with cHL has yet to be established. We hypothesise that measurements of quantitative tumour characteristics will improve diagnostic and predictive accuracy of interim PET. Thus, more successful candidates will be identified by interim PET for novel treatment approaches. The systematic review protocol described here has an extensive search strategy. It seeks to clarify the role of quantitative interim PET in cHL prognostication and influence practice by informing physician recommendations. Preliminary searches as of January 2016 were conducted using the MEDLINE, Prospero, JBI Library and Cochrane databases to establish whether previous systematic reviews on this topic were publicly available. No systematic reviews or guidelines related to this issue were discovered.

Objective

The objective of this review will be to compare diagnostic test accuracies between quantitative and qualitative interim PET methods with the aim of improving cHL prognostication.

METHODS AND ANALYSIS

Methods

This systematic review protocol was developed according to: (1) the Preferred Reporting Items for Systematic Reviews and Meta-Analysis Protocols (PRISMA-P), and (2) the Joanna Briggs Institute (JBI) methodology for systematic reviews of diagnostic test accuracy. It has been enrolled with the PROSPERO prospective register of systematic reviews: CRD42016027953.

Study eligibility

Types of participants

The systematic review will consider all studies that investigated adult cHL (determined with the WHO diagnostic criteria), who were treated according to the current international guidelines. Studies that included adolescents (≤18 years old) will be excluded.

Index test

The systematic review will consider all studies that measure one or more of the following as an index test: QT PET, quantitative evaluation of interim FDG-PET by metabolic tumour volume, TLG or SUVmax. Only studies which used standardised international criteria for interim FDG-PET interpretation will be analysed.

Reference test

The systematic review will consider studies that perform qualitative FDG-PET (QL PET) or visual evaluation of interim PET as a reference test.

Diagnosis of interest

The systematic review will consider studies that evaluate prognostic accuracy of QT PET in patients with cHL as calculated by changes in negative and positive predictive values when compared with QL PET.

Types of studies

The systematic review will only include diagnostic cross-sectional study designs.
Search strategy

A search strategy will be developed using medical subject headings (eg, MeSH for MEDLINE) and then adopted to query each database. Keywords related to the overarching topic will also be identified. The search strategy seeks to identify and include both published and unpublished work, and will therefore use a three-step search strategy. First, limited searches of MEDLINE and EMBASE will be undertaken, followed by analyses of keywords contained in the title, abstract and the index terms used to describe an article. Second, all identified keywords and index terms will be searched across all relevant databases. Third, reference lists from the newly identified reports and articles will be searched for additional studies. All studies with title and abstract in English will be considered for inclusion, regardless of the language used in the body of the manuscript. Studies published with no time restriction will also be considered for inclusion.

The databases to be searched include:
- MedLine@Ovid, MEDLINE(R), Tripdatabase, Pedro, EMBASE, Cochrane Central Register of Controlled Trials, CINAHL and Web of Science.
- Searches for unpublished studies will be performed using:
  - Open Grey, Current Controlled Trials, MedNar, ClinicalTrials.gov, Cos Conference Papers Index and International Clinical Trials Registry Platform of the WHO.

Example search strategy (MedLine@Ovid interface):
1. Hodgkin*;
2. Quantitative PET OR Metabolic Tumour Volume OR Total Tumour Glycolysis OR Standardized Uptake Value;
3. Qualitative PET OR Visual evaluation PET OR Visual analysis PET;
4. Diag* OR sensitivity OR specificity OR predictive;
5. 1 AND 2 AND 3 AND 4.

Data collection process

Data will be independently extracted by two reviewers (VP and MK) from studies included in the review using standardised data extraction tools from JBI-DATARI (see online supplementary appendix II). Extracted data will include characteristics of the populations, index tests, reference tests and the diagnoses relevant to the systematic review objectives. Disagreements will be resolved during team discussions, as necessary.

Data items/dealing with missing data

Both generic and trade names of the index tests will be extracted. Diagnostic accuracy of index versus reference tests will be compared using sensitivity, specificity and receiver operating characteristics (ROC) readouts, as well as patient characteristics (eg, age, gender, given disease). Study authors will be contacted, as necessary, to provide relevant information for comparative assessments.

Outcomes and prioritisation

The primary outcome of this systematic review will be to compare diagnostic and prognostic accuracy of quantitative and qualitative PET results in patients with cHL.

We will seek data answering the following specific questions:
1. What was the rate of 5-year progression-free survival (PFS; followed from enrolment through the end of the study period)?
2. What is the predicted rate of treatment failure?

Data synthesis

All available diagnostic data will be pooled into a statistical meta-analysis using JBI-DATARI. Results from the included studies will be subjected to double data entry. Meta-analysis results will be presented with two graphical techniques. First, forest plots will illustrate sensitivity and specificity of each selected primary study by graphing the means and CIs. Means and CIs will also be in numeric form. Additionally, true-positive, false-positive, true-negative and false-negative values will be listed. Second, summary ROC curves will be created. The bivariate model for performing meta-analyses will be used.

Assessment of heterogeneity

Initially, clinical heterogeneity will be assessed by determining whether study inclusion criteria are sufficiently similar to the pooled results. If heterogeneity is found, characteristics of the differing studies will be carefully investigated. If it seems that heterogeneity is due to the existence of specific risks of bias in some studies, then the meta-analysis will be restricted to studies that do not contain those risks. To ensure sensitivity analysis, we will exclude all studies that are appraised as having a high risk of bias.
Subgroup analysis
Subgroup analysis will be used for different age and gender characteristics. Another subgroup analysis will be used for chL and different comorbidities according to their type and severity. Another subgroup analysis will be used for initial disease stage and type of chemotherapy given. If the data are available in primary studies, we will perform subgroup analysis according to: PFS; standardised PET using Body Phantom experiments.

Meta-bias assessment
To show potential reporting bias, we will use funnel plots if more than 10 studies are available. Begg’s rank correlation and Egger’s regression tests will be used for detecting and correcting publication bias.

Confidence in cumulative evidence
On the basis of the results and quality of evidence, the ‘Grading of Recommendation Assessment, Development and Evaluation’ (GRADE) tool will be used.26 Quality of evidence will be assessed across the domains of: risk of bias, consistency, directness, precision and publication evidence will be assessed across the domains of: risk of bias, consistency, directness, precision and publication bias. Quality will be assessed at: high (further research is very unlikely to alter confidence in the accuracy estimate), moderate (further research will most likely impact confidence in the accuracy estimate, and may change the estimate), low (further research is very likely to impact confidence in the accuracy estimate, and will most likely change the estimate) or very low (the accuracy estimate is very uncertain).

ETHICS AND DISSEMINATION
This systematic review protocol was crafted in February 2016. Next, the systematic review development team will begin performing the protocol described herein. Dissemination of results will be targeted at patients and oncology practitioners through publication in a peer-reviewed journal. Ethical assessment is unnecessary as only existing sources of literature will be queried and evaluated.

Author affiliations
1Faculty of Medicine and Dentistry, Department of Hemato-Oncology, Palacký University in Olomouc, Olomouc, Czech Republic
2Faculty of Medicine and Dentistry, Department of Social Medicine and Public Health, Palacky University in Olomouc, Olomouc, Czech Republic
3Faculty of Medicine and Dentistry, The Czech Republic (Middle European) Centre for Evidence-Based Health Care: An Affiliated Centre of the Joanna Briggs Institute, Palacky University in Olomouc, Olomouc, Czech Republic
4Division of Hematology, Oncology and Transplantation University of Minnesota, Minneapolis, Minnesota, USA

Acknowledgements
The authors would like to thank the Charlesworth Group for professional editing of this manuscript. Grant funding for this work was provided by the Faculty of Medicine and Dentistry, Palacky University in Olomouc, Czech Republic (IGA_LF_2016_001 and RVO: 61989592) and by Takeda Pharmaceuticals International AG (ISSR-2015-101289).

Contributors
VP and MK conceptualised and designed the study. All authors contributed to selection criteria development, risk of bias assessment strategy and data extraction. MK, JK and DT were methodologists. VP, VB and TP were Hodgkin lymphoma content experts. All authors (VP, MK, VB, TP) read, provided feedback and approved the final manuscript.

Funding
Grant funding for this manuscript was provided by Faculty of Medicine and Dentistry, Palacky University in Olomouc, Czech Republic (IGA_LF_2016_001 and RVO: 61989592) and Takeda Pharmaceuticals International AG (ISSR-2015-101289). The final review will disclose all financial and non-financial sources of support (ie, pharmaceutical companies); financial support will be drawn only from the publicly funded sources aforementioned.

Competing interests
None declared.

Provenance and peer review
Not commissioned; externally peer reviewed.

Data sharing statement
The only data which the authors will miss could be missing data from primary studies. In this case, the authors will contact authors of primary research to ask for sharing the additional data with them.

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/

REFERENCES


