Cochrane Library

Cochrane Database of Systematic Reviews

Interim PET-results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies (Review)

Aldin A, Umlauff L, Estcourt LJ, Collins G, Moons KGM, Engert A, Kobe C, von Tresckow B, Haque M, Foroutan F, Kreuzberger N, Trivella M, Skoetz N

Aldin A, Umlauff L, Estcourt LJ, Collins G, Moons KG, Engert A, Kobe C, von Tresckow B, Haque M, Foroutan F, Kreuzberger N, Trivella M, Skoetz N.
Interim PET-results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies.
Cochrane Database of Systematic Reviews 2020, Issue 1. Art. No.: CD012643.
DOI: 10.1002/14651858.CD012643.pub3.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Prognosis Review]

Interim PET-results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies

Angela Aldin[1], Lisa Umlauff[1], Lise J Estcourt[2], Gary Collins[3], Karel GM Moons[4], Andreas Engert[5], Carsten Kobe[6], Bastian von Tresckow[5], Madhuri Haque[1], Farid Foroutan[7], Nina Kreuzberger[1], Marialena Trivella[3a], Nicole Skoetz[8b]

[1]Cochrane Haematology, Department I of Internal Medicine, Center for Integrated Oncology Aachen Bonn Cologne Duesseldorf, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. [2]Haematology/Transfusion Medicine, NHS Blood and Transplant, Oxford, UK. [3]Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK. [4]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands. [5]Department I of Internal Medicine, Center for Integrated Oncology Aachen Bonn Cologne Duesseldorf, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. [6]University of Cologne, Faculty of Medicine and University Hospital Cologne, Department for Nuclear Medicine, Cologne, Germany. [7]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada. [8]Cochrane Cancer, Department I of Internal Medicine, Center for Integrated Oncology Aachen Bonn Cologne Duesseldorf, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany

[a]contributed equally. [b]contributed equally

Contact address: Angela Aldin, angela.aldin@uk-koeln.de.

Editorial group: Cochrane Haematology Group.
Publication status and date: Edited (no change to conclusions), published in Issue 8, 2020.

Citation: Aldin A, Umlauff L, Estcourt LJ, Collins G, Moons KG, Engert A, Kobe C, von Tresckow B, Haque M, Foroutan F, Kreuzberger N, Trivella M, Skoetz N. Interim PET-results for prognosis in adults with Hodgkin lymphoma: a systematic review and meta-analysis of prognostic factor studies. *Cochrane Database of Systematic Reviews* 2020, Issue 1. Art. No.: CD012643. DOI: 10.1002/14651858.CD012643.pub3.

ABSTRACT

Background

Hodgkin lymphoma (HL) is one of the most common haematological malignancies in young adults and, with cure rates of 90%, has become curable for the majority of individuals. Positron emission tomography (PET) is an imaging tool used to monitor a tumour's metabolic activity, stage and progression. Interim PET during chemotherapy has been posited as a prognostic factor in individuals with HL to distinguish between those with a poor prognosis and those with a better prognosis. This distinction is important to inform decision-making on the clinical pathway of individuals with HL.

Objectives

To determine whether in previously untreated adults with HL receiving first-line therapy, interim PET scan results can distinguish between those with a poor prognosis and those with a better prognosis, and thereby predict survival outcomes in each group.

Search methods

We searched MEDLINE, Embase, CENTRAL and conference proceedings up until April 2019. We also searched one trial registry (ClinicalTrials.gov).

Selection criteria

We included retrospective and prospective studies evaluating interim PET scans in a minimum of 10 individuals with HL (all stages) undergoing first-line therapy. Interim PET was defined as conducted during therapy (after one, two, three or four treatment cycles). The

minimum follow-up period was at least 12 months. We excluded studies if the trial design allowed treatment modification based on the interim PET scan results.

Data collection and analysis

We developed a data extraction form according to the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS). Two teams of two review authors independently screened the studies, extracted data on overall survival (OS), progression-free survival (PFS) and PET-associated adverse events (AEs), assessed risk of bias (per outcome) according to the Quality in Prognosis Studies (QUIPS) tool, and assessed the certainty of the evidence (GRADE). We contacted investigators to obtain missing information and data.

Main results

Our literature search yielded 11,277 results. In total, we included 23 studies (99 references) with 7335 newly-diagnosed individuals with classic HL (all stages).

Participants in 16 studies underwent (interim) PET combined with computed tomography (PET-CT), compared to PET only in the remaining seven studies. The standard chemotherapy regimen included ABVD (16) studies, compared to BEACOPP or other regimens (seven studies). Most studies (N = 21) conducted interim PET scans after two cycles (PET2) of chemotherapy, although PET1, PET3 and PET4 were also reported in some studies. In the meta-analyses, we used PET2 data if available as we wanted to ensure homogeneity between studies. In most studies interim PET scan results were evaluated according to the Deauville 5-point scale (N = 12).

Eight studies were not included in meta-analyses due to missing information and/or data; results were reported narratively. For the remaining studies, we pooled the unadjusted hazard ratio (HR). The timing of the outcome measurement was after two or three years (the median follow-up time ranged from 22 to 65 months) in the pooled studies.

Eight studies explored the independent prognostic ability of interim PET by adjusting for other established prognostic factors (e.g. disease stage, B symptoms). We did not pool the results because the multivariable analyses adjusted for a different set of factors in each study.

Overall survival

Twelve (out of 23) studies reported OS. Six of these were assessed as low risk of bias in all of the first four domains of QUIPS (study participation, study attrition, prognostic factor measurement and outcome measurement). The other six studies were assessed as unclear, moderate or high risk of bias in at least one of these four domains. Four studies were assessed as low risk, and eight studies as high risk of bias for the domain other prognostic factors (covariates). Nine studies were assessed as low risk, and three studies as high risk of bias for the domain 'statistical analysis and reporting'.

We pooled nine studies with 1802 participants. Participants with HL who have a negative interim PET scan result probably have a large advantage in OS compared to those with a positive interim PET scan result (unadjusted HR 5.09, 95% confidence interval (CI) 2.64 to 9.81, I^2 = 44%, moderate-certainty evidence). In absolute values, this means that 900 out of 1000 participants with a negative interim PET scan result will probably survive longer than three years compared to 585 (95% CI 356 to 757) out of 1000 participants with a positive result.

Adjusted results from two studies also indicate an independent prognostic value of interim PET scan results (moderate-certainty evidence).

Progression-free survival

Twenty-one studies reported PFS. Eleven out of 21 were assessed as low risk of bias in the first four domains. The remaining were assessed as unclear, moderate or high risk of bias in at least one of the four domains. Eleven studies were assessed as low risk, and ten studies as high risk of bias for the domain other prognostic factors (covariates). Eight studies were assessed as high risk, thirteen as low risk of bias for statistical analysis and reporting.

We pooled 14 studies with 2079 participants. Participants who have a negative interim PET scan result may have an advantage in PFS compared to those with a positive interim PET scan result, but the evidence is very uncertain (unadjusted HR 4.90, 95% CI 3.47 to 6.90, I^2 = 45%, very low-certainty evidence). This means that 850 out of 1000 participants with a negative interim PET scan result may be progression-free longer than three years compared to 451 (95% CI 326 to 569) out of 1000 participants with a positive result.

Adjusted results (not pooled) from eight studies also indicate that there may be an independent prognostic value of interim PET scan results (low-certainty evidence).

PET-associated adverse events

No study measured PET-associated AEs.

Authors' conclusions

This review provides moderate-certainty evidence that interim PET scan results predict OS, and very low-certainty evidence that interim PET scan results predict progression-free survival in treated individuals with HL. This evidence is primarily based on unadjusted data. More studies are needed to test the adjusted prognostic ability of interim PET against established prognostic factors.

PLAIN LANGUAGE SUMMARY

Imaging with positron emission tomography (PET) during chemotherapy to predict outcome in adults with Hodgkin lymphoma

Review question

This Cochrane Review aimed to find out whether the results of a positron emission tomography (PET) during therapy in people with Hodgkin lymphoma (HL) can help to distinguish between those with a poor prognosis and those with a better prognosis, and predict survival outcomes in each group.

Background

Hodgkin lymphoma is a cancer which affects the lymphoid system of the body. It is considered a relatively rare disease (two to three cases per 100,000 people every year in Western countries), that is most common in young adults in their twenties, but it can also occur in children and elderly people. As treatment options have improved, most people with HL can now be cured. It is important that individuals receive the treatment with the greatest efficacy and least toxicity possible. PET is an imaging tool for assessing the disease stage of an individual, and monitoring tumour activity. It has been suggested that PET performed during therapy (so-called interim PET, e.g. after two cycles of chemotherapy) can distinguish between people who respond well to therapy and those who do not respond well. The aim of this review was to demonstrate the prognostic ability to distinguish between these groups, and predict survival outcomes in each group, to help clinicians make an informed decision on the treatment pathway to improve long-term outcomes and safety for people with HL.

Study characteristics

We included 23 studies to explore the association between interim PET scan results after one to four cycles of chemotherapy and survival outcomes in adults with HL (all stages). We contacted 10 authors, and six provided us with relevant information and/or data.

Key results

In 16 included studies, participants received either ABVD chemotherapy or BEACOPP chemotherapy (four studies) only, with or without radiotherapy. In 16 studies, participants underwent an interim PET scan in combination with a computed tomography (CT) (PET-CT), which have higher accuracy in detecting primary and secondary cancers than a PET scan alone. In the remaining seven studies, PET-only was conducted. Twenty-one studies conducted interim PET scans after two cycles (PET2) of chemotherapy.

Eight studies did not report enough data on our outcomes or population of interest, so we reported the results from these studies narratively. We combined individual study results in meta-analyses to provide robust evidence for our outcomes of interest overall survival and progression-free survival. No study measured PET-associated adverse events (harms).

For overall survival, combined results from nine studies (1802 participants) show that there is probably a large advantage in overall survival for people with a negative interim PET scan compared to people with a positive interim PET scan. For progression-free survival, combined results from 14 studies (2079 participants) show that interim PET-negative people may have an advantage for progression-free survival, compared to interim PET-positive people, but we are uncertain about this result. These are unadjusted results, where interim PET was tested as the only prognostic factor.

Eight studies reported adjusted results, where the independent prognostic ability of interim PET was assessed against other established prognostic factors (e.g. disease stage, B symptoms). We could not combine individual study results because the studies did not include identical sets of covariates. Nevertheless, their results indicate a probable independent prognostic ability of interim PET to predict both outcomes.

Certainty of the evidence

Regarding the unadjusted results, we rated our certainty of the evidence as 'moderate' for overall survival. This means that the true effect is likely to be close to the estimated effect, but there is a possibility that it is substantially different. For progression-free survival, we rated our certainty of the evidence as 'very low', meaning that we have little confidence in the effect estimate, and that the true effect is likely to be substantially different from the estimated effect.

Regarding the adjusted results, we rated our certainty of the evidence as 'moderate' for overall survival, and 'low' for progression-free survival.

How up-to-date is this review?

We searched data bases up until 2 April 2019, and one trial registry on 25 January 2019.

Summary of findings 1. Comparison of interim PET-negative and interim PET-positive individuals with Hodgkin Lymphoma

Comparison of interim PET-positive and interim PET-negative participants with Hodgkin lymphoma

Population: Individuals with Hodgkin lymphoma
Setting: Eleven studies recruited participants from a total of 28 haemato-oncology treatment centres/hospitals in Brazil (N = 1), China (N = 1), Denmark (N = 4), France (N = 4), Italy (N = 3), Poland (N = 11), UK (N = 2) and the USA (N = 2). One study (Straus 2011) included participants from 29 institutions, but did not report the countries. One study (Simon 2016) reported the country (Hungary) but not the number of centres. One multi-centre study (Hutchings 2014) recruited participants from four countries (USA, Italy, Poland and Denmark). One RCT (Kobe 2018) included participants from 301 hospitals and private practices in Germany, Switzerland, Austria, the Netherlands, and the Czech Republic.

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with Interim PET-negative	Risk with Interim PET-positive				
Overall survival Follow up: 3 years	Low		HR 5.09 (2.64 to 9.81)	1802 (9 studies)	⊕⊕⊕⊖ MODERATE [2 3 4]	
	900 per 1.000 [1]	585 per 1.000 [1] (356 to 757)				
	High					
	980 per 1.000 [1]	902 per 1.000 [1] (820 to 948)				
Progression-free survival Follow up: 3 years	Low		HR 4.90 (3.47 to 6.90)	2079 (14 studies)	⊕⊖⊖⊖ VERY LOW [6 7 8]	
	850 per 1.000 [5]	451 per 1.000 [5] (326 to 569)				
	High					
	940 per 1.000 [5]	738 per 1.000 [5] (653 to 807)				
Adverse events associated with PET - not reported	No study measured PET-associated adverse events.		-	-	-	

Overall survival (adjusted effect estimate)	Two studies reported an adjusted effect estimate for overall survival after interim PET2: a hazard ratio of 3.2 (95% CI 1.3 to 8.4, P = 0.02) (Kobe 2018) and 11.51 (95% CI 3.14 to 42.86, P < 0.001) (Simon 2015) indicates the independent prognostic value of interim PET over and above other clinically relevant prognostic factors.	-	843 (2 studies)	⊕⊕⊕⊖ MODERATE [9]
Progression-free survival (adjusted effect estimate)	Eight studies conducted a multivariable analysis to test the independent prognostic value of interim PET over and above other clinically relevant prognostic factors. Four of these studies reported a hazard ratio as the adjusted effect estimate, of which the value ranges from 2.4 to 36.89, indicating the independent prognostic value of interim PET2.[10]	-	996 (4 studies)[10]	⊕⊕⊖⊖ LOW [11 12]

*__The survival in the PET-positive group__ (and its 95% confidence interval) is based on the assumed survival in the PET-negative group.

CI: Confidence interval; **HR:** Hazard ratio; **PET:** positron emission tomography

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

[1] The assumed event-free survival in the control group is based on the survival rate of the interim PET-negative participants at 3 years in the studies included (the lowest survival rate from Cerci 2010 and the highest survival rate from Kobe 2018).
[2] High risk of bias in seven studies for the domain 'other prognostic factors (covariates)', and high risk of bias in three studies for the domain 'statistical analysis and reporting'. Downgraded by 1 point for risk of bias.
[3] For one study we used the reported hazard ratio. For seven studies we had to estimate the hazard ratio and for one study we re-calculated it (Trivella 2006). Downgraded by 1 point for imprecision.
[4] Upgraded by one point due to the large effect showing the large difference between interim PET-negative and interim PET-positive participants (HR 5.09, CI 2.64 to 9.81).
[5] The assumed event-free survival in the control group is based on the survival rate of the interim PET-negative participants at 3 years in the studies included (the lowest survival rate from Rossi 2014 and the highest survival rate from Kobe 2018).
[6] High risk of bias in eight studies for the domain 'other prognostic factors (covariates)', and high risk of bias in six studies for the domain 'statistical analysis and reporting'. Downgraded by 1 point for risk of bias.
[7]The definition of PFS varied across studies, downgraded by 1 point for inconsistency
[8] For three studies we used the reported hazard ratio. For ten studies we had to estimate the value, and for one study we had to re-calculate it (Trivella 2006). Downgraded by 1 point for imprecision.
[9] High risk of bias for the domains 'other prognostic factors (covariates)' and statistical analysis and reporting for one study (Simon 2016). Downgraded by 1 point for risk of bias.
[10] Hutchings 2006; Kobe 2018; Mesguich 2016; Simon 2016.
[11] High risk of bias for the domains 'other prognostic factors (covariates)' and statistical analysis and reporting for one study (Simon 2016). Also high risk of bias for the domain study participation in one study (Hutchings 2006). Downgraded by 1 point for risk of bias.
[12] Studies included a heterogenous set of covariates in the adjusted analyses. Downgraded by 1 point for inconsistency.

BACKGROUND

Description of the condition

Hodgkin lymphoma (HL) is a cancer of the lymph nodes and the lymphoid system with possible involvement of other organs such as the liver, lung, bone or bone marrow (Lister 1989). With an annual incidence of approximately two to three per 100,000 inhabitants in Western countries, HL is a comparatively rare disease, but it is one of the most common malignancies in young adults (Howlader 2015). In industrialised countries, the age distribution of HL shows a first peak in the third decade and a second peak after the age of 50 (Thomas 2002).

The World Health Organization (WHO) Classification of Tumours of Haematopoietic and Lymphoid Tissues distinguishes between two types of HL: classical HL, representing about 95% of all HL; and lymphocyte-predominant HL, representing about 5% of all HL (Swerdlow 2008). Both types differ in morphology, phenotype and molecular features, and therefore in clinical behaviour and presentation (Re 2005).

The Ann Arbor Classification is used for staging and distinguishes between four different tumour stages. Stages one to three indicate the degree of lymph node and localised extranodal organ involvement, or both, and stage four includes disseminated organ involvement, which can be found in 20% of cases. Factors associated with a poor prognosis include a large mediastinal mass, three or more involved lymph node areas, a high erythrocyte sedimentation rate, extranodal lesions, B symptoms (weight loss > 10%, fever, drenching night sweats) and advanced age, but the factors considered as significant vary slightly between different study groups (German Study Hodgkin Lymphoma Study Group (GHSG); European Organization for Research and Treatment of Cancer (EORTC); National Cancer Institute of Canada (NCIC)). The Cotswold modification of the Ann Arbor Classification also takes into consideration the occurrence of bulky disease (largest tumour diameter greater than 10 cm) (Lister 1989). Hodgkin lymphoma is classified into early favourable, early unfavourable and advanced stage (Engert 2007; Klimm 2005). In Europe, the early favourable-stage group usually comprises Ann Arbor stages I and II without risk factors. The early unfavourable-stage group includes individuals with Ann Arbor stages I or II and one or more risk factors. Most individuals with stages IIB, III or IV disease are included in the advanced-stage risk group (Engert 2003).

With cure rates of up to 90%, HL is one of the most curable cancers worldwide (Engert 2010; Engert 2012; Rancea 2013a; von Tresckow 2012). A combination of adriamycin, bleomycin, vinblastine and dacarbazine (ABVD) is widely accepted as the standard chemotherapy regimen in early-stage HL (Bröckelmann 2018, Canellos 1992; Engert 2010). Individuals in this stage usually receive a combination of chemotherapy and involved-field radiation therapy (IF-RT) (Engert 2010; von Tresckow 2012), whereas those with advanced-stage disease receive an intensified regimen, such as BEACOPP (bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine and prednisone) (Skoetz 2017a; Borchmann 2011; Engert 2012; Skoetz 2013), or ABVD. A large randomised study showed that two cycles of ABVD followed by 20 Gy of IF-RT is sufficient for the treatment of early-favourable HL (Engert 2010), which is implemented into current standard treatment, whereas four cycles of chemotherapy followed by 30 Gy IF-RT is more suitable for individuals with early-unfavourable HL. Approximately 10% of people with HL will be refractory to initial treatment or will relapse; this is more common in people with advanced stage or bulky disease. These individuals can be treated with high-dose chemotherapy and autologous stem cell transplantation (Rancea 2013). Immunotherapy for relapsed HL as another possible approach is under active investigation (Moskowitz 2018).

The current treatment approach for HL aims to maximise progression-free and OS and to minimise acute and long-term toxicities like cardiac and pulmonary damage, infertility and secondary cancers. Development of a secondary cancer is one of the major causes of morbidity and mortality once the risk of progression and relapse of HL is over, i.e. from about five years after first-line treatment onwards. In a large systematic review based on individual patient data in people with HL, Franklin and colleagues demonstrated that treatment de-intensification by avoiding additional radiotherapy reduces the risk of a secondary cancer (Franklin 2005).

Description of the index (prognostic) factor

A prognostic factor is a characteristic of a patient or the disease (e.g. age, sex, co-morbidities, disease stage, blood or imaging results) that is likely to predict patient outcomes or health events, often related to OS and disease-free survival (Moons 2009; Riley 2013). Prognostic information ultimately provides a basis for the determination of treatment and also helps to stratify individuals for treatment according to their risk of future outcomes (Riley 2013). Established prognostic factors in HL include age, gender, B symptoms, Ann Arbor disease stage, bulky disease, albumin level, anaemia and white blood cell count, amongst others (Cuccaro 2014; Josting 2010; Kılıçkap 2013). Particularly male gender, advanced disease stage or age, and a low level of albumin, for example, are associated with worse prognosis and survival outcomes (Cuccaro 2014; Josting 2010).

The prognostic factor to be examined in this review is the tumour's metabolic activity, its stage, and progression as captured by [18F]-fluorodeoxy-D-glucose (FDG)-positron emission tomography (PET, also called PET scanning), which is an imaging tool. The principle of FDG-PET is based on a radio-labelled glucose analogue being a good indicator of the glucose metabolism of a tissue. It comprises two parts: a vector (2-deoxy-D-glucose) taken up by cells with a high metabolic rate, and 18F, a positron-emitting nuclide, which is detected by scintigraphy. FDG-PET scanning provides the opportunity to identify the state and degree of progression of FDG-avid tumours and has therefore become a standard imaging tool for various cancers (Boellaard 2010). Hodgkin lymphoma is a FDG-avid tumour; in a study of 233 people with HL, 100% were FDG-avid (Weigler-Sagie 2010). However, as the field of imaging continuously evolves, it is now widely accepted to use PET in combination with a computed tomography (CT), known as PET-CT (Barrington 2014). The combination of PET-CT is argued to provide clearer imaging and a more accurate measurement of nodal size (Cheson 2014). Nevertheless, in the studies included in this review, the use of PET or PET-CT varied.

Over the last few decades FDG-PET has been used more and more for staging, prognosis, treatment planning and response evaluation in individuals with HL, and is a widely accepted procedure (Barrington 2017a; Cheson 2014; Fitzgerald 2019; Kobe 2010a; Markova 2009; Meignan 2009; Radford 2015; Specht 2007).

FDG-PET is primarily used for the pretreatment assessment in order to determine the stage of the disease of an individual and thereby to decide on the appropriate treatment regimen (Cheson 2014; Meignan 2009). However, it is now argued that PET should also be conducted during first-line chemotherapy in individuals with HL, namely interim PET after a few cycles of chemotherapy (Barrington 2017a; Bröckelmann 2018; Meignan 2009). The result of the interim PET scan (positive or negative) is believed to be a good predictor of outcome, aiding the distinction between individuals with a poor prognosis from those with a better prognosis, while undergoing early treatment (Gallamini 2007; Kobe 2010; Markova 2012). Therapy adaptation based on interim PET results was introduced after detailed exploration of the FDG-PET procedure (Engert 2012; Kobe 2008a), the idea being to achieve maximum efficacy in terms of OS and progression-free survival (PFS).We will refer to the prognostic factor henceforth as 'interim PET'.

Why it is important to do this review

There is a need to systematically explore the prognostic ability of the factor (interim PET) in conditions where there is no treatment adaptation. The 'no treatment adaptation' clause is a rather important point in the prognostic exploration as adapting treatment based on interim PET results in daily practice when its prognostic ability is not yet proven is not desired. There is one systematic review on the prognostic value of interim PET without treatment adaptation in individuals with HL (Adams 2015a). However, this review looked at 'treatment failure' as an outcome of the interim PET scan, which is different to the outcomes the current review explored. Moreover, and despite the fact that it is entitled as a review of prognosis studies, the methodology used is akin to diagnostic test evaluation (with calculations of diagnostic odds ratio, specificity and sensitivity), rather than using established prognostic methodology and crucially, the confidence in the calculated estimates was not rated. Moreover, the review included studies published before December 2014 and, therefore, important research published since that time is not included.

One Cochrane Review on the role of PET-adapted treatment modification for people with HL found some evidence that PFS was decreased in people with early-stage HL and a negative PET scan receiving only chemotherapy (PET-adapted therapy) compared to those receiving radiotherapy in addition to chemotherapy (which is the standard therapy regimen) (Sickinger 2015). A similar result was found in another Cochrane Review (Blank 2017). The authors compared the effects of chemotherapy alone versus chemotherapy plus radiotherapy on outcome and safety for adults with early stage HL. They found moderate evidence that when individuals receive the same number of chemotherapy cycles, the addition of radiotherapy can improve PFS. However, both reviews were not able to give definite conclusions on the effect on OS. Another systematic review suggests the change of therapy after interim PET in advanced-stage individuals only (Amitai 2018). In the current German guideline for the treatment of HL, for example, it is recommended that patients with advanced HL receive an interim PET scan after two cycles of chemotherapy. The result of the interim PET scan can then be used to guide further treatment for patients in advanced stages of HL (Bröckelmann 2018). Hence, the disease stage is an additional key prognostic factor for patients with HL. Several randomised controlled trials (RCTs) have recently been published that investigated the consequences of treatment adaptation based on interim PET scan results on outcome and safety for individuals with HL (Andre 2017; Casasnovas 2019; Kobe 2018; Johnson 2016; Radford 2015).

Hence, the prognostic role of interim PET in individuals with HL undergoing first-line chemotherapy is very important and will strongly influence decision-making particularly regarding the choice of subsequent treatments. Therefore, we have summarised all available data from identified studies and included these in a meta-analysis when they were sufficiently homogeneous. Our aim was to produce robust evidence based on the improved power that a meta-analysis provides over the limitations of individual primary studies, and grade the evidence. A reliable answer to the question of the prognostic value of interim PET scan to predict survival outcomes in individuals with HL will strongly influence decision-making at a crucial point of an individual's treatment pathway. Moreover, grading the evidence on the prognostic value of interim PET will provide readers with an estimate of how much they can rely on the calculated results.

The aim of this systematic review was to determine whether in previously untreated adults with HL receiving first-line therapy, interim PET scan results can distinguish between those with a poor prognosis and those with a better prognosis, and whether it can predict survival outcomes in each group. Thereby, we assessed the prognostic value of interim PET scan results. Meta-analyses and grading of the evidence allow a conclusion of whether interim PET is a prognostic factor. This comprehensive overview will have a great impact on international guidelines and clinical pathways, and will contribute to a high-grade support in clinical decision-making for effective, supportive strategies for the individual patient.

OBJECTIVES

To determine whether in previously untreated adults with Hodgkin lymphoma (HL) receiving first-line therapy, interim positron emission tomography (PET) scan results can distinguish between those with a poor prognosis and those with a better prognosis, and thereby predict survival outcomes in each group.

Primary objective

To identify all studies evaluating interim PET scan results as a prognostic factor, describe the characteristics and risk of bias of included studies and meta-analyse results on the association between PET scan results and overall survival (OS), progression-free survival (PFS) and PET-associated adverse events.

PICOTS

We used the PICOTS (population, index, comparator, outcome(s), timing, setting) system to describe the key items for framing this review and its objective and methodology (Table 1) (Debray 2017; Riley 2019).

Table 1. PICOTS system

Population	Index (prognostic) factor	Comparator	Outcome(s)	Timing	Setting

• People with classic HL, at any stage of the disease • Newly diagnosed individuals undergoing first-line therapy • Adults, as defined in the studies	Interim PET scan results	Not applicable to this review	• Overall survival (OS) • Progression-free survival (PFS) • PET-associated adverse events (AEs) The outcome should be measured after a minimum follow-up of 12 months.	• Interim PET scan should be conducted during chemotherapy (after one, two, three or four cycles of chemotherapy)	Hospital/treatment centre

METHODS

This is a systematic review of prognostic factor studies.

Criteria for considering studies for this review

Types of studies

We included retrospective and prospective studies evaluating interim PET scan results in a minimum of 10 individuals with Hodgkin lymphoma (HL) undergoing first-line therapy.

We excluded studies that modified the treatment regimen based on the interim PET scan results in order to draw an unbiased conclusion of the ability of interim PET to predict the outcomes under study.

Participants

We included studies on adults with newly diagnosed classic HL receiving first-line therapy. If in a study a percentage of the included participants were adolescents but received adult treatment regimen and dosage, and the study considered them as adults, then we also accepted this 'adult' definition.

All participants received an interim PET scan during chemotherapy (e.g. after one, two, three and/or four cycles of chemotherapy), and continued with the planned chemotherapy regimen, without treatment adaptation due to the interim PET scan result.

Index (prognostic) factor

We included studies that assessed interim PET scan results as the index (prognostic) factor to predict survival outcomes. We expected the interim PET scan to be conducted during first-line treatment of adults with HL, and without interim PET-guided treatment adaptation, meaning participants should be treated in the same way regardless of the interim PET scan result. We accepted all studies that conducted a PET or PET-CT (see Background 'Description of index (prognostic) factor').

In the literature, it is generally recommended to use a five-point scale to assess the grade of uptake and report the PET scan result (Meignan 2009). Generally, scores 1-3 indicate PET-negativity, while scores 4-5 indicate PET-positivity (Barrington 2014). Most of the included studies used a validated scale, such as the 5-PS Deauville criteria (Meignan 2009), the Lugano classification (Cheson 2014), the Imaging Subcommittee of International Harmonization Project in Lymphoma criteria (Juweid 2007) or the joint Italian-Danish study criteria (Gallamini 2007).

Type of outcome measures

Primary outcome

- Overall survival (OS), defined as the time to death due to any cause.

We chose OS as our primary outcome because it has the greatest clinical relevance and is most important for individuals with HL. Furthermore, death due to any cause is an objective endpoint not susceptible to bias by the outcome assessor.

Secondary outcomes

- Progression-free survival (PFS), defined as the time to disease progression, relapse, death due to any cause or last follow-up.
- Adverse events (AEs), defined as any event associated with the index factor (e.g. radiation safety).

To report meaningful findings, the required minimum follow-up period was 12 months for each outcome.

Search methods for identification of studies

Electronic searches

Reporting and therefore retrieval of prognostic factor studies is very poor, as evaluation of guidelines on reporting of prognostic markers in cancer have shown (Altman 2012; Mallett 2010; McShane 2005). Moreover, no specific search filter exists for this new methodological approach, therefore published filters have to be combined for a sensitive search strategy (Geersing 2012). However, as PET scans often are not reported as a prognostic factor, we did not combine our search strategy with a filter for prognosis research. Therefore, the search strategy was not very specific and the results were screened independently and in detail by two teams of two review authors. Furthermore, we did not apply a language restriction in order to reduce the language bias, according to chapter six of the *Cochrane Handbook for Systematic Reviews of Interventions* (Lefebvre 2011).

We searched the following databases.

- Databases of medical literature
  - Cochrane Central Register of Controlled Trials (CENTRAL; 2 April 2019, Issue 11) (Appendix 1)
  - MEDLINE Ovid SP (1946 until 2 April 2019) (Appendix 2)
  - Embase (1990 until 2 April 2019) (Appendix 2)
- Conference proceedings of annual meetings of the following societies for abstracts (2000 to 2019)
  - American Society of Hematology

- European Hematology Association
- International Symposium on Hodgkin Lymphoma

- We searched ClinicalTrials.gov (on 25 January 2019 using the query PET and Hodgkin lymphoma) to identify clinical trials.

Searching other resources

- Handsearching of references
  * We searched the references of all identified studies, relevant review articles and current treatment guidelines for further literature to find other relevant studies and to identify associated articles.
- Personal contacts
  * We contacted 10 principal investigators of included studies for further information, of whom six replied and answered our questions for clarification. Two out of these six provided us also with relevant data to conduct our analyses.

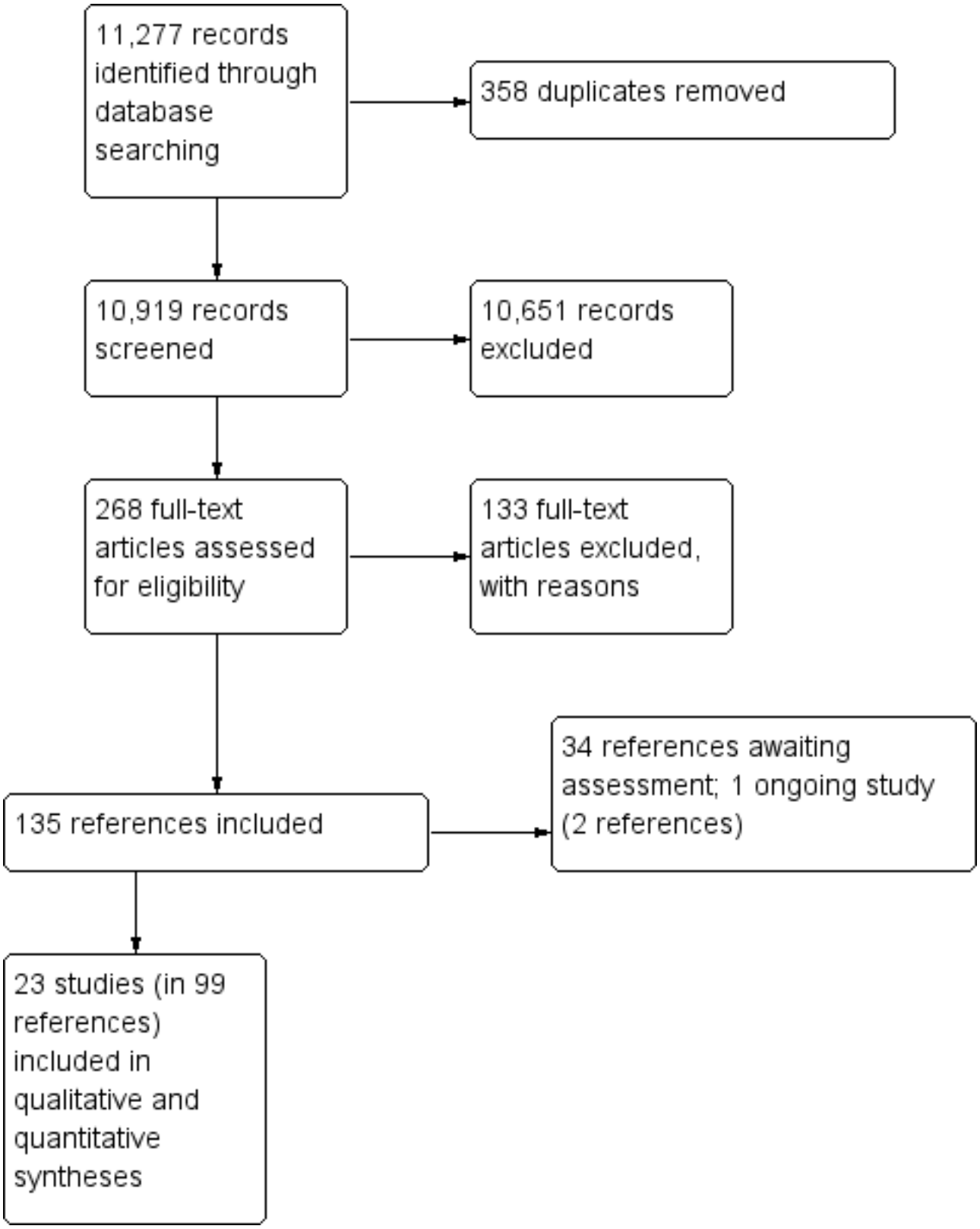
Data collection and analysis

Selection of studies

Two teams of two review authors (AA, LE, MHT, NS) independently screened the results of the search strategies to identify eligible studies by reading the titles and abstracts in Covidence (Covidence). In case of disagreements, consensus between the two review authors was reached by discussion of the full-text publication. When consensus could not be reached, a third review author was consulted for final decision (Higgins 2011).

We documented the study selection process in a flow chart as recommended in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher 2009), showing the total numbers of retrieved references and the numbers of included and excluded studies (Figure 1).

Figure 1. Study flow diagram according to PRISMA



Data extraction and management

We developed a data extraction form specific to studies of prognostic factors based on the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) (Moons 2014). The form was piloted using four of the included studies, and then further assessed during several teleconferences between the review authors to discuss required changes. After several amendments of the form, two teams of two review authors (AA, LE, MHT, NS) independently extracted all relevant data from the included studies. After data extraction, we contacted 10 principal investigators of included studies to request additional information.

Our form included the following items (in short).

- General information
  - i.e. Author, title, source, publication date, country, language, duplicate publications
- Source of data
  - i.e. Cohort, prospective planned study, randomised study participants, or registry data
- Participants
  - Participant eligibility and recruitment method (e.g. consecutive participants, location, number of centres, setting, inclusion and exclusion criteria)
  - Participant description (e.g. age, gender, stage of disease)
  - Details of treatments received
  - Study dates
- Prognostic factor
  - Definition and method for measurement of prognostic factor
  - Timing of prognostic factor measurement (number of chemotherapy cycles before and after measurement of the prognostic factor)
- Outcomes to be predicted
  - Definition and method for measurement of outcome
  - Was the same outcome definition (and method for measurement) used in all individuals?
  - Was the outcome assessed without knowledge of the prognostic factor (i.e. blinded)?
  - Time of outcome occurrence or summary of duration of follow-up
- Sample size
  - Number of participants and number of outcomes/events
- Missing data
  - Number of participants with any missing value (include predictors and outcomes)
  - Handling of missing data (e.g. complete-case analysis, imputation, or other methods)
- Reported results
  - Overall survival (OS) (including duration of follow-up)
  - Progression-free survival (PFS) (including duration of follow-up)
  - Adverse events (AEs) (including duration of follow-up)

Risk of bias

In the protocol for this review we prespecified that we will use the Quality in Prognostic Studies (QUIPS) tool (Hayden 2013) for the risk of bias assessment. However, recent methodological developments for the systematic review of prognostic factor studies (Riley 2019; Riley 2019b) led us to consider amending this tool. In the light of this we consulted the primary author (Hayden 2013) of the QUIPS tool and following discussions decided to add to the three bias ratings ('low', 'moderate' and 'high' risk of bias) a fourth 'unclear' option. This was necessary due to the inconsistent reporting of the included studies, when information was clearly missing, and hence, without an 'unclear' category, risk of bias assessment would not be feasible.

Following further discussions, we additionally decided to rename the fifth domain 'study confounding' to 'other prognostic factors (covariates)' in order to highlight the important distinction between confounding (the preferred term when seeking estimates of causal effect of a specific etiologic factor) and adjusting for other important prognostic factors, namely covariates (advocated when seeking the independent prognostic ability of index prognostic factors). As said, in the context of our review (adults with Hodgkin lymphoma), the disease stage is a key factor that is taken into account together with the interim PET scan result when decisions about treatment adaptation are made in daily clinical practice (Bröckelmann 2018). Hence, we assessed studies that only included participants within one disease stage (e.g. only early stages or only advanced stages of HL) as 'low' risk of bias, as such patient sampling can be considered as accounting for disease stage as another prognostic factor. Studies that included participants within all disease stages, but offered adjusted results including disease stage as another prognostic factor, were also assessed as 'low' risk of bias. Studies with participants of all disease stages, not accounting for disease stage, were assessed as 'high' risk of bias in this domain. This latter modification is also reflected in the GRADE assessment. Regardless of whether meta-analysis of adjusted or unadjusted (crude) effects of the prognostic factor of interest (interim PET scan results) was possible, we included this domain's risk of bias assessment in our GRADE judgement.

Two teams of two review authors (AA, LE, MHT, NS) independently assessed the risk of bias of the included studies according to the domains of the QUIPS tool. We judged each domain by taking into account the criteria listed for each domain in the QUIPS tool (Hayden 2013), and also provided a brief statement supporting our judgement.

We made the following judgements.

- **Low risk of bias**: the relationship between the prognostic factor and outcome is unlikely to be different for participants and eligible non-participants.
- **Moderate risk of bias**: the relationship between the prognostic factor and outcome may be different for participants and eligible non-participants.
- **High risk of bias**: the relationship between the prognostic factor and outcome is very likely to be different for participants and eligible non-participants.
- **Unclear risk of bias**: the study does not provide sufficient information that allows a clear judgement for this domain.

Furthermore, we decided to assess the risk of bias per outcome in each study because not all studies reported all of our outcomes of interest, and even studies reporting at least two of our outcomes showed differences in their outcome reporting.

We judged the following domains and criteria.

- Study participation
  - Adequate participation in the study by eligible persons
  - Description of the source population or population of interest
  - Description of the baseline study sample
  - Adequate description of the sampling frame and recruitment
  - Adequate description of the period and place of recruitment
  - Adequate description of inclusion and exclusion criteria
- Study attrition
  - Adequate response rate for study participants
  - Description of attempts to collect information on participants who dropped out
  - Reasons for loss to follow-up are provided
  - Adequate description of participants lost to follow-up
  - There are no important differences between participants who completed the study and those who did not
- Prognostic factor measurement
  - A clear definition or description of the prognostic factor is provided
  - Method of prognostic factor measurement is adequately valid and reliable
  - Continuous variables are reported or appropriate cut points are used
  - The method and setting of measurement of prognostic factor is the same for all study participants
  - Adequate proportion of the study sample has complete data for the prognostic factor
  - Appropriate methods of imputation are used for missing prognostic factor data
- Outcome measurement
  - A clear definition of the outcome is provided
  - Method of outcome measurement used is adequately valid and reliable
  - The method and setting of outcome measurement is the same for all study participants
- Other prognostic factors (covariates)
  - Other prognostic factors (covariates) are measured
  - Clear definitions of the important prognostic factors (covariates) measured are provided
  - Measurement of all important prognostic factors (covariates) is adequately valid and reliable
  - The method and setting of prognostic factor measurement are the same for all study participants
  - Appropriate methods are used if imputation is used for missing data
  - Important potential prognostic factors (covariates) are accounted for in the study design
  - Important potential prognostic factors (covariates) are accounted for in the analysis
- Statistical analysis and reporting
  - Sufficient presentation of data to assess the adequacy of the analytic strategy
  - Strategy for model building is appropriate and is based on a conceptual framework or model
  - The selected statistical model is adequate for the design of the study
  - There is no selective reporting of results

Reporting deficiencies

Methods and reporting in prognostic research often do not follow current methodological recommendations, limiting retrieval, reliability and applicability of these publications (Bouwmeester 2012; Peat 2014). There is evidence suggesting that prognosis research in cancer is cluttered with false-positive studies, which would not have been published if the results were negative (Kyzas 2005; Kyzas 2007; Sauerbrei 2005). Moreover, studies evaluating prognostic factors are usually not prospectively registered and no protocol is published (Peat 2014; Riley 2013), resulting in difficulties to identify all studies and to assess potential risks of publication bias. We used sensitive search filters for the disease (HL) and the prognostic factor (interim PET scan results) without any specific filter for research on prognosis in order to increase retrieval.

Due to the expected large effect of hazard ratios (HRs), tests for funnel plot asymmetry could result in publication bias being incorrectly indicated by the test (Macaskill 2010). Therefore, we decided not to evaluate the risk of publication bias by funnel plot asymmetry and describe reporting deficiencies instead.

Data synthesis

We performed analyses according to the recommendations of Cochrane, and the Cochrane Prognosis Methods Group in particular, and used the Cochrane statistical package Review Manager 5 (Deeks 2011; Review Manager 2014). We are aware that since the protocol development, the methodology on assessing studies of prognosis has evolved; hence, some differences between the published protocol and this full review may exist to account for the updated guidance. We have listed these in Differences between protocol and review.

We pooled unadjusted (crude) HRs for OS and PFS by applying meta-analysis using the RevMan's generic inverse variance methods random-effects model. Due to reporting inefficiencies and the expected heterogeneity between studies, we only combined studies that were sufficiently similar (e.g. most studies used ABVD as the main therapy regimen, or most studies conducted interim PET after two cycles of chemotherapy). Studies did not always provide an HR and associated standard error (SE), which are the parameters needed for meta-analysis. Where these values were not available, we estimated them from other available data where possible using an in-house calculator based on published methods for recovering survival data (Altman 1999; Parmar 1998; Tierney 2007). Recovered data included information and results reported in the text, tables, and Kaplan-Meier (K-M) curves. We also contacted 10 principal investigators of included studies to either ask for additional data, or to clarify issues regarding the studies.

As prespecified in the protocol, we would have also pooled adjusted HRs of the interim PET scan-result (the index prognostic factor) from multivariable analyses of the included studies as

adjusted prognostic effects (e.g. HRs) indicate the independent prognostic value of the prognostic factor over and above other clinically relevant prognostic factors (Riley 2019). However, pooling of adjusted estimates is recommended only if the same (largely) prognostic factors (covariates) are adjusted for in multivariable analyses (Riley 2019; Riley 2019b). As, said clinically relevant prognostic factors in individuals with HL particularly include the disease stage, as well as age, gender, and B symptoms (Cuccaro 2014). Regardless of whether pooling of adjusted or unadjusted effects of interim PET scan results was possible, we always assessed the risk of bias for all studies using the QUIPS tool, including the fifth domain 'other prognostic factors (covariates)', where we considered the disease stage as an important covariate to be taken into account.

Detailed description of the estimation of hazard ratios (HRs) and standard errors (SEs)

We used unadjusted HRs as the effect measure for OS and PFS. In cases where the HR and SE were not reported, we estimated them from available data using an in-house calculator (Trivella 2006), based on methods reported by Tierney 2007, Altman 1999 and Parmar 1998, or contacted authors to request additional data (Higgins 2011b). Recovered data included sample size, number of events, results such as the logrank P-value and confidence intervals (CIs), which were reported in the text, tables, and K-M curves. We kept detailed records of how the HR and SEs were calculated for each outcome in each included study. We identified the following six categories of HR precision.

1. HR was provided in the study, and the SE was either provided or easily estimated from reported CIs, and/or using the RevMan inbuilt calculator.
2. HR was provided but on checking while attempting to obtain the SE, there were errors and/or discrepancies with related provided data and we re-estimated the HR.
3. HR and SE were not provided but all necessary data for their estimation were available in the study.
4. HR and SE were not provided. Other necessary data were available but not an exact logrank P value, hence the nearest value was used in the estimation. For example, if they reported P < 0.001, then the nearest exact value was used, in this case P = 0.0009.
5. HR and SE were not provided. Other necessary data were available but the number of events was estimated from the K-M curves.
6. IPD data were available and HR and SE were accurately calculated.

We are aware that categories four and five are likely to over- or under-estimate the HR and associated SE. However, they were the best estimates we could obtain. We consider the remaining categories as precise. We explored the precision of the estimates in a post-hoc sensitivity analysis where the imprecise studies were temporarily removed to examine the robustness of the pooled result.

Grading the evidence

According to the recommendations of the GRADE working group, we rated and described the confidence in estimates for each outcome by assessing potential risk of bias, inconsistency, imprecision, indirectness and publication bias. We applied an approach that has been proposed for prognosis studies by the GRADE working group, suggesting that the starting point is one of high certainty of the evidence for observational studies (Iorio 2015).

Dealing with missing data

We dealt with missing data as suggested in Chapter 16 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011b). We contacted ten principal investigators of included studies to answer our questions regarding the studies and/or to provide us with additional data. Six principal investigators replied and answered our questions, of which two also provided us with additional data necessary to perform our analyses. One investigator kindly provided us with individual participant data for the whole data set. In some studies, the description of the methodology was rather unclear or relevant information was missing. In addition, some studies did not fully report their statistical analyses and data were missing, which complicated a full assessment of the study. We performed sensitivity analysis to assess how sensitive the results were to reasonable changes in the assumptions that were made, and addressed the potential impact of missing data on the findings of this review in the Discussion.

Furthermore, we noticed that most studies applied exclusion criteria on the baseline population (such as unavailability of interim PET or descriptive information) without providing a description of the size of this population and/or reasons for missing information. We treated this as a potential source of selection bias in the domain study participation of the QUIPS tool.

Investigation of heterogeneity

We investigated and discussed clinical and statistical heterogeneity and design aspects of included studies as mentioned in the section 'Data extraction and data management'. We assessed between-study heterogeneity using the I^2 statistic (an I^2 greater than 50% = moderate heterogeneity; an I^2 greater than 80% = considerable heterogeneity) (Deeks 2011). As most studies of prognosis are observational in nature, we are aware that they are prone to higher and/or inflated heterogeneity. Hence, we also assessed the Tau^2 values from the meta-analyses to be able to make a more robust judgment on the degree of statistical heterogeneity.

As specified in the protocol, we explored potential causes of heterogeneity by subgroup analysis. We considered the following parameters.

- Study design (e.g. prospective versus retrospective)
- Disease stage (e.g. early versus advanced stages)
- Type of chemotherapy (e.g. ABVD versus BEACOPP)
- Type of radiotherapy (e.g. involved field versus involved site)
- Type of PET measurement (e.g. PET versus PET-CT) *(post-hoc)*

In addition, we conducted a post hoc sensitivity analysis for the timing of the interim PET, as well as the availability/estimation of HR and SE to explore the robustness of the pooled results.

RESULTS

Results of the search

Our literature search in CENTRAL, MEDLINE and Embase (until 2 April 2019, see Appendix 1, Appendix 2 and Appendix 3, respectively) and one trial registry (ClinicalTrials.gov on 25 January

2019), identified 11,277 potentially relevant publications. After removal of 358 duplicates, we screened titles and abstracts of 10,919 references using inclusion and exclusion criteria defined at the protocol stage. These criteria led to the exclusion of 10,651 references, and 268 references were then included for full-text screening. Before starting full-text screening, we discussed and determined exclusion reasons. Full-text screening led to the exclusion of 133 references. Thirty-four references that were identified are still awaiting assessment (see Studies awaiting classification), and one study is still ongoing (see Ongoing studies). Hence, we finally included 23 studies (from 99 references) in this review. The overall number of publications screened, identified, selected and included in this review is shown in Figure 1

Description of studies

Included studies

See also Characteristics of included studies.

We included 23 studies in this review (Andre 2017; Annunziata 2016; Barnes 2011; Casasnovas 2019; Cerci 2010; Gallamini 2014; Gandikota 2015; Hutchings 2005; Hutchings 2006; Hutchings 2014; Kobe 2018; Markova 2012; Mesguich 2016; Oki 2014; Okosun 2012; Orlacchio 2012; Rossi 2014; Simon 2016; Straus 2011; Touati 2014; Ying 2014; Zaucha 2017; Zinzani 2012), which added up to a total of 99 references when secondary citations were included. To avoid duplication and overlapping of participant data in our analyses, we grouped those publications that assessed the same population (or groups from the same population). In such cases, we chose the publication with the greatest number of participants and/or most information as the primary publication. Duplicate or overlapping study populations were found for eight studies (Andre 2017; Barnes 2011; Gallamini 2014; Kobe 2018; Markova 2012; Simon 2016; Straus 2011; Zinzani 2012). Four studies did not report the duration of follow-up (Andre 2017; Annunziata 2016; Orlacchio 2012; Straus 2011). The earliest study recruited participants between 1993 and 2004 (Hutchings 2005), and the most recent between 2007 and 2014 (Annunziata 2016).

There was considerable heterogeneity between the included studies, particularly with regard to: stages of disease; treatment regimens; and the timing and criteria for evaluation of the interim PET scans, which are described in detail in the sections below. For meta-analyses, we only grouped studies that were homogenous enough in order to ensure comparability, and conducted subgroup analyses to explore the potential impact of heterogeneity on our results (see Methods 'Investigation of heterogeneity').

Study design

Of the 23 included studies, seven studies were retrospective single-centre studies (Annunziata 2016; Markova 2012; Oki 2014; Orlacchio 2012; Rossi 2014; Touati 2014; Ying 2014). Five studies were retrospective multi-centre studies (ranging between two to 17 centres) (Barnes 2011; Gallamini 2014; Mesguich 2016; Okosun 2012; Zinzani 2012). Two retrospective studies did not report the number of centres from which participants were recruited (Gandikota 2015; Simon 2016). Out of eight studies with a prospective study design, one study was a single-centre study (Cerci 2010), three were multi-centre studies (including between four and 11 centres, with Hutchings 2014 not reporting the number of study centres) (Hutchings 2006; Hutchings 2014; Zaucha 2017), and four were clinical trials (Andre 2017; Casasnovas 2019; Kobe 2018; Straus 2011). One study did not report the study design (Hutchings 2005).

For more details see Characteristics of included studies.

Sample size

The smallest study included 23 participants (Okosun 2012) and the largest study included 1945 participants (Kobe 2018).

Location

The included studies were conducted in a variety of countries, including Austria, Belgium, Brazil, Croatia, Czech Republic, Denmark, France, Germany, Hungary, Italy, the Netherlands, Poland, Slovakia, Switzerland, the United Kingdom (UK), the United States of America (USA), and the People's Republic of China. Four studies reported the country but not the study centre (Annunziata 2016; Hutchings 2014; Markova 2012; Simon 2016), and two studies reported neither country nor study centre (Gandikota 2015; Straus 2011).

Participants

This review included a total of 7335 male and female consecutive participants who were newly diagnosed with classic HL and received first-line therapy. Out of these, a total of 2205 participants were included in meta-analyses.

Follow-up

There were differences in the follow-up time between studies. Three studies did not report follow-up time (Annunziata 2016; Orlacchio 2012; Straus 2011). Two studies reported follow-up time per subgroup, i.e. surviving participants only (Kobe 2018; Zaucha 2017). The median follow-up time for the remaining 18 studies ranged from 23 to 66 months. The total raw range of follow-up time was between two to 195 months.

Stages of disease

Fifteen studies included all stages of the disease. Four studies included only early stages (Andre 2017; Barnes 2011; Gandikota 2015, Straus 2011) and four studies only advanced stages (Casasnovas 2019; Kobe 2018; Markova 2012; Okosun 2012).

Treatment/therapy

The following chemotherapy regimens were administered.

- ABVD (adriamycin/doxorubicin, bleomycin, vinblastine and dacarbazine) in 16 studies (Andre 2017; Annunziata 2016; Barnes 2011; Cerci 2010; Gallamini 2014; Hutchings 2005; Hutchings 2006; Hutchings 2014; Mesguich 2016; Oki 2014; Okosun 2012; Orlacchio 2012; Simon 2016; Touati 2014; Zaucha 2017; Zinzani 2012).
- Either ABVD or BEACOPP in one study (Ying 2014).
- $BEACOPP_{escalated}$ (bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine and prednisone in escalated doses) in one trial (Casasnovas 2019).
- $BEACOPP_{escalated}$ or $BEACOPP_{escalated}$ with rituximab in one trial (Kobe 2018).
- $BEACOPP_{escalated}$ or time-condensed $BEACOPP14_{baseline}$ (BEACOPP in standard, non-escalated doses repeated on day 15) in one study (Markova 2012).

- AVG (doxorubicin, vinblastine and gemcitabine) in one trial (Straus 2011).
- ABV/MOPP (adriamycin, bleomycin, vinblastine, mechlorethamine, vincristine, procarbazine and prednisone), ABVD/COPP (ABVD plus cyclophosphamide, vincristine, procarbazine and prednisone), eBEACOPP, or PVAG (prednisone, vinblastine, doxorubicin and gemcitabine) in subgroups of participants in three studies (Hutchings 2005; Hutchings 2006; Touati 2014).
- Anthracycline-based chemotherapy not further specified in one study (Rossi 2014).

The following number of chemotherapy cycles were administered.

- Two, three, four, six or eight cycles of chemotherapy alone or combined with radiotherapy in 15 studies (Andre 2017; Annunziata 2016; Barnes 2011; Casasnovas 2019; Cerci 2010; Gallamini 2014; Hutchings 2014; Markova 2012; Mesguich 2016; Orlacchio 2012; Rossi 2014; Simon 2016; Straus 2011; Zaucha 2017; Zinzani 2012). The number of cycles usually depended on the stage of the disease.
- Four, six or eight cycles of chemotherapy, depending on the interim PET scan results, in one trial (Kobe 2018). A protocol amendment during the trial introduced a reduction of standard therapy from eight to six cycles.
- Six cycles of chemotherapy combined with antiretroviral therapy due to HIV-positive study population in one study (Okosun 2012).

Six studies did not report the number of cycles (Gandikota 2015; Hutchings 2005; Hutchings 2006; Oki 2014; Touati 2014; Ying 2014).

The following radiotherapy techniques were used either in all or a subgroup of participants.

- Involved-field radiotherapy in eight studies (Barnes 2011; Gallamini 2014; Hutchings 2005; Hutchings 2006; Hutchings 2014; Mesguich 2016; Rossi 2014; Simon 2016), and either involved-field radiotherapy or extended-field radiotherapy in one study (Gandikota 2015).
- Involved-node radiotherapy in three studies (Andre 2017; Annunziata 2016; Zaucha 2017).
- Involved-site radiotherapy in two studies (Touati 2014; Zinzani 2012).
- Radiotherapy without further specification in five studies (Cerci 2010; Kobe 2018; Markova 2012; Orlacchio 2012; Ying 2014).
- No radiotherapy in three studies (Oki 2014; Okosun 2012; Straus 2011).

Stem cell transplantation was conducted in participants who relapsed after first-line therapy despite treatment escalation or salvage therapy.

- Autologous stem cell transplantation in eight studies (Cerci 2010; Gallamini 2014; Hutchings 2014; Mesguich 2016; Touati 2014; Ying 2014; Zaucha 2017).
- Autologous and/or allogeneic stem cell transplantation in one study (Zinzani 2012).
- Type of stem cell transplantation not specified in four studies (Hutchings 2005; Hutchings 2006; Markova 2012; Orlacchio 2012).
- No stem cell transplantation reported in 10 studies (Andre 2017; Annunziata 2016; Barnes 2011; Gandikota 2015; Kobe 2018; Oki 2014; Okosun 2012; Rossi 2014; Simon 2016; Straus 2011).

Index (prognostic) factor

Participants in 16 out of 23 studies underwent PET combined with computed tomography (CT), contrast enhanced CT, or multi detector CT (MDCT), compared to PET-only for participants in the other studies. Participants in 13 studies underwent PET-CT (Annunziata 2016; Cerci 2010; Gallamini 2014; Gandikota 2015; Hutchings 2014; Kobe 2018; Mesguich 2016; Okosun 2012; Rossi 2014; Simon 2016; Touati 2014; Ying 2014; Zaucha 2017). Participants in another study underwent either PET or PET-CT (Barnes 2011); participants in one study underwent PET with contrast-enhanced CT (Markova 2012); and participants in another study underwent PET/MDCT (Orlacchio 2012). In the remaining seven studies, participants underwent a PET scan only (Andre 2017; Casasnovas 2019; Hutchings 2005; Hutchings 2006; Oki 2014; Straus 2011; Zinzani 2012).

Timing of interim PET

The timing of interim PET imaging varied between studies. In most studies, participants underwent an interim PET scan after two cycles (PET2) of chemotherapy (Andre 2017; Casasnovas 2019; Cerci 2010; Gallamini 2014; Hutchings 2005; Hutchings 2006; Kobe 2018; Mesguich 2016; Oki 2014; Okosun 2012; Orlacchio 2012; Rossi 2014; Simon 2016; Straus 2011; Touati 2014; Zinzani 2012). In another study, participants underwent an interim PET scan after the first cycle (PET1) of chemotherapy only (Annunziata 2016). In one study, participants underwent interim PET scans after the first and second cycle of chemotherapy, but the study protocol was amended after interim analysis to limit PET2 scans to participants with positive results after PET1 (Zaucha 2017). In one multi-centre study, participants from two centres underwent both PET1 and PET2, whereas participants from the remaining two centres underwent PET2 only if PET1 was positive (Hutchings 2014). Three retrospective studies included participants who underwent interim PET after two to four cycles of chemotherapy (Barnes 2011; Gandikota 2015; Ying 2014), and in another study participants underwent interim PET after four cycles (PET4) of chemotherapy (Markova 2012). For meta-analyses, we used information at PET2 whenever available in order to ensure homogeneity across studies.

Evaluation of PET scans

In most studies, two nuclear medicine physicians evaluated the PET scans individually, and disagreements in scoring were solved in a consensus meeting (Annunziata 2016; Barnes 2011; Cerci 2010; Hutchings 2005; Hutchings 2006; Hutchings 2014; Mesguich 2016; Orlacchio 2012; Rossi 2014; Ying 2014; Zinzani 2012). Evaluation of PET scans was performed by only one expert in one study (Markova 2012); and by a panel consisting of three to six experts in eight studies (Andre 2017; Casasnovas 2019; Gallamini 2014; Kobe 2018; Oki 2014; Okosun 2012; Straus 2011; Zaucha 2017). Three studies did not report the number or qualification of persons who performed evaluation of PET scans (Gandikota 2015; Simon 2016; Touati 2014). Nine out of 13 multi-centre studies reported that evaluation of PET scans took place centrally (Andre 2017; Gallamini 2014; Hutchings 2006; Kobe 2018; Mesguich 2016; Okosun 2012; Straus 2011; Zaucha 2017; Zinzani 2012), and two studies did not report how reviewing of PET scans was performed across centres (Barnes 2011; Hutchings 2014).

In 11 studies, outcome assessors were blinded to the outcome (Kobe 2018; Gallamini 2014; Gandikota 2015; Hutchings 2006; Hutchings 2014; Mesguich 2016; Oki 2014; Rossi 2014; Straus 2011; Zaucha 2017; Zinzani 2012). The remaining studies did not report blinding.

Criteria for evaluation

Most studies reported the use of a standardised scale for the evaluation of the PET scans, but the scoring systems and cut-off points between studies varied.

- In 12 studies, the Deauville 5-point scoring system for evaluation of PET scans was used: in nine studies, Deauville scores 1 - 3 were considered as PET-negative, and Deauville scores 4 - 5 as PET-positive (cut-off ≥4) (Annunziata 2016; Casasnovas 2019; Gallamini 2014; Hutchings 2014; Oki 2014; Okosun 2012; Rossi 2014; Simon 2016; Zaucha 2017); in two studies, both cut-off points for evaluation of the PET scans were used by scoring each image twice, and comparing performance of interim PET between both scales (Kobe 2018; Mesguich 2016); and in one study, it was reported that the PET scans were re-interpreted retrospectively using the Deauville criteria, but it was not indicated which cut-off points were used (Touati 2014).
- In one study, the International Harmonization Project criteria were used: a PET scan was considered positive when the residual mass is ≥ 2 cm or, if less than 2 cm, positive if its activity is above that of the surrounding background (Andre 2017). A negative PET scan corresponds to Deauville score 1 (no uptake) and score 2 (uptake ≤ mediastinum).
- In two studies, the scoring systems were not specified, but similar scales and cut-off points as the Deauville scoring system were used: in one study, PET scans were reviewed using a 4-point scale (Barnes 2011), and in another study using a 5-point scale (Gandikota 2015).
- In three studies, other standardised scales for the evaluation of PET scans were used: one study used the Juweid criteria (Zinzani 2012), and two studies used the International Harmonization Project guidelines (Orlacchio 2012; Straus 2011).
- Two studies did not report how PET scans were evaluated (Hutchings 2005; Hutchings 2006); and four studies reported performance of visual evaluation but did not indicate the use of a standardised scoring system (Cerci 2010; Markova 2012; Touati 2014; Ying 2014).

Outcomes

Primary outcome

Overall survival (OS)

Univariable analyses

Twelve out of 23 included studies reported unadjusted results for our primary outcome OS (Barnes 2011; Casasnovas 2019; Cerci 2010; Gallamini 2014; Hutchings 2005; Hutchings 2006; Hutchings 2014; Kobe 2018; Simon 2016; Touati 2014; Zaucha 2017; Zinzani 2012). Of these, nine provided sufficient information and data to be included in meta-analysis. One study reported an HR that we used (Kobe 2018). Another study reported an HR, but we still re-calculated it due to discrepancies in values between the graph and table (Simon 2016). For the other seven studies, we estimated the HR using other available data from the publications (Barnes 2011; Cerci 2010; Hutchings 2005; Hutchings 2014; Touati 2014; Zaucha 2017; Zinzani 2012).

Multivariable analyses

Two studies reported adjusted results for OS (Kobe 2018; Simon 2016). Two additional studies planned, but did not conduct the analysis for different reasons (Gallamini 2014; Hutchings 2005).

Secondary outcomes

Progression-free survival (PFS)

Univariable analyses

Twenty-one out of 23 studies reported unadjusted results for PFS (Andre 2017; Annunziata 2016; Barnes 2011; Casasnovas 2019; Cerci 2010; Gallamini 2014; Hutchings 2005; Hutchings 2006; Hutchings 2014; Kobe 2018; Markova 2012; Mesguich 2016; Oki 2014; Okosun 2012; Rossi 2014; Simon 2016; Straus 2011; Touati 2014; Ying 2014; Zaucha 2017; Zinzani 2012). Of these, 15 provided sufficient information and data to be included in meta-analysis. Three studies provided an HR which we used (Annunziata 2016; Kobe 2018; Simon 2016). Another three studies reported an HR, but we still re-calculated it due to unclear description of the statistical methods used (Hutchings 2006), reporting discrepancies between graphs and tables (Mesguich 2016) or general uncertainties in the reported values (Rossi 2014). For eight studies we estimated the HR using other available data (Barnes 2011; Cerci 2010; Hutchings 2005; Straus 2011; Touati 2014; Ying 2014; Zaucha 2017; Zinzani 2012).

Multivariable analyses

Eight studies reported adjusted results for PFS (Casasnovas 2019; Gallamini 2014; Hutchings 2005; Hutchings 2006; Kobe 2018; Mesguich 2016; Rossi 2014; Simon 2016). Three studies took the importance of adjustment into account, but did not actually conduct a multivariable analysis (Annunziata 2016; Hutchings 2014; Oki 2014).

Definitions of Progression-free survival (PFS)

The definition of the progression outcome varied between studies. Four studies that reported PFS did not provide a definition (Hutchings 2014; Simon 2016; Straus 2011; Zaucha 2017). One study analysed event-free survival (Cerci 2010), which was identical with PFS and, therefore, included in the analysis. Table 2 presents an overview of definitions used for progression outcome. Studies with identical definitions were grouped.

Table 2. Definitions of progression outcomes

Study	Definition of progression outcome
Andre 2017	Progression-free survival, defined – from the date of random assignment to date of progression – as experiencing relapse after previous complete remission or progression after reaching partial remission (50% decrease and resolution of B symptoms and no new lesions); progressive dis-

	ease (50% increase from nadir of any previous partial remission lesions or appearance of new lesions) on CT scan measurements during protocol treatment; or death from any cause, whichever occurred first.
Casasnovas 2019	Progression-free survival defined as the time from randomisation to first progression, relapse, or death from any cause or last follow-up.
Annunziata 2016	The primary endpoint was PFS, with progression during treatment, lack of complete remission at the end of the first-line treatment, and relapse counted as adverse events.
Barnes 2011; Ying 2014; Zinzani 2012	Progression-free survival is defined as the time from diagnosis to progression or death from any cause.
Kobe 2018	Progression-free survival is defined as the time from completion of staging until progression, relapse, or death from any cause, or to the day when information was last received on the patient's disease status.
Cerci 2010	Three-year event-free survival was chosen as the endpoint and defined as the time from diagnosis to treatment failure or last follow-up. Treatment failure was defined as an incomplete response after first-line treatment, progression during therapy, relapse, or death.
Gallamini 2014; Markova 2012; Mesguich 2016; Oki 2014;	Progression-free survival is defined as the time from diagnosis to either disease progression or relapse, or to death as a result of any cause, whichever occurred first.
Hutchings 2005; Hutchings 2006	Progression-free survival is defined as the time from diagnosis to first evidence of progression or relapse, or to disease-related death.
Okosun 2012	Progression-free survival is defined as the time from diagnosis to disease progression or relapse or last follow-up.
Rossi 2014	Progression-free survival is defined as the time from the beginning of treatment until progression, relapse, or death from any cause or the date of last follow-up. Time-to-progression (TTP) is defined as the time from the date of the first course of chemotherapy to any treatment failure, including progression, relapse, or death related to lymphoma, or the date of the last follow-up.
Touati 2014	Progression-free survival is defined as the time from diagnosis to relapse or death.
Hutchings 2014; Simon 2016; Straus 2011; Zaucha 2017	Definition not reported.

Adverse events (AEs)

None of the included studies measured PET-associated AEs.

Conflict of interest

Two studies reported potential conflicts of interest (Andre 2017; Casasnovas 2019). Fourteen studies declared that the investigators had no conflict of interest (Annunziata 2016; Barnes 2011; Hutchings 2006; Hutchings 2014; Kobe 2018; Mesguich 2016; Oki 2014; Okosun 2012; Orlacchio 2012; Rossi 2014; Simon 2016; Straus 2011; Zaucha 2017; Zinzani 2012). Seven studies did not report investigators' disclosures of potential conflicts of interest (Cerci 2010; Gallamini 2014; Gandikota 2015; Hutchings 2005; Markova 2012; Touati 2014; Ying 2014).

Excluded studies

After screening titles and abstracts, we excluded 10651 references that did not match our inclusion criteria. In addition, we excluded a total of 133 references after full-text screening for the following reasons.

- Fifty-six references had a study design or publication type that did not match our inclusion criteria, i.e. letters and commentaries, case studies with a small sample size or

validation studies (Adams 2016; Adams 2017; Adams 2018; Adams 2018a; Adams 2018b; Adams 2019; Afanasyev 2017; Ansell 2016; Barrington 2017; Bar-Shalom 2003; Basu 2009; Becherer 2002; Bednaruk-Mlynski 2015; Biggi 2012; Bishop 2015; Bodet-Milin 2009; Boisson 2007; Borchmann 2016; Bucerius 2006; Cremerius 1999; D'Urso 2018; Dann 2018; deAndres-Galiana 2015; Diehl 2007; El-Galaly 2012; Evens 2014; Fanti 2008; Friedberg 2002; Friedberg 2004; Gallamini 2008; Gallamini 2018a; Gallowitsch 2008; Guidez 2016; Hagtvedt 2015; Hartmann 2012; Hartridge-Lambert 2013; Kobe 2008; Kobe 2014; Lowe 2002; Milgrom 2017; Mocikova 2010; NCT02292979; Pichler 2000; Reinhardt 2005; Rigacci 2002; Rigacci 2017; Rubello 2015; Sakr 2017; Specht 2007; Spinner 2018; Strigari 2016; Tirelli 2015; Xie 2018; Yasgur 2015; Zabrocka 2016; Zaucha 2009).

- Thirty-nine references adapted the treatment based on PET-results (Albano 2017; Albano 2018; Biggi 2017; Carras 2018; Ciammella 2016; Cuccaro 2016; Damlaj 2017; Damlaj 2019; Danilov 2017; Dann 2009; Dann 2010; Dann 2010a; Dann 2012; Dann 2013; Dann 2016; Dann 2017; Fornecker 2017; Gallamini 2017; Gallamini 2018; Greil 2018; Illidge 2015; Johnson 2015; Johnson 2016; Kamran 2016; Kamran 2018; Moskowitz 2015; NCT00784537; NCT00795613; NCT01358747; NCT01652261; Nguyen 2017; Paolini 2007; Pavlovsky 2019; Simontacchi 2015; Straus 2018; Torizuka 2004; Trotman 2017; Villa 2018; Zinzani 2016).
- Eighteen references also included participants with other types of lymphoma and did not report data for HL separately (Awan 2013; Blum 2002; Bodet-Milin 2008; Cremerius 2001; Filmont 2003; Freudenberg 2004; Fruchart 2006; Goldschmidt 2011; Haioun 2005; Honda 2014; Iagaru 2008; Kostakoglu 2006; Li 2013; Slaby 2002; Tomita 2015; Torizuka 2004; Zinzani 1999; Zinzani 2002).
- Ten references included participants who received treatment other than first-line therapy, i.e. second-line therapy for relapsed or refractory disease (Bjurberg 2006; Front 1999; Huic 2006; Mocikova 2010; Mocikova 2011; Schot 2007; Sucak 2011; Tseng 2012; Weidmann 1999; Yoshimi 2008).
- Eight references reported only end-of-chemotherapy PET-results (Advani 2007; Hueltenschmidt 2001; Hutchings 2007; Jerusalem 2003; Molnar 2010; Naumann 2001; Panizo 2004; Spaepen 2001).
- Two were duplicates (Freudenberg 2004; Kobe 2014).

These publications are described in Characteristics of excluded studies.

Risk of bias in included studies

We assessed the risk of bias at outcome level (OS and PFS) for each study using the QUIPS tool. No study reported PET-associated AE. The detailed assessment can be found in the 'Risk of bias (QUIPS)' section in the Characteristics of included studies.

Risk of bias in studies included in meta-analyses

The 'Risk of bias' summary (Figure 2) presents the combined judgement made by the review authors in a cross-tabulation. Studies included in meta-analysis are highlighted in bold.

Figure 2. 'Risk of bias' assessment according to QUIPS (Quality in Prognostic Studies) by outcome.

Outcome	Study	Quality in Prognostic Studies (QUIPS)					
		Study participation	Study attrition	Prognostic factor measurement	Outcome measurement	Other prognostic factors (covariates)	Statistical analysis and reporting
Overall survival	**Barnes 2011**	Unclear	Low	Moderate	High	Low	High
	Casasnovas 2019	Low	Low	Low	Low	Low	Low
	Cerci 2010	Low	Low	Low	Low	High	High
	Gallamini 2014	Low	Low	Low	Low	Low	Low
	Hutchings 2005	Unclear	Moderate	Low	Low	High	Low
	Hutchings 2006	High	Low	Low	Low	High	Low
	Hutchings 2014	Low	Low	Low	Low	High	Low
	Kobe 2018	Low	Low	Low	Low	Low	Low
	Simon 2016	Unclear	Low	Low	Low	High	High
	Touati 2014	Unclear	Low	Moderate	Low	High	Low
	Zaucha 2017	Low	Low	Moderate	Low	High	Low
	Zinzani 2011	Low	Low	Low	Low	High	Low
Progression-free survival	Andre 2017	Low	Low	Moderate	Low	Low	Low
	Annunziata 2016	Unclear	Unclear	Low	Low	High	High
	Barnes 2011	Unclear	Low	Moderate	High	Low	High
	Casasnovas 2019	Low	Low	Low	Low	Low	Low
	Cerci 2010	Low	Low	Low	Low	High	High
	Gallamini 2014	Low	Low	Low	Low	Low	Low
	Hutchings 2005	Unclear	Moderate	Low	Low	Low	Low
	Hutchings 2006	High	Low	Low	Low	Low	Low
	Hutchings 2014	Low	Low	Low	Low	High	Low
	Kobe 2018	Low	Low	Low	Low	Low	Low
	Markova 2012	Low	Low	Moderate	Low	Low	Low
	Mesguich 2016	Low	Low	Low	Low	Low	Low
	Oki 2014	Low	Low	Low	Low	High	High
	Okosun 2012	Low	Low	Low	Low	Low	High
	Rossi 2014	Low	Low	Low	Low	High	Low
	Simon 2016	Unclear	Low	Low	Low	High	High
	Straus 2011	Low	Low	Low	Low	Low	Low
	Touati 2014	Unclear	Low	Moderate	Low	High	Low
	Ying 2014	Low	Low	Moderate	Low	High	High
	Zaucha 2017	Low	Low	Moderate	High	High	High
	Zinzani 2011	Low	Low	Low	Low	High	Low

Overall survival (OS)

For our primary outcome OS, one out of nine studies included in meta-analysis was assessed as 'low' in all risk of bias domains (Kobe 2018). Four studies were assessed as 'unclear' for the domain study participation (Barnes 2011; Hutchings 2005; Simon 2016; Touati 2014), mostly due to a lack of information about the baseline population from which the study sample originated. Most studies had defined exclusion criteria to sample participants from the baseline population (e.g. unavailability of interim PET2) without providing a description of the original population or reasons for missing information. Considering this a potential source of selection bias, we assessed this domain as 'unclear' when information about the baseline population was missing. For the domains study attrition, prognostic factor measurement and outcome measurement, risk of bias was assessed as 'low' in most studies. Two studies did not report the use of standardised criteria for prognostic factor measurement, therefore we assessed the risk of bias as 'moderate' (Barnes 2011; Touati 2014). One study was assessed as 'moderate' risk because PET2 availability was dependent on PET1 result (Zaucha 2017). Due to inconsistency in reporting of the timing of the interim PET measurement, the risk of bias for outcome measurement was assessed as 'high' in one study (Barnes 2011), while the remaining studies were all assessed as 'low'. Two studies were assessed as 'low' risk of bias in the domain other prognostic factors (covariates) because they only included participants within one disease stage (e.g. early or advanced stages) (Barnes 2011; Kobe 2018), while the remaining seven studies were assessed as 'high' risk of bias for this domain because they included all disease stages without adjusting for stage (Cerci 2010; Hutchings 2005; Hutchings 2014; Simon 2016; Touati 2014; Zaucha 2017; Zinzani 2012). Six studies provided sufficient information about the methods used for univariable analysis (Hutchings 2005; Hutchings 2014; Kobe 2018; Touati 2014; Zaucha 2017; Zinzani 2012), therefore we assessed the risk of bias for statistical analysis and reporting as 'low'. The same domain was assessed as 'high' in three studies due to discrepancies between text and figures and/or tables (Barnes 2011; Cerci 2010; Simon 2016).

Progression-free survival (PFS)

For our secondary outcome PFS, two out of 14 studies included in meta-analysis were assessed as 'low' risk of bias in all domains (Casasnovas 2019; Mesguich 2016). Eight studies provided clear descriptions of study characteristics and participants (Cerci 2010; Kobe 2018; Mesguich 2016; Rossi 2014; Straus 2011; Ying 2014; Zaucha 2017; Zinzani 2012), so we assessed the risk of bias as 'low'. Five studies did not report inclusion and/or exclusion criteria (Annunziata 2016; Barnes 2011; Hutchings 2005; Simon 2016; Touati 2014), so we assessed the risk of bias for study participation as 'unclear'. One study reported a high number of participants with unavailable interim PET scans without further information (Hutchings 2006), so we assessed the risk of bias as 'high' in the same domain. Most studies had no loss to follow-up to report or

provided a clear description of how missing data were handled, so we assessed the risk of bias for study attrition as 'low' in the majority of studies. One study was assessed as 'unclear' due to a lack of information regarding loss to follow-up (Annunziata 2016); another study was assessed as 'moderate' because no explanation was provided as to why some participants were lost to follow-up (Hutchings 2005). The risk of bias for the domains prognostic factor measurement and outcome measurement was assessed as 'low' in most studies. Three studies did not report the use of standardised criteria for prognostic factor measurement, therefore we assessed the risk of bias as 'moderate' (Barnes 2011; Touati 2014; Ying 2014). A fourth study was assessed as 'moderate' risk because PET2 availability was dependent on PET1 result (Zaucha 2017). Due to lack of outcome definition or inconsistency in the reporting of the timing of the interim PET measurement, the risk of bias for outcome measurement was assessed as 'high' in one study (Barnes 2011). In another study, this domain was also assessed as 'high' because the outcome was not defined (Zaucha 2017). The remaining studies were all assessed as 'low' for the domain outcome measurement. For the domain other prognostic factors (covariates), six studies were assessed as 'low' risk of bias, because they either included participants within one disease stage only, or if all disease stages were included, the authors adjusted for disease stage (Barnes 2011; Hutchings 2005; Hutchings 2006; Kobe 2018; Mesguich 2016; Straus 2011). The remaining eight studies were assessed as 'high' risk of bias for this domain (Annunziata 2016; Cerci 2010; Rossi 2014; Simon 2016; Touati 2014; Ying 2014; Zaucha 2017; Zinzani 2012). Eight studies provided sufficient information about the methods used for univariable analysis (Hutchings 2005; Hutchings 2006; Kobe 2018; Mesguich 2016; Rossi 2014; Straus 2011; Touati 2014; Zinzani 2012), so we assessed the risk of bias for statistical analysis and reporting as 'low'. Five studies were assessed as 'high' for this domain because of the poor reporting of results (Annunziata 2016; Barnes 2011; Cerci 2010; Simon 2016; Ying 2014), including discrepancies between text and figures and/or tables in some studies. Another study was also assessed as 'high' because the method of analysis was not sufficiently described (Zaucha 2017).

Risk of bias in studies reported narratively

The risk of bias for all studies reported narratively is included in Figure 2.

Overall survival (OS)

The results for OS from three studies are reported narratively in this review (Casasnovas 2019; Gallamini 2014; Hutchings 2006). For two studies (Casasnovas 2019; Gallamini 2014) we assessed the risk of bias as 'low' in all six domains of the QUIPS tool. For Hutchings 2006, the first four domains were assessed as 'low' risk of bias. For the domain study participation, the study was assessed as 'high' risk because a great number of participants initially included in the study did not undergo an early interim PET. The study was also assessed as 'high' risk for the domain other prognostic factors (covariates) because participants within all disease stages were included.

Progression-free survival (PFS)

For PFS, the results from seven studies are reported narratively (Andre 2017; Casasnovas 2019; Gallamini 2014; Hutchings 2014; Markova 2012; Oki 2014; Okosun 2012). Out of these, two studies (Casasnovas 2019; Gallamini 2014) were assessed as 'low' risk of bias in all six domains of the QUIPS tool. From the remaining five studies, all were assessed as 'low' risk of bias for the domain study participation. For the domains study attrition, prognostic factor measurement and outcome measurement, three studies were assessed as a 'low' risk of bias (Hutchings 2014; Oki 2014; Okosun 2012). For the other two studies (Andre 2017; Markova 2012), the domain prognostic factor measurement was assessed as 'moderate' risk because the prognostic factor was measured differently in some participants. For the domain other prognostic factors (covariates), five studies were assessed as 'low' risk of bias (Andre 2017; Casasnovas 2019; Gallamini 2014; Markova 2012; Okosun 2012). The other two studies were assessed as 'high' risk for this domain because they included all disease stages without adjusting for disease stage (Hutchings 2014; Oki 2014). Regarding the domain statistical reporting and analysis, five studies were assessed as 'low' risk because they used appropriate methods for the planned analysis (Andre 2017; Casasnovas 2019; Gallamini 2014; Hutchings 2014; Markova 2012). The remaining two studies were assessed as 'high' risk due to inconsistent conduct and reporting of the analyses (Oki 2014; Okosun 2012).

Other potential sources of bias

Reporting deficiencies and selective reporting

We detected reporting deficiencies in some of the studies, particularly when not all analyses that were planned in the methods were actually conducted. In some cases, this was due to the low number of events (i.e. in PET-negative participants) that did not allow for further analyses. In other cases, it was unclear why certain analyses were performed and others not. This was particularly the case with regard to multivariable analyses, when studies planned to assess the independent prognostic ability of the interim PET in a prognostic model including other clinically relevant prognostic factors (covariates). Studies either did not perform such an analysis even though they initially planned to, or they did not consider adjustment. None of the studies stated clearly their rationale for the choice of covariates; in some cases, the choice was based on their significance in univariable analysis. For example, in studies that only included two or less covariates in the model in addition to interim PET, the interim PET was always independent in its performance. However, how interim PET possibly performed in comparison to other covariates remains unclear. Hence, it is particularly important to state why certain covariates were taken into account. Thus, we cannot be sure that studies did not only report certain positive ('significant') results, which can be an issue of selective reporting.

In addition, we detected discrepancies in the reporting of results within the texts of some studies, or between text and the corresponding graph(s) (i.e. in the reporting of the HR or number of events). In these cases, we tried to contact the corresponding principal investigator(s) for clarification in order to have a better understanding of the results.

Blinding of prognostic factor assessor

Eleven studies reported that the clinicians evaluating the interim PET scans were blinded to the outcome (Kobe 2018; Gallamini 2014; Gandikota 2015; Hutchings 2006; Hutchings 2014; Mesguich 2016; Oki 2014; Rossi 2014; Straus 2011; Zaucha 2017; Zinzani 2012).

Results of the analyses

Twenty-three studies evaluated interim PET as a prognostic factor in individuals with HL. Two studies did not report data for our outcomes of interest (Gandikota 2015; Orlacchio 2012) and we have not been able to either obtain or estimate any relevant data. None of the included studies reported PET-associated AEs. Fifteen studies were included in meta-analyses. Another six of the included studies in this review reported results for OS and/or PFS, but we were not able to pool results because, despite our approaches for possible estimation of missing data items, there was a lack of accurate information or data to do so (Andre 2017; Casasnovas 2019; Gallamini 2014; Markova 2012; Oki 2014; Okosun 2012). For all studies that were not included in meta-analyses, we reported the main results narratively in this review.

Overall survival (OS)

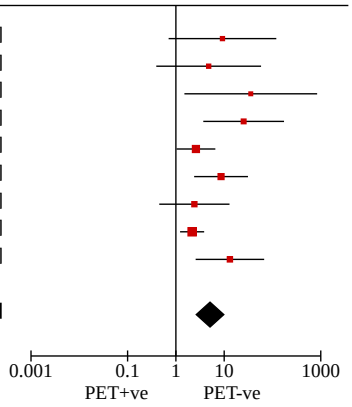
Meta-analysis of unadjusted results

We included nine studies with 1802 participants in meta-analysis for OS (Barnes 2011; Cerci 2010; Hutchings 2005; Hutchings 2014; Kobe 2018; Simon 2016; Touati 2014; Zaucha 2017; Zinzani 2012). There were 475 interim PET-positive and 1327 interim PET-negative participants. Meta-analysis shows a clear advantage in OS for participants with a negative interim PET scan compared to participants with a positive interim PET scan (HR 5.09, 95% CI 2.64 to 9.81, I^2 44%, moderate certainty of evidence) (Analysis 1.1) (Figure 3).

Figure 3. Forest plot of comparison: 1 Univariable comparison of PET+ve vs. PET-ve, outcome: 1.1 Overall survival

Study or Subgroup	log[Hazard Ratio]	SE	PET+ve Total	PET-ve Total	Weight	Hazard Ratio IV, Random, 95% CI
Barnes 2011	2.220623	1.309795	17	79	5.3%	9.21 [0.71 , 120.03]
Cerci 2010	1.563053	1.274372	30	74	5.6%	4.77 [0.39 , 58.02]
Hutchings 2005	3.570366	1.614442	22	63	3.7%	35.53 [1.50 , 841.02]
Hutchings 2014	3.231135	0.981951	37	89	8.3%	25.31 [3.69 , 173.42]
Kobe 2018	0.9555	0.4724	236	486	18.6%	2.60 [1.03 , 6.56]
Simon 2016	2.153725	0.654523	32	89	13.9%	8.62 [2.39 , 31.08]
Touati 2014	0.881751	0.854282	24	44	10.1%	2.42 [0.45 , 12.89]
Zaucha 2017	0.774	0.2928	24	152	24.1%	2.17 [1.22 , 3.85]
Zinzani 2012	2.575668	0.833487	53	251	10.4%	13.14 [2.57 , 67.31]
Total (95% CI)			**475**	**1327**	**100.0%**	**5.09 [2.64 , 9.81]**

Heterogeneity: $Tau^2 = 0.38$; $Chi^2 = 14.39$, df = 8 (P = 0.07); $I^2 = 44\%$
Test for overall effect: Z = 4.87 (P < 0.00001)
Test for subgroup differences: Not applicable



Subgroup analysis

We conducted subgroup analyses to explore the underlying clinical heterogeneity between the studies.

For subgroup analysis by radiotherapy, we found evidence on subgroup difference between the groups (P = 0.05, INRT/ISRT in three studies: N = 548, IFRT in four studies: N = 428, RT not further specified in two studies: N = 826). Results still show an advantage in OS for PET-negative participants, irrespective of the type of radiotherapy they received (Analysis 2.1).

For the remaining subgroups, there was no evidence of subgroup differences.

- Different study designs (P = 0.28; three prospective studies: N = 406, four retrospective studies: N = 589, one RCT: N = 722) (Analysis 2.2). One study (Hutchings 2005) was not included in this subgroup analysis because they did not explicitly state their study design.
- Different chemotherapy regimens (P = 0.33; ABVD in five studies: N = 801, ABVD and other in three studies: N = 279, BEACOPP in one study: N = 722) (Analysis 2.3). Chemotherapy-regimen in the included studies was mainly ABVD, with differentiating numbers of cycles, with or without radiotherapy (Barnes 2011; Cerci 2010; Hutchings 2014; Simon 2016; Zaucha 2017; Zinzani 2012). In Hutchings 2005, the majority of participants received ABVD, while the remaining received MOPP or MOPP/ABV, or another regimen which was not specified. Some participants also received additional radiotherapy. In Kobe 2018, all participants received eBEACOPP. In Touati 2014, the regimens included ABVD, MOPP/ABV hybrid or BEACOPP. If separate data had been available for each type of chemotherapy, we could have performed more specific subgroup analysis to test for differences between chemotherapies.
- PET-CT versus PET (P = 0.66; PET-CT in five studies: N = 595, PET only in three studies: N = 1111) (Analysis 2.4). One study (Barnes 2011) was not included in this subgroup analysis because they conducted PET in some participants and PET-CT in the other participants.
- Different stages of disease (P = 0.33; early stages with A or B symptoms in one study: N = 96, all stages in seven studies: N = 984, advanced stages in one study: N = 722) (Analysis 2.5). One study included disease stages IA, IB, IIA and IIB (Barnes 2011) and another study included advanced-stages only (Kobe 2018). The remaining seven studies included participants representing all disease stages of HL.

Sensitivity analysis

We conducted sensitivity analyses for the timing of interim PET (removing those that did not conduct a PET2), and the precision of

the estimated HR and SE (removing the studies with imprecise HR and SE estimation).

Regarding the timing of the interim PET, interim PET2 was conducted in six studies (N = 1495 participants in total) (Cerci 2010; Kobe 2018; Simon 2016; Touati 2014; Zinzani 2012; Zaucha 2017). In three studies (N = 307 participants in total), interim PET was conducted at other timings: in Barnes 2011, 41 participants received PET2 while the rest of the participants received PET3; in Hutchings 2005, 55 participants received PET2 and 35 participants received PET3; and in Hutchings 2014, PET1 was conducted for all participants (N = 126). Although 89 out of 126 also received a PET2, we used the data for PET1 as the publication provided us with the most information on PET1. At sensitivity analysis, temporarily removing studies that did not perform a PET2 slightly affected the pooled OS (overall: HR 5.09, 95% CI 2.64 to 9.81; sensitivity: HR 3.53, 95% CI 1.97 to 6.32) (Analysis 2.6). It seems that there was an overestimation of the HR for the studies that did not perform a PET2. However, the direction of the effect is firm and unchanged. This difference may also be partly explained by the very wide follow-up ranges within the studies. Hence, following the sensitivity analysis, we consider the overall OS to be robust.

Regarding the precision of the HR estimation, we were able to either obtain or estimate a precise HR and SE for seven studies (N = 1638 participants in total) (Cerci 2010; Hutchings 2005; Hutchings 2014; Kobe 2018; Simon 2016; Zaucha 2017; Zinzani 2012). For two studies (N = 164 participants in total) (Barnes 2011; Touati 2014), we were only able to provide imprecise estimations of the HR and SE. Temporarily removing the imprecise studies during sensitivity analysis barely affected the pooled results for OS, indicating that the measurements obtained from our imprecise method were quite accurate after all (overall: HR 5.09, 95% CI 2.64 to 9.81; sensitivity: HR 5.70, 95% CI 2.60 to 12.48) (Analysis 2.7). Hence, we concluded that the overall pooled OS is robust.

Narrative reporting of results

Univariable analyses

Three studies (Casasnovas 2019; Gallamini 2014; Hutchings 2006) that reported results for OS were not included in meta-analysis due to lack of adequate data for estimating the HR and associated SE (Table 3).

Table 3. Narrative reporting of results from univariable analysis for OS

Study	No. of participants + stages	Timing of interim PET scan	Unadjusted results for interim PET scan
Casasnovas 2019	Standard arm N = 413	PET2	PET2 results (N = 398) ***Intention-to-treat analysis*** 5-year OS for entire arm = 95·2% (95% CI 91·1 to 97·4), 13 events ***Per-procotol analysis*** N = 372 participants 5-year OS for entire arm = 95.6% (95% CI 91.2 to 97.8), 10 events **Comment:** Separate results for PET2-negative and PET2-positive participants in the standard arm were not reported for this outcome.
Gallamini 2014	260 (stages IIA - IVB)	PET2	PET-negative N = 215, 2 deaths, 3-year OS = 99% PET-positive N = 45, 6 deaths, 3-year OS = 87% **Comment:** Logrank test for difference between groups was not reported and could not be obtained.
Hutchings 2006	77 (all stages)	PET2 and PET4	PET2 results (N = 77) PET-negative N = 61, no deaths PET-positive N = 16, 2 deaths Logrank test for difference between groups: P < .01 PET4 results (N = 64) PET-negative N = 51, no deaths PET-positive N = 13, 2 deaths

Comment: Logrank test for difference between groups after PET4 was not reported and could not be obtained.

Multivariable analyses

Two studies (Kobe 2018; Simon 2016) reported adjusted effect estimates to test the prognostic ability of PET2 in addition to other prognostic factors. Table 4 displays a list of established prognostic factors (Cuccaro 2014; Josting 2010; Kılıçkap 2013), and shows which were considered as covariates in the final multivariable model. The selection of prognostic factors (covariates) for the final model was either based on the literature (Simon 2016), or on their significance in univariable analysis (Kobe 2018). However, pooling of adjusted data was not possible. In Simon 2016, only the results of those covariates that remained independent prognostic markers in multivariable analysis, namely LMR and PET2-positivity, were reported. It is unclear whether, or which other covariates were included in the final model. A full list of study-specific, candidate covariates can be found in the respective table for each study in the Characteristics of included studies.

The statistical methods used were Cox proportional hazards regression model and logistic regression model, which are the appropriate methods for a multivariable analysis.

Table 4. Adjusted results from final multivariable model for OS

Study	Prognostic factors								Adjusted results for interim PET
	Interim PET	Age	Gender	Disease stage	B symptoms	Bulky disease	IPS	Other study-specific factors	
Kobe 2018	x	-	-	-	-	-	x	x	Interim PET-positivity (DS 4) HR 3.2 (95% CI 1.3 to 8.4), P = 0.02 **Comment:** Adjusted results indicate an independent prognostic impact of PET2.
Simon 2016	x	-	-	-	-	-	-	x	Interim PET-positivity HR = 11.51 (95% CI 3.14 to 42.86), P < 0.001 **Comment:** Adjusted results indicate the independent prognostic impact of PET2.

x = prognostic factor considered for adjustment in the final model

- = prognostic factor was not considered in the final model

Progression-free survival (PFS)

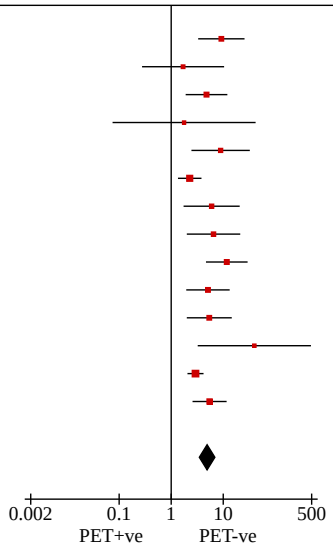
Meta-analysis of unadjusted results

We included 14 studies with 2079 participants in meta-analysis for PFS (Annunziata 2016; Barnes 2011; Cerci 2010; Hutchings 2005; Hutchings 2006; Kobe 2018; Mesguich 2016; Rossi 2014; Simon 2016; Straus 2011; Touati 2014; Ying 2014; Zaucha 2017; Zinzani 2012). There were 529 interim PET-positive and 1550 interim PET-negative participants. Meta-analysis shows a clear advantage in PFS for participants with a negative interim PET scan compared to participants with a positive interim PET scan (HR 4.90, 95% CI 3.47, 6.90, I^2 = 45%, very low certainty of evidence) (Analysis 1.2) (Figure 4).

Figure 4. Forest plot of comparison: 1 Univariable comparison of PET+ve vs. PET-ve, outcome: 1.2 Progression-free survival

Study or Subgroup	log[Hazard Ratio]	SE	PET+ve Total	PET-ve Total	Weight	Hazard Ratio IV, Random, 95% CI
Annunziata 2016	2.2192	0.5231	12	56	7.1%	9.20 [3.30 , 25.65]
Barnes 2011	0.5261	0.9261	17	79	3.0%	1.69 [0.28 , 10.39]
Cerci 2010	1.5624	0.4706	30	74	8.1%	4.77 [1.90 , 12.00]
Hutchings 2005	0.570366	1.6144	22	63	1.1%	1.77 [0.07 , 41.87]
Hutchings 2006	2.187	0.6587	16	61	5.2%	8.91 [2.45 , 32.40]
Kobe 2018	0.8198	0.2651	236	486	13.4%	2.27 [1.35 , 3.82]
Mesguich 2016	1.7891	0.6333	16	60	5.5%	5.98 [1.73 , 20.70]
Rossi 2014	1.873	0.6031	13	46	5.9%	6.51 [2.00 , 21.22]
Simon 2016	2.4596	0.4697	32	89	8.1%	11.70 [4.66 , 29.38]
Straus 2011	1.6268	0.49	24	64	7.7%	5.09 [1.95 , 13.29]
Touati 2014	1.685	0.5075	24	44	7.4%	5.39 [1.99 , 14.58]
Ying 2014	3.6783	1.278	10	25	1.7%	39.58 [3.23 , 484.51]
Zaucha 2017	1.0753	0.1812	24	152	16.0%	2.93 [2.05 , 4.18]
Zinzani 2012	1.6982	0.3845	53	251	10.0%	5.46 [2.57 , 11.61]
Total (95% CI)			**529**	**1550**	**100.0%**	**4.90 [3.47 , 6.90]**

Heterogeneity: Tau² = 0.16; Chi² = 23.52, df = 13 (P = 0.04); I² = 45%
Test for overall effect: Z = 9.07 (P < 0.00001)
Test for subgroup differences: Not applicable



Subgroup analysis

We conducted subgroup analyses to explore the underlying clinical heterogeneity between the studies.

Regarding the disease stage, we detected a significant difference between the groups (P = 0.02, early stages with A or B symptoms in two studies: N = 184, all stages in eleven studies: N = 1173, advanced stages in one study: N = 722). Results still showed an advantage for PFS in PET-negative participants in any stage of the disease (Analysis 3.4). Twelve studies included all disease stages, while one study included stages IA - IIB (Barnes 2011), and another study included advanced-stages only (Kobe 2018).

For the remaining subgroups, there was no evidence of subgroup differences.

- Different study designs (P = 0.29, three prospective studies: N = 357, eight retrospective studies: N = 827, two RCTs: N = 165) (Analysis 3.1). One study (Hutchings 2005) was not included in this subgroup analysis because they did not explicitly state their study design.
- Different chemotherapy regimen (P = 0.43; ABVD in seven studies: N = 945, ABVD and other chemotherapy in four studies: N = 265, other chemotherapies in three studies: N = 869) (Analysis 3.2). Chemotherapy-regimen was ABVD in seven studies, with or without radiotherapy (Annunziata 2016; Barnes 2011; Cerci 2010; Mesguich 2016; Simon 2016; Zaucha 2017; Zinzani 2012). In two studies, participants received either ABVD, ABV/MOPP, ABVD/COPP, BEACOPP esc., PVAG or radiotherapy only (Hutchings 2005; Hutchings 2006). In Touati 2014, the regimens included ABVD, MOPP/ABV hybrid or BEACOPP. In Ying 2014, participants received either ABVD or BEACOPP. In Kobe 2018, all participants received eBEACOPP. In Rossi 2014, all participants received anthracycline-based chemotherapy, and in Straus 2011, all participants received AVG.
- PET versus PET-CT (P = 0.30; PET-CT in eight studies: N = 707, PET only in five studies: N = 1276) (Analysis 3.3). One study (Barnes 2011) was not included in this analysis because they conducted PET in some participants and PET-CT in the other participants.
- Different radiotherapy (P = 0.29; INRT/ISRT in five studies: N = 651, IFRT in six studies: N = 514, RT not specified in two studies: N = 826, no RT given in one study: N = 88) (Analysis 3.5).

In addition, we detected variations between the studies with regard to the definition of PFS. However, all trials included in meta-analysis reported some progression endpoint such as treatment failure, progression or relapse. We have provided the exact reported definitions in Table 2.

Sensitivity analysis

Regarding the timing of interim PET, interim PET was conducted after two cycles of chemotherapy (PET2) in nine studies (N = 1677 participants in total) (Cerci 2010; Hutchings 2006; Kobe 2018; Rossi 2014; Simon 2016; Straus 2011; Touati 2014; Zaucha

2017; Zinzani 2012). In five studies (N = 402 participants in total), interim PET was conducted at other timings: in Annunziata 2016 all participants received PET1; in Barnes 2011 and Hutchings 2005 participants received either PET2 or PET3; and in Hutchings 2006 and Mesguich 2016 participants received either a PET2, PET3 or PET4. At sensitivity analysis, temporarily removing studies that did not perform a PET2 barely affected the results for PFS (overall: HR 4.90, 95% CI 3.47 to 6.90; sensitivity: HR 4.68, 95% CI 3.14 to 6.98) (Analysis 3.6). Hence, the timing of the interim PET measurement (when conducted at a time point other than PET2) did not affect the overall pooled result for PFS.

Regarding the precision of the HR estimation, we were able to provide a precise estimation of the HR and SE for nine studies (N = 1450 participants in total) (Annunziata 2016; Barnes 2011; Hutchings 2005; Kobe 2018; Rossi 2014; Simon 2016; Straus 2011; Ying 2014; Zaucha 2017). For five studies (N = 629 participants in total) we were only able to provide a slightly imprecise estimation of the HR and SE (Cerci 2010; Hutchings 2006; Mesguich 2016; Touati 2014; Zinzani 2012). However, at sensitivity analysis we found that the imprecise HRs did not significantly affect the pooled results. Temporarily removing the imprecise studies during sensitivity analysis barely affected the pooled results (overall: HR 4.90, 95% CI 3.47 to 6.90; sensitivity: HR 4.69, 95% CI 2.84 to 7.73) (Analysis 3.7). Hence, we concluded that the overall pooled PFS is robust and was not affected by our slightly imprecise method of HR and SE estimation.

Narrative reporting of results

Univariable analyses

Seven studies that reported results for PFS were not included in meta-analysis (Andre 2017; Casasnovas 2019; Gallamini 2014; Hutchings 2014; Markova 2012; Oki 2014; Okosun 2012). Table 5 presents the results from these studies narratively. We extracted all data that were available and relevant to us (i.e. number of interim PET-negative and interim PET-positive participants, number of events and percentages for PFS). Due to strong differences in the reporting between studies, the table presents more information for some studies compared to others.

Table 5. Narrative reporting of results from univariable analysis for PFS

Study	No. of participants analysed	Timing of interim PET scan	Unadjusted results for interim PET
Andre 2017	Favourable: N = 371 standard arm Unfavourable: N = 583 standard arm	PET2	*PET-negative Favourable group: N = 2 events (both relapses) in the ABVD + INRT arm, ITT 5-year PFS rate was 99.0% (95% CI 3.8 to 66.1) Unfavourable group: N = 22 events (16 relapses and 6 deaths not related to HL), ITT 5-year PFS rate was 92.1% (95% CI 88.0 to 94.8) *Results presented here are only for participants without interim PET adaptation (ABVD + INRT arm). Unclear how many of these participants were PET-positive or PET-negative. In total (all participants included in the study), there were 465 PET-negative participants and 361 PET-positive participants. *PET-positive N = 41 events (36 relapses and 5 deaths not related to HL) in the ABVD + INRT arm, ITT 5-year PFS rate was 77.4% (95% CI 70.4 to 82.9)
Casasnovas 2019	Standard arm N = 413	PET2	PET2 results (N = 398) ***Intention-to-treat analysis*** PET2-negative N = 349 participants (88%), 5-year PFS = 88.4% (95% CI 83.3 to 92) PET2-positive N = 49 participants (12%), 5-year PFS = 73.5% (95% CI 58.7 to 83.6) Results for entire standard arm 5-year PFS = 86.2% (95% CI 81.6 to 89.8) 41 participants relapsed or progressed, 14 deaths

			Comment: Logrank test for difference between groups in the standard arm after PET2 was not reported and could not be obtained. ***Per-protocol analysis*** N = 372 participants 5-year PFS = 86.7% (95% CI 81.9 to 90.3) for entire arm
Gallamini 2014	260 (stages IIA - IV)	PET2	PET-negative N = 215, 12 events (progression N = 7, relapse N = 5), 3-year PFS = 95% PET-positive N = 45, 33 events (progression N = 27, relapse N = 6), 3-year PFS = 28% Logrank test for difference between groups: P < 0.0001
Hutchings 2014	121 (all stages)	PET1 (N = 121) PET 2 (N = 89)	PET1 results (N = 126) PET-negative N = 89, 5 events (relapse), 2-year PFS = 94.1% PET-positive N = 37, 22 events (17 primary refractory disease, 5 relapses), 2-year PFS = 40.8% Log-rank test for difference between groups: P < 0.01 PET1 vs. PET2 results (N = 89) Participants scanned after PET1 and 2 PET1-negative 2-year PFS = 98.3% PET1-positive 2-year PFS = 38.5% PET2-negative 2-year PFS = 90.2% PET2-positive 2-year PFS = 23.1% 14 PET1-positive converted to a PET2-negative (6 progressed). All PET1-negative were also PET2-negative.
Markova 2012	69 (advanced stages)	PET4	PET-negative N = 51, 2 events (1 relapse and 1 death), % of PFS not reported PET-positive N = 18, 4 events (progression or relapse), % of PFS not reported Log-rank test for difference between groups: P = 0.016
Oki 2014	229 (all stages)	PET2	3-year PFS rates in PET2-negative versus PET–positive by disease subgroups Early stage favourable: 100% vs. 100% Early stage unfavourable: 91.5% vs. 56.3% (P < 0.0001) Early stage non-bulky: 95.9% vs. 76.9% (P = 0.0018) Stage II bulky: 83.3% vs. 20% (P = 0.017) Advanced stage with IPS≤2: 77.0% vs. 30.0% (P < 0.001) Advanced stage with IPS≥3: 71.0% vs. 44.4% (P = 0.155)
Okosun 2012	23 (stages II - IV)	PET2 or PET3	PET-negative: N = 21, no events, 2-year PFS = 100%

PET-positive: N = 2, 1 event (treatment failure), 2-year PFS = 50%

Log-rank test for difference between groups: P = 0.0012

Multivariable analyses

Eight studies reported adjusted effect estimates for PFS (Casasnovas 2019; Gallamini 2014; Hutchings 2005; Hutchings 2006; Kobe 2018; Mesguich 2016; Rossi 2014; Simon 2016). Table 6 shows which prognostic factors (covariates) were considered in the final multivariable model of the studies. In two studies, only the results of those covariates that remained independent prognostic factors in multivariable analysis were reported (Gallamini 2014; Simon 2016). It is unclear whether, or which other covariates were included in the final multivariable model. The selection of prognostic factors (covariates) for adjustment in the studies was either based on their significance in univariable analysis (Casasnovas 2019; Hutchings 2006; Kobe 2018), or on the literature (established prognostic factors) (Hutchings 2005; Rossi 2014; Simon 2016). In two studies, the rationale for the covariates was not clearly stated (Gallamini 2014; Mesguich 2016).

As there are no final models with an identical set of covariates, pooling of adjusted effect estimates was not feasible. A full list of study-specific, candidate covariates can be found in the respective table for each study in the Characteristics of included studies.

The statistical methods used were Cox proportional hazards regression model and logistic regression model.

Table 6. Adjusted results from final multivariable model for PFS

Study	Prognostic factors								Adjusted results for interim PET
	Interim PET	Age	Gender	Disease stage	B symptoms	Bulky disease	IPS	Other study-specific factors	
Casasnovas 2019	x	-	x	x	x	x	x	x	Multivariable analysis not reported separately for standard treatment group.
Gallamini 2014	x	-	-	-	-	-	-	x	PET2 HR N/A, P < 0.01 (Sig. 0.000), 95% CI 3.136 to 7.917 **Comment:** Adjusted results indicate the independent prognostic impact of interim PET2.
Hutchings 2005	x	-	-	x	-	-	-	x	Early interim PET Wald 19.05, HR N/A, P-value = 0.00007 **Comment:** Adjusted results indicate the independent prognostic impact of early interim PET.
Hutchings 2006	x	-	-	x	-	-	-	x	*Model 1 (interim PET2 + clinical stage + extranodal disease)* PET2 HR = 36.281 (95% CI 7.179 to 183.4), P < .001 *Model 2 (interim PET2 + extranodal disease)* PET2 HR = 36.887 (95% CI 7.338 to 185.4), P < .001 **Comment:** Adjusted results indicate the independent prognostic impact of interim PET2.
Kobe 2018	x	-	-	-	-	-	x	x	Interim PET-positivity (DS 4) HR 2.4 (95% CI 1.4 to 4.1), P = 0.002

									Comment: Adjusted results indicate an independent prognostic impact of PET2.
Mesguich 2016	**x**	-	-	**x**	-	**x**	-	-	*Model 1 (interim PET + disease stage)* Positive interim PET HR = 3.73 (95% CI 1.35 to 10.35), P = 0.0112 *Model 2 (interim PET + bulky disease)* Positive interim PET HR = 3.62 (95% CI 1.30 to 10.05), P = 0.0138 **Comment:** Adjusted results indicate the independent prognostic impact of interim PET.
Rossi 2014	**x**	-	-	-	-	-	-	**x**	SUVmax PET0-PET2 Relative risk = 7.9 (95% CI 2.9 to 22.9), P = 0.0001 **Comment:** Adjusted results indicate the independent prognostic impact of SUVmax PET0-PET2.
Simon 2016	**x**	-	-	-	-	-	-	**x**	Interim PET-positivity HR = 17.74, P < 0.001, 95% CI 6.61 to 47.57 **Comment:** Adjusted results indicate the independent prognostic impact of PET2.

x = prognostic factor considered for adjustment in the final model

- = prognostic factor was not considered in the final model

Adverse events (AEs)

None of the included studies measured PET-associated AE.

Studies not reporting our outcomes

Two studies (Gandikota 2015; Orlacchio 2012) did not report data for our outcomes of interest, but were still included in this review as they fit our inclusion criteria. Their investigated outcomes were very close to our review outcomes and potentially the authors could have measured them, but did not report them in their publication. However, it has not been possible to obtain the relevant information; therefore, they are reported narratively in this review. Table 7 presents the results from these studies narratively.

Table 7. Narrative reporting of results from studies not reporting our outcomes of interest

Study	No. of participants	Outcomes/comparison	Results
Gandikota 2015	77 (stages IIA - IIB)	• Analysis of imaging at different time points: Baseline imaging, imaging during (after two to four cycles of ABVD) and at the end of treatment, follow-up imaging • Need for surveillance imaging	Analysis of imaging at different time points *Baseline imaging* • 77 participants had baseline PET-CT scans, 1 had only chest X-ray due to pregnancy at baseline *Imaging during and at the end of treatment* • 77 participants had interim PET-CT during chemotherapy (N = 34) or after chemotherapy before initiation of radiotherapy (N = 43) • Out of 77, 4 remained PET-positive, scans after completion of radiotherapy showed a complete response in 2/4, inflammation in 1/4, resolution of all adenopathy in 1/4, 0/4 relapsed during follow-up *Follow-up imaging* • Median follow-up: 46 months (range 24 to 126) • Total of 466 scans in 78 participants (PET-CT in N = 42) • No relapses occurred in the entire cohort, N = 3 were diagnosed with a second primary malignancy by either imaging or clinical presentation, N = 6 had false-positive imaging findings (3/6 PET-CT) requiring further supplementary imaging or biopsy/surgery Need for surveillance imaging Quote: “No relapse of cHL was detected at a median follow-up of 46 months. […] Routine imaging (either CT or PET-CT) for the early detection of relapse does not appear necessary or justified in these participants.”
Orlacchio 2012	132 (all stages)	Interim PET2 vs. end PET (three months after the end of chemo- and radiotherapy).	Interim PET results • Negative interim PET2: 104 • Positive interim PET2: 28 End PET results *Negative interim PET2 group* • Negative final PET: 102/104 • Positive final PET: 2/104 *Positive interim PET2 group* • Negative final PET: 16/28 • Positive final PET: 12/28

Interim PET vs. end PET

Negative interim PET2 group

- Quote: “Final PET confirmed the negative results in 102 cases (98%) and revealed pathological uptake in the remaining two cases (2%).”

Positive interim PET2 group

- Of the 28 interim PET-positive participants, 19 showed a partial response and nine had disease stability or progression. Twelve of the 28 interim PET-positive participants had a positive final PET. Hence, the remaining 16 had a negative final PET.

NPV and PPV

- Quote: “Interim PET had a NPV of 98%, with 85.7% sensitivity, 86.4% specificity and 86.4% diagnostic accuracy.”
- Quote: “[In univariable analysis] the only independent predictor is the result of interim PET. […] PET had a PPV of 42%."

DISCUSSION

Summary of main results

In this systematic review, we summarised unadjusted data for interim positron emission tomography (PET) scan results as a prognostic factor in individuals with classic Hodgkin lymphoma (HL). The results of an interim PET scan during therapy, e.g. after two cycles of chemotherapy, has been suggested as a good predictor of outcome. Interim PET scan results have also been suggested as an indicator to guide further treatment in order to achieve the best possible outcome in those that have a poor prognosis and those that have a good prognosis, while also minimising adverse events due to the toxicity of the chemotherapy. The results of our review are summarised in the Summary of findings 1.

The findings emerging from meta-analyses are as follows.

- Unadjusted results for overall survival (OS) show a large advantage for participants with a negative interim PET scan result compared to participants with a positive interim PET scan result. We rated the certainty of the evidence as 'moderate'.
- Unadjusted results for progression-free survival (PFS) show an advantage for participants with a negative interim PET scan result compared to participants with a positive interim PET scan result, but the evidence is very uncertain. We rated the certainty of the evidence as 'very low'.

The findings of the adjusted results from multivariable analyses, reported narratively in this review, are as follows.

- Adjusted results for OS indicate an independent prognostic ability of interim PET beyond other associated factors. We rated our certainty of the evidence as 'moderate'.
- Adjusted results for PFS indicate that there may be an independent prognostic ability of interim PET beyond other associated factors. We rated our certainty of the evidence as 'low'.

No study measured adverse events (AEs) associated with PET.

Overall completeness and applicability of the evidence

The evidence in this review mostly applies to adults who were newly diagnosed with classic HL, and who receive a PET scan in combination with CT (PET-CT) after two cycles of chemotherapy (PET2). The studies included in this review addressed our research question in a total of 7335 male and female participants representing all stages of classic HL (Ann Arbor stages I - IV with A or B symptoms). Nine studies included individuals aged 18 years or older, while the remaining studies also included adolescents and young adults (the youngest being 13 years of age, although most studies started from the age of 16 and onwards). Overall, the findings from this review support the statement that in this group of individuals, interim PET scan results can predict OS and PFS. Most participants in the included studies received ABVD (adriamycin/doxorubicin, bleomycin, vinblastine and dacarbazine) chemotherapy, which is the standard treatment regimen for early-stage disease (Bröckelmann 2018; Engert 2010). However, as participants can have different therapy regimens, which is decided based on their disease stage and other clinical or individual characteristics, results should always be interpreted with caution for different patient groups, and this naturally restrains the applicability of the evidence for all people with classic HL. Twelve out of 23 studies reported our primary outcome of interest OS, while 21 studies reported PFS. No study reported PET-associated AE. As the main aim of the review was to identify the prognostic value of interim PET results to predict survival outcomes, it is unlikely that studies on prognosis will measure or report AE.

Heterogeneity between the studies was also found with regard to the evaluation of the interim PET scan, as studies used different criteria for the interpretation of the results. Most studies used the Deauville five-point scale (DS 1 - 5) for the evaluation of the PET scans. However, different cut-off values were used for PET-positivity. Most studies considered scores one to three (DS 1-3) for PET-negativity, and scores four to five (DS 4-5) for PET-positivity. In some studies, however, DS3 was also considered (or tested) for

PET-positivity. Results from these studies should be interpreted with caution, as using a score of ≥3 can have an important impact on the results and possibly introduce bias. Firstly, using this cut-off can lead to an increased number of false-positive results for interim PET (Casasnovas 2019). This can have a relevant impact for the individual, if treatment would be modified based on the interim PET scan results (such as in the studies by Andre 2017; Casasnovas 2019; Kobe 2018). Furthermore, using this cut-off can lead to an overestimation of the positive outcomes in the interim PET-positive group. In the study by Kobe 2018, in which cut-off DS3 and DS4 were tested for PET-positivity, the results showed no significant difference in DS1-2 compared to DS3, but a significant difference between DS1-3 and 4. Thereby, the authors argue for DS4 as the cut-off value for PET-positivity, which is interpreted as an [18F]-fluorodeoxy-D-glucose (FDG) uptake higher than in the liver, instead of an uptake higher than in the mediastinum (corresponding DS3) (Kobe 2018). Hence, the implementation of a commonly used cut-off in clinical practice is important in order to improve interobserver reliability and agreements between central reviewers, and is also highly crucial for the individual (Kobe 2018; Meignan 2009a). In the remaining studies included in this review, either different criteria were used (e.g. International Harmonization Project in Lymphoma criteria (Juweid 2007)), or no specific scale was indicated. However, in most studies, at least two nuclear medicine physicians independently interpreted the interim PET scan results.

One of the greatest issues regarding the prognostic factor studies in this review relates to the difficult reporting of their statistical analyses. Even when the methods of the statistical analyses were appropriate for the study design, the data were insufficiently reported in many of the included studies. We used hazard ratios (HRs) as the effect measure for time-to-event data in this review. We were able to pool data from only 15 studies, either because the HR and associated standard error (SE) were not reported, or because we did not have separate data for our participants or outcomes of interest. Out of these 15 studies, six studies reported an HR, but we still re-calculated the value for four of them for different reasons. For example, values were re-calculated either when we detected discrepancies between the text and corresponding graph(s) and table(s), or when they were simply not reported, while other relevant data were, helping us to estimate the HR and SE. For the remaining studies, we estimated the HR using other available data where possible (Altman 2012; Parmar 1998; Tierney 2007; Trivella 2006). For this reason, we contacted 10 principal investigators to clarify our questions and provide us with additional information or data, or both. This step was particularly helpful for deciding which data to pool.

We prespecified in our protocol that we would only pool adjusted associations of the index prognostic factor if analyses were based on an identical set of covariates. Although this was not feasible for our review, we suggest that future authors of systematic reviews of prognostic factor studies consider pre-specifying a core set of covariates (established prognostic factors) that are important to the disease under review, and should be investigated in the included studies (Riley 2019; Riley 2019b). In this way, authors may be able to pool adjusted effect estimates, if studies are homogenous enough in the adjustment set of the other prognostic factors. In addition, we have moderate between-study heterogeneity, which is reflected in the I^2 and wide confidence intervals (CIs). We took these issues around the reporting in the studies into account when we assessed risk of bias and GRADE.

Furthermore, the pooled estimates of the prognostic effect of the interim PET scan result in our analyses are based on crude HRs (no adjustment for covariates), therefore the reported results are at risk of overestimating the prognostic ability of the interim PET scan result. Hence, in light of the absence of adjustment for other prognostic factors, and considering the risk of bias assessment for the fifth domain of the QUIPS tool, we downgraded the strength of the evidence in our GRADE assessment. This is because it is widely acknowledged that adjusting the predictive effect of a specific prognostic factor for the contribution of other prognostic factors strengthens the robustness of the evidence on the clinically relevant prognostic ability of that factor (Riley 2019; Riley 2019b).

Lastly, although we did not conduct a test for funnel plot asymmetry as this type of test is not necessarily recommended for survival data due to issues of censoring (Debray 2018), we cannot exclude potential publication bias and the presence of small-study effects in our review (Riley 2019). Firstly, we assume that publication bias may be present in our review as most studies in our analyses have rather small sample sizes, of which all present positive results on the prognostic ability of interim PET scan results. Secondly, most studies included in this review are retrospective studies that have not been pre-registered, for example, in trial registries. Studies are also not always labelled or indexed as prognosis studies, and search filters for studies on prognosis are still under development, which is the main reason as to why we conducted a broad search with the disease (HL) and prognostic factor (PET) of interest. This led to a high number of search results that had to be screened. Thirdly, we identified a great number of conference abstracts on studies for which we could not find full-text publications (see Characteristics of studies awaiting classification). Hence, based on these experiences, we cannot preclude that more studies may exist that have either not been published, or not indexed properly.

Certainty of the evidence

Our certainty of the evidence is presented in the Summary of findings 1.

Unadjusted results

For our primary outcome OS, we judged the certainty of the evidence as 'moderate'. We included nine studies in the meta-analysis, of which eight were observational studies and one was a clinical trial. We used the data of participants from the standard arm (no treatment adaptation) of this trial. We judged the certainty of the evidence as 'moderate' due to some methodological issues. We downgraded by one point for risk of bias due to a high risk of bias in seven studies for the domain other prognostic factors (covariates), as well as a high risk of bias in three studies for the domain statistical analysis and reporting. In addition, we downgraded by one point for imprecision because the HR had to be estimated in seven studies, and re-calculated in one study. Hence, only one out of nine studies reported a HR that we used. Nevertheless, we upgraded by one point for a large effect showing the large difference in the OS between interim PET-positive and interim PET-negative participants (HR 5.09, CI 2.64 to 9.81).

For the outcome PFS, we judged the certainty of the evidence as 'very low'. We included 14 studies in the meta-analysis, of which 12

were observational studies and two were clinical trials (participants from the standard arms). For this outcome, we downgraded by one point for inconsistency because the definition of PFS varied across the studies. We also downgraded by one point for imprecision because the HR had to be estimated in 10 studies and re-calculated in one study. Hence, we were able to use a reported HR for only three out of 14 studies. In addition, we downgraded by one point for risk of bias, because of a high risk of bias in eight studies for the domain other prognostic factors (covariates), and high risk of bias in six studies for the domain 'statistical analysis and reporting'.

Adjusted results

For the outcome OS, two studies reported adjusted results from multivariable analyses including established prognostic factors (e.g. International Prognostic Score) in individuals with HL, and the results of both studies indicate the independent prognostic ability of interim PET to predict OS. We judged our certainty in the evidence as 'moderate' for this outcome due to some methodological issues. We downgraded by one point for risk of bias due to a high risk of bias in the domains other prognostic factors (covariates) and statistical analysis and reporting for one study.

For the outcome PFS, there were eight studies that reported adjusted results (adjusted for e.g. disease stage or B symptoms). All studies found that interim PET scan results have an independent prognostic ability to predict PFS. However, we rated our certainty in the evidence as 'low' for this outcome. We downgraded by one point for risk of bias due to a high risk of bias in the domain study participation in one study, as well as a high risk of bias in the domains other prognostic factors (covariates) and statistical analysis and reporting in a second study. Furthermore, we downgraded by one point for inconsistency because the studies included a heterogenous set of covariates in the multivariable analyses, which made the pooling of adjusted results not feasible.

Potential biases in the review process

To prevent bias in this review, two teams of two review authors independently performed all relevant processes (i.e. screening, data extraction, risk of bias and GRADE assessment). Due to the complexity of assessing bias in prognostic factor studies, as well as assessing the certainty of the evidence from these types of studies, we conducted several teleconferences with different experts in the field of prognosis to discuss our assessments. We consulted Jill Hayden (Hayden 2013) for the 'Risk of bias' assessment, and the GRADE for Prognosis working group for the GRADE assessment. In particular, the methods for grading the evidence from prognosis studies are still under development.

For the 'Risk of bias' assessment, we are aware that adding 'unclear' as a fourth possible rating, thereby setting an example for future authors, can lead to a potential bias in the assessment. However, for our assessment we only used 'unclear' when relevant information was evidently missing, thereby making it difficult to make a fair and transparent judgement for the respective study and domain. We felt that rating a domain as high risk of bias in such cases would be inappropriate. We clearly advise against the use of 'unclear' as a default option and want to recommend future authors of reviews of prognosis studies to use this fourth rating carefully (if the fourth rating will be included in an update of the QUIPS tool).

Our analyses included post-hoc subgroup analyses on the type of PET measurement (PET versus PET-CT), as well as post-hoc sensitivity analyses on the timing of the interim PET and the type of estimation (see Methods) used to estimate missing values. These analyses were necessary due to the heterogeneity between the studies. Results should be interpreted in light of differences that can exist when participants receive a PET-CT as compared to a PET scan only. Furthermore, the timing of the interim PET is crucial, as PET1 and PET2 may provide different results compared to PET3 and PET4.

Regarding the adjusted results, we refrained from pooling results because, although the studies looked at established prognostic factors, they did not include identical sets of covariates. As the studies are already very heterogeneous, pooling of the adjusted results was not feasible for our review, as the comparison and interpretation of these results may be problematic in this case. To avoid this in the future, we suggest pre-defining a core set of covariates in order to enable pooling of adjusted results (Riley 2019).

Agreements and disagreements with other studies or reviews

In our review, we included studies that have assessed the prognostic value of interim PET in HL participants without treatment modification. Overall, the findings from this review are in agreement with similar reviews and studies that have investigated the prognostic value of interim PET. Our results are also in agreement with the literature that interim PET can be used for disease and therapy monitoring (Barrington 2017a). Some reviews and studies have investigated this in participants in whom the treatment was changed based on the interim PET scan result, and have come to similar conclusions that interim PET can predict outcome in the different groups (PET-negative and PET-positive participants).

We are aware of three systematic reviews (Adams 2015a; Amitai 2018; Sickinger 2015) that have investigated interim PET as a prognostic factor. Adams 2015a included ten studies with limited-, intermediate- and advanced-stage HL participants in whom the treatment regimen was not modified based on the interim PET scan results. In fact, nine out of these 10 studies are also included in our review. One study was not included in our review because they only included children. The authors of this review concluded that a negative interim PET cannot exclude treatment failure, but that a positive interim PET can identify and predict treatment failure. The authors assessed the quality of the studies with the QUIPS tool (as we did in our review) and judged the overall methodological quality of the included studies as moderate. We have compared their QUIPS assessment with ours for each individual study, and identified that for the domains study participation and study attrition in particular, we found agreements between the authors and our review that there is a low risk of bias in the studies. Disagreement was found regarding the domain prognostic factor measurement, for which the authors judged the quality as moderate mainly due to the heterogeneity between the studies regarding the use of PET-CT versus PET only, which is an issue that we have also addressed in our review by subgroup analysis.

Comparison of interim PET with end PET

Nine of the included studies compared the performances of interim PET and end-of treatment PET (end PET) (Barnes 2011; Hutchings 2006; Hutchings 2014; Markova 2012; Mesguich 2016; Orlacchio

2012; Straus 2011; Ying 2014; Zinzani 2012), as omitting one of the two can have an impact on radiation safety for the patient. However, results between studies are rather contradictory. For example, in Barnes 2011 the authors could not detect a significant difference in OS and PFS between interim PET-negative and interim PET-positive participants. In their analyses, interim PET-positive participants that were negative at end PET had the same good outcomes as participants who were negative both at interim and end PET. In addition, after end PET, the difference between end PET-positive and end PET-negative participants was fairly high, with a greater four-year OS and PFS for end-PET-negative participants. In this study, 74 (end PET) out of 79 participants (interim PET) remained PET-negative, while nine (end PET) out of 17 (interim PET) participants remained PET-positive. The authors concluded that end-PET (after six cycles of chemotherapy) predicts outcome, rather than interim PET (after two or four cycles of chemotherapy). In Hutchings 2006, interim PET was conducted after two and four cycles of chemotherapy (total number of cycles was six to eight). Results show that PET2 and PET4 were similarly successful in predicting outcome in participants, but the authors of the study still argue that treatment modifications should be indicated as early as possible (e.g. after PET2) in order to achieve the best possible outcome. In the study by Mesguich 2016, interim PET was also lower in its predictive ability compared to end PET. Out of 60 interim PET-negative participants, seven converted to a positive end PET. Out of 16 interim PET-positive participants, seven converted to a negative end PET. In addition, treatment failure was most common in participants with a positive end PET as compared to participants with a positive interim PET. The sensitivity of interim PET was measured as 47% compared to 80% of end PET (Mesguich 2016).

Contrastingly, Orlacchio 2012 detected a very high negative predictive value (NPV) of 98% for interim PET2, with an overall diagnostic accuracy of 86.4%. Out of 104 interim PET-negative participants, 102 were still negative after end PET. Out of 28 interim PET-positive participants, however, 16 converted to a negative end PET. A high NPV for interim PET was also found in Hutchings 2005 (interim PET2/3) as interim PET-negative participants rarely relapsed. In Hutchings 2014, 89 participants had an interim PET1 and PET2, and both show a strong prognostic ability for predicting outcome. In this study, none of the participants in early stages that had a negative interim PET1 progressed or relapsed. Advanced-stage participants with a negative interim PET1 had a long-term PFS of more than 90%. The three-year PFS of interim PET1-positive participants was 30%. In total, 89 participants had both PET1 and PET2. Out of these, 62 were PET1-negative, and after treatment, 60 were in complete remission. Twenty-seven participants were PET1-positive, of which 15 were in complete remission. To compare, 76 participants were PET2-negative, of whom 70 were in complete remission. Thirteen participants were PET2-positive, of which five were in complete remission. The negative predictive value of PET1 was reported as 96.8%, while the positive predictive value was 44.4%. Zinzani 2012 also reported that interim PET after two cycles is highly predictive of OS and PFS. In their study, 92% of the interim PET-negative participants (n = 251) were in continuous complete remission as compared to 24.5% of the interim PET-positive participants (n = 53). These conclusions are supported by Ying 2014, although their sample size (n = 35) is too small to provide definite answers. Straus 2011 supported these statements particularly for participants in early stages (as included in their study), as participants with a negative interim PET2 result had a PFS of about 90%, compared to 50% for interim PET-positive participants, at two years. Markova 2012 reported similar findings for interim PET4, which had a high NPV of 98%. Out of 68 participants in total, 50 had a negative interim PET, but 59 a negative end PET. In other words, nine interim PET-positive participants were end PET-negative after chemotherapy. The other nine participants who were interim PET-positive were also end PET-positive. At both timings (PET4 and PET6/8) the authors found a significant difference in the survival between PET-positive and PET-negative participants. The high NPV of interim PET supports early de-escalation of chemotherapy, or omitting radiotherapy, in order to reduce the risk of toxicity and adverse events related to the harsh treatment.

Treatment adaptation based on interim PET

Although not an aim of our review, we considered it important to discuss some results from recently published randomised controlled trials (RCTs) in which the interim PET scan result was used to adapt the therapy for individuals with HL in order to improve outcomes (Andre 2017; Casasnovas 2019; Johnson 2016; Kobe 2018), based on the premise that interim PET scan results are indeed prognostic. For example, in the trial by Johnson 2016, the primary aim was to test the omission of bleomycin due to its toxic effects. All participants (N = 1214, advanced stages) started with ABVD chemotherapy. After interim PET2, PET-positive (DS4-5) participants (N = 182) were assigned to BEACOPP, and PET-negative (DS1-3) participants (N = 935) were randomised to receive either ABVD or AVD. Results show that three-year PFS was slightly better in the ABVD group compared to the AVD group (85.7% versus 84.4%, respectively). Regarding three-year OS, the ABVD group reached 97.2% compared to 97.6% in the AVD group. Hence, there were no significant subgroup differences. However, grade 3 and 4 AEs due to the chemotherapy were more common in the ABVD group. In the PET-positive group, which was escalated to BEACOPP chemotherapy, 3-year PFS was 67.5% and 3-year OS was 87.8%.

In another example by Casasnovas 2019, 823 advanced-stage HL participants were randomly assigned to standard treatment group or PET-driven treatment group. All participants received two cycles of $BEACOPP_{escalated}$ as the initial therapy and interim PET was conducted thereafter. PET-positive participants in both groups, as well as PET-negative participants in the standard group continued with the initial therapy after PET2. PET-negative participants in the experimental arm, however, were switched to two cycles of ABVD. Results of five-year PFS show a similar survival of PET-negative participants in the standard group and experimental group: 88.4% and 89.4%, respectively.

Several systematic reviews were also published that investigated treatment adaptation based on interim PET scan results. Amitai 2018 included 13 studies (of which four were RCTs) that investigated interim PET-adapted treatment in advanced-staged HL. Their findings support the statement that PET-adapted treatment is an appropriate strategy and that it should be considered as standard care for advanced HL (Amitai 2018). This finding is supported by a Phase II RCT (Carras 2018), which assessed interim PET-response adapted treatment strategy in advanced-stage HL. The authors concluded that early salvage therapy and high-dose chemotherapy or autologous stem cell transplant (ASCT) for PET2-positive participants is safe and can lead to similar positive outcomes as in PET2-negative participants (Carras 2018). To compare, Sickinger 2015 included studies in which the treatment was also modified, but concluded that PFS was shorter in individuals with early-stage HL and a negative PET

scan receiving chemotherapy only (PET-adapted therapy) than in those receiving additional RT (standard therapy). This finding was confirmed in another review by Blank 2017, showing improved PFS in early-stage participants receiving radiotherapy in addition to chemotherapy. However, the overall methodological quality of the included studies in both reviews was judged as moderate (for PFS) to very low (for OS). Constrasting evidence on the clinical and prognostic value of interim PET-adapted treatment was also found in non-systematic reviews, which particularly acknowledge the heterogeneity between available studies that makes it difficult to give definite conclusions (Adams 2016a; Berriolo-Riedinger 2018).

AUTHORS' CONCLUSIONS

Implications for practice

This review provides moderate-certainty evidence that interim positron emission tomography (PET) scan results predict overall survival (OS), and very low-certainty evidence that interim PET scan results predict progression-free survival (PFS) in individuals with Hodgkin lymphoma (HL) (evidence of the pooled, unadjusted results). The evidence on the ability of interim PET scan results to distinguish between individuals with a poor prognosis and individuals with a good prognosis can aid decision-making for clinicians and diagnosed individuals, and the evidence may be used in international treatment guidelines for individuals with HL.

Implications for research

Multivariable analyses and prognostic models

Thus far, the prognostic value of interim PET has mostly been assessed in univariable analyses, in which its prognostic ability of determining survival outcomes in individuals with HL has been shown. However, using one single factor is usually not sufficient to give a satisfactory prediction of an outcome, and clinicians, therefore, usually additional factors to give an accurate prediction of an individual's disease progression and health outcome (Moons 2009). Hence, it is important to assess the independent prognostic value of the prognostic factor of interest (in this case interim PET) against established prognostic factors such as disease stage, age, sex, B symptoms or other relevant clinical and individual factors in multivariable analyses as well (Moons 2009; Riley 2019). In such analyses, the independent prognostic ability of a factor, as well as its incremental value on top of other prognostic factors, can be assessed (Moons 2009). In a next step, prognostic models can be built that include multiple prognostic factors that have been proven to be predictive of outcome. Such models are built for risk adaptation and treatment stratification for participants who present those specific factors included in a prediction model for a specific disease, and thereby enables more individualised disease monitoring and treatment guidance. Using a combination of factors, rather than one factor only, allows for a more individual and accurate estimate of the risk of a patient to experience a certain health event (or outcome) within a specific period of time (Moons 2009; Steyerberg 2013).

With regard to our index prognostic factor, we could pool adjusted results in meta-analyses in an update of this review if new studies would adjust for the same set of prognostic factors (covariates). There is a number of different established clinical and individual prognostic factors that can be used to predict survival outcomes in individuals with HL (Cuccaro 2014; Josting 2010; Kılıçkap 2013). In order to enable pooling of adjusted results, future authors of systematic reviews of prognostic factor studies could define a core set of covariates a priori (Riley 2019).

Study design

There is some evidence from retrospective studies that interim PET scan results can predict outcome in individuals during chemotherapy. However, it is commonly agreed that the true prognostic value of this factor can best be assessed in randomised controlled trials (RCTs), in which participants are randomly assigned to a standard or an experimental arm. In the standard arm, participants continue with the planned therapy regimen independent of the interim PET scan result. In the experimental arm, however, different treatments are given according to the interim PET scan result, e.g. de-escalation of treatment in interim PET-negative participants. Hence, RCTs are the most suitable study design, with results from experimental arms in which participants receive therapy adaptation based on the interim PET scan result providing the most robust evidence on whether outcome can be approved, while treatment can be safer, by this strategy. Although assessing therapy modification was not an aim of our review, we judged it important to present and discuss some results of published trials that evaluated the impact of PET-adapted treatment on survival outcomes.

ACKNOWLEDGEMENTS

This review was published in collaboration with the Cochrane Fast-Track Service. We particularly thank Helen Wakeford (Managing Editor) and Heather Maxwell (copy-editor) for their support. The Fast-Track Service team made comments on the review and managed the editorial process.

We thank Nicola Köhler of the Cochrane Haematological Malignancies (CHM) Editorial Base as well as the editors Dr. Julia Bohlius and PD Dr. Sebastian Theurich and the consumer editor Anne Lyddiatt for commenting on the protocol of this review.

We thank Ina Monsef of CHM for developing the search strategy and conducting the search for this review, and Marius Goldkuhle and Vanessa Piechotta for commenting on the first submitted draft of this review.

We thank Jill Hayden for her help with the QUIPS tool, and particularly for agreeing to our adaptation of the tool for the purpose of this review.

We thank Yu-Tian Xiao for the translation of the Chinese publication (Ying 2014).

We thank all principal investigators who have replied to all of our questions and provided us with more information or data, thereby helping us to include as many studies as possible in our analysis. Hence, we want to thank Prof. Dr. Jan Maciej Zaucha, Prof. Dr. Andrea Gallamini, Prof. Dr. Peter Borchmann and his colleague Helen Görgen, Prof. Dr. Pier Luigi Zinzani, Dr. Martin Hutchings, Prof. Dr. Marc André and his colleague Dr. John Raemaekers.

We thank Kym Snell and Sarah Hodgkinson for commenting on the first submitted draft of this review, as well as Robin Featherstone for commenting on the search strategy.

We thank all external peer-reviewers who read and commented on this review. We thank Robert Wolff, Bob Phillips, Michel Meignan,

Tim Illidge, Julian Higgins and Ulrike Holtkamp who greatly helped to improve this review.

The research was supported by NHS Blood and Transplant and the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

REFERENCES

References to studies included in this review

Andre 2017 *{published data only}*

Adams HJ, Kwee TC. Interim fluorodeoxyglucose positron emission tomography-adapted therapy is not an efficient approach to improving outcome in early-stage Hodgkin lymphoma. *Journal of Clinical Oncology* 2017;**35**(24):2850-1.

* Andre ME, Girinsky T, Federico M, Reman O, Fortpied C, Gotti M, et al. Early positron emission tomography response-adapted treatment in stage I and II Hodgkin lymphoma: final results of the randomized EORTC/LYSA/FIL H10 trial (NCT00433433). *Journal of Clinical Oncology* 2017;**35**:1786-94.

Andre ME, Reman O, Federico M, Brice P, Brusamolino E, Girinski T, et al. First report on the H10 EORTC/GELA/IIL randomized intergroup trial on early FDG-PET scan guided treatment adaptation versus standard combined modality treatment in patients with supra-diaphragmatic stage I/II Hodgkin's lymphoma, for the Groupe d'Etude Des Lymphomes De l'Adulte (GELA), European Organisation for the Research and Treatment of Cancer (EORTC) Lymphoma Group and the Intergruppo Italiano Linfomi (IIL). *Blood* 2009;**114**(22):97.

Andre ME, Reman O, Federico M, Girinski T, Brice P, Brusamolino E, et al. Interim analysis of the randomized EORTC/LYSA/FIL intergroup H10 trial on early PET-scan driven treatment adaptation in stage I/II Hodgkin lymphoma. *Blood* 2012;**120**(21):549.

Cottereau AS, Versari A, Loft A, Casanovas O, Bellei M, Ricci R, et al. Prognostic value of baseline metabolic tumor volume in early-stage Hodgkin lymphoma in the standard arm of the H10 trial. *Journal of Nuclear Medicine* 2018;**131**(13):1456-63.

Fornecker LM, Lazarovici J, Aurer I, Casasnovas RO, Gac A C, Bonnet C, et al. PET-based response after 2 cycles of brentuximab vedotin in combination with AVD for first-line treatment of unfavorable early-stage Hodgkin lymphoma: First analysis of the primary endpoint of BREACH, a randomized phase II trial of LYSA-FIL-EORTC intergroup. *Blood* 2017;**130**(Suppl 1):736.

Hindie E, Mesguich C, Zanotti-Fregonara P. On the role of interim fluorine-18-labeled fluorodeoxyglucose positron emission tomography in early-stage favorable Hodgkin lymphoma. *Journal of Clinical Oncology* 2017;**35**(24):2851-2.

Raemaekers JM, Andre MP, Federico M, Girinsky T, Oumedaly R, Brusamolino E, et al. Omitting radiotherapy in early positron emission tomography-negative stage I/II Hodgkin lymphoma is associated with an increased risk of early relapse: Clinical results of the preplanned interim analysis of the randomized EORTC/LYSA/FIL H10 trial. *Journal of Clinical Oncology* 2014;**32**:1188-94.

Annunziata 2016 *{published data only}*

Annunziata S, Calcagni M, Rufini V, Cuccaro A, Massini G, Bartolomei F, et al. The prognostic role of interim FDG-PET/CT and CD68+ cells count in the treatment of Hodgkin lymphoma. *European Journal of Nuclear Medicine and Molecular Imaging* 2014;**41**(Suppl. 2):S182.

Annunziata S, Calcagni ML, Indovina L, Rufini V. Measurement uncertainty and clinical impact of target-to-background ratios derived by interim FDG-PET/CT in Hodgkin Lymphoma: reply to Laffon and Martan. *European Journal of Nuclear Medicine and Molecular Imaging* 2017;**44**(12):2140-1.

Annunziata S, Cuccaro A, Calcagni M, Indovina L, Hohaus S, Giordano A, et al. Could the ratio between lesion and liver SUVmax (rPET) be a prognostic factor in patients with Hodgkin lymphoma undergoing interim FDG-PET/CT? A retrospective study. In: European Journal of Nuclear Medicine and Molecular Imaging. Vol. 42. 2015:S686.

* Annunziata S, Cuccaro A, Calcagni ML, Hohaus S, Giordano A, Rufini V. Interim FDG-PET/CT in Hodgkin lymphoma: the prognostic role of the ratio between target lesion and liver SUVmax (rPET). *Annals of Nuclear Medicine* 2016;**30**:588-92.

Annunziata S, Cuccaro A, Hohaus S, Calcagni M L, Giordano A, Rufini V. Interim FDG-PET/CT in patients with Hodgkin lymphoma: The prognostic role of the ratio between target lesion and liver SUVmax (rPET). *Journal of Nuclear Medicine* 2016;**57**(Supp 2):651.

Annunziata S, Cuccaro A, Hohaus S, Giordano A, Rufini V. Semi-quantitative parameters could improve positive predictive value of interim FDG-PET/CT in Hodgkin lymphoma. *European Journal of Nuclear Medicine and Molecular Imaging* 2016;**43**(Suppl 1):S312.

Annunziata S, Cuccaro A, Hohaus S, Rufini V. Interim FDG-PET/CT in Hodgkin lymphoma: The prognostic value of ratios between target lesion and background SUV. *Clinical and Translational Imaging* 2017;**5**(Suppl 1):S44-5.

Annunziata S, Cuccaro A, Rufini V, Calcagni M, Giachelia M, Martini M, et al. The prognostic role of interim FDG-PET/CT, CD68 1 cells count and levels of plasma thymus and activation-regulated chemokine in the treatment of Hodgkin lymphoma. *Clinical and Translational Imaging* 2015;**3**(Suppl 1):S18-19.

Cuccaro A, Annunziata S, Rufini V, Calcagni M L, Giordano A, Hohaus S. Can the ratio between lesion and liver SUVmax improve outcome prediction in Hodgkin lymphoma with respect to the 5-point Deauville score? *Haematologica* 2015;**100**(Suppl 3):89.

Barnes 2011 *{published data only}*

* Barnes JA, LaCasce AS, Zukotynski K, Israel D, Feng Y, Neuberg D, et al. End-of-treatment but not interim PET scan predicts outcome in nonbulky limited-stage Hodgkin's lymphoma. *Annals of Oncology* 2011;**22**:910-5.

Sher DJ, Mauch PM, Van Den Abbeele A, LaCasce AS, Czerminski J, Ng AK. Prognostic significance of mid- and post-ABVD PET imaging in Hodgkin's lymphoma: the importance of involved-field radiotherapy. *Annals of Oncology* 2009;**20**:1848-53.

Casasnovas 2019 *{published data only}*

* Casasnovas O, Bouabdallah R, Brice P, Lazarovici J, Ghesquieres H, Stamatoullas A, et al. PET-adapted treatment for newly diagnosed advanced Hodgkin lymphoma (AHL2011): a randomised, multicentre, non-inferiority, phase 3 study. *Lancet Oncology* 2019;**20**(2):202-15.

Casasnovas O, Brice P, Bouabdallah R, Salles G, Stamatoulas A, Dupuis J, et al. Final analysis of the AHL2011 randomized phase III LYSA study comparing an early pet driven treatment de-escalation to a not pet-monitored strategy in patients with advanced stages Hodgkin lymphoma. *HemaSphere* 2018;**2**(Suppl. 2):7.

Casasnovas O, Brice P, Bouabdallah R, Salles GA, Stamatoulas A, Dupuis J, et al. Randomized phase III study comparing an early PET driven treatment de-escalation to a not PET-monitored strategy in patients with advanced stages Hodgkin lymphoma: final analysis of the AHL2011 LYSA study. *Journal of Clinical Oncology* 2018;**36**(15):7503.

Casasnovas O, Brice P, Bouabdallah R, Salles GA, Stamatoullas A, Dupuis J, et al. Randomized phase III study comparing an early pet driven treatment de-escalation to a not pet-monitored strategy in patients with advanced stages hodgkin lymphoma: interim analysis of the AHL2011 LYSA study. *Blood* 2015;**126**(23):577.

Casasnovas O, Kanoun S, Tal I, Cottereau AS, Edeline V, Brice P, et al. Baseline total metabolic volume (TMTV) to predict the outcome of patients with advanced Hodgkin lymphoma (HL) enrolled in the AHL2011 LYSA trial. *Journal of Clinical Oncology* 2016;**34**(15):7509.

Casasnovas O, Meignan M, Reman O, Gaillard I, Stamatoullas A, Brice P, et al. AHL 2011: A LYSA randomized phase III study of a treatment driven by early PET response compared to a standard treatment in patients with Ann Arbor stage III-IV or high-risk IIB Hodgkin lymphoma. *Journal of Clinical Oncology* 2013;**31**(Suppl. 15):8615.

Cerci 2010 *{published data only}*

Cerci JJ, Pracchia LF, Linardi CC, Pitella FA, Delbeke D, Izaki M, et al. 18F-FDG PET after 2 cycles of ABVD predicts event-free survival in early and advanced Hodgkin lymphoma. *Journal of Nuclear Medicine* 2010;**51**:1337-43.

Gallamini 2014 *{published data only}*

Adams HJ, Kwee TC. Prevention of large-scale implementation of unnecessary and expensive predictive tests in Hodgkin's lymphoma. *Lancet Haematology* 2017;**4**(2):e63-4.

Agostinelli C, Gallamini A, Stracqualursi L, Agati P, Tripodo C, Fuligni F, et al. The combined role of biomarkers and interim PET scan in prediction of treatment outcome in classical Hodgkin's lymphoma: a retrospective, European, multicentre cohort study. *Lancet Haematology* 2016;**3**(10):e467-79.

Biggi A, Bergesio F, Bianchi A, Menga M, Chauvie S, Fallanca F, et al. Semi-quantitative scan assessment improves the accuracy of the deauville 5-point scale? *Clinical and Translational Imaging* 2017;**5**(Suppl. 1):S43-4.

Biggi A, Gallamini A, Chauvie S, Hutchings M, Kostakoglu L, Gregianin M, et al. International validation study for interim PET in ABVD-treated, advanced-stage Hodgkin lymphoma: interpretation criteria and concordance rate among reviewers. *Journal of Nuclear Medicine* 2013;**54**:683-90.

Biggi A, Kostakoglu L, Barrington S, Chauvie S, Gregianin M, Meignan M, et al. How the threshold of positive results influence progression free survival (PFS) in advanced Hodgkin's lymphoma treated with ABVD: Experience from the International Validation Study. *European Journal of Nuclear Medicine and Molecular Imaging* 2012;**39**(Suppl. 2):S222.

* Gallamini A, Barrington SF, Biggi A, Chauvie S, Kostakoglu L, Gregianin M, et al. The predictive role of interim positron emission tomography for Hodgkin lymphoma treatment outcome is confirmed using the interpretation criteria of the Deauville five-point scale. *Haematologica* 2014;**99**:1107-13.

Gallamini A, Hutchings M, Rigacci L, Specht L, Merli F, Hansen M, et al. Early interim 2-[18F]fluoro-2-deoxy-D-glucose positron emission tomography is prognostically superior to international prognostic score in advanced-stage Hodgkin's lymphoma: a report from a joint Italian-Danish study. *Journal of Clinical Oncology* 2007;**25**:3746-52.

Gallamini A, Rigacci L, Merli F, Nassi L, Bosi A, Capodanno I, et al. The predictive value of positron emission tomography scanning performed after two courses of standard therapy on treatment outcome in advanced stage Hodgkin's disease. *Haematologica* 2006;**91**:475-81.

Meignan M. High-risk interim PET negative patients in Hodgkin's lymphoma. *Lancet Haematology* 2016;**3**(10):e449-50.

Rigacci L, Puccini B, Zinzani PL, Biggi A, Castagnoli A, Merli F, et al. The prognostic value of positron emission tomography performed after two courses (INTERIM-PET) of standard therapy on treatment outcome in early stage Hodgkin lymphoma: A multicentric study by the fondazione italiana linfomi (FIL). *American Journal of Hematology* 2015;**90**:499-503.

Gandikota 2015 *{published data only}*

Gandikota N, Hartridge-Lambert S, Migliacci JC, Yahalom J, Portlock CS, Schoder H. Very low utility of surveillance imaging in early-stage classic Hodgkin lymphoma treated with a combination of doxorubicin, bleomycin, vinblastine, and dacarbazine and radiation therapy. *Cancer* 2015;**121**:1985-92.

Hutchings 2005 *{published data only}*

Hutchings M, Mikhaeel NG, Fields PA, Nunan T, Timothy AR. Prognostic value of interim FDG-PET after two or three cycles of chemotherapy in Hodgkin lymphoma. *Annals of Oncology* 2005;**16**:1160-8.

Hutchings 2006 *{published data only}*

Hutchings M, Loft A, Hansen M, Pedersen LM, Buhl T, Jurlander J, et al. FDG-PET after two cycles of chemotherapy predicts treatment failure and progression-free survival in Hodgkin lymphoma. *Blood* 2006;**107**:52-9.

Hutchings 2014 *{published data only}*

Hutchings M, Kostakoglu L, Zaucha J M, Malkowski B, Biggi A, Danielewicz I, et al. Early determination of treatment sensitivity in Hodgkin lymphoma: FDG-PET/CT after one cycle of therapy has a higher negative predictive value than after two cycles of chemotherapy. *Annals of Oncology* 2011;**22**(Suppl. 4):138-9.

* Hutchings M, Kostakoglu L, Zaucha JM, Malkowski B, Biggi A, Danielewicz I, et al. In vivo treatment sensitivity testing with positron emission tomography/computed tomography after one cycle of chemotherapy for Hodgkin lymphoma. *Journal of Clinical Oncology* 2014;**32**:2705-11.

Kobe 2018 *{published data only}*

Adams HJ, Kwee TC. Interim FDG-PET does not predict outcome in advanced-stage Hodgkin lymphoma patients treated with BEACOPP. *British Journal of Haematology* 2018;**185**(4):758-60.

Adams HJ, Kwee TC. Interim FDG-PET/CT in Hodgkin lymphoma: what are we actually looking at? *Acta Oncologica* 2018;**57**(8):1128-30.

Borchmann P, Eichenauer DA, Pluetschow A, Haverkamp H, Kreissl S, Fuchs M, et al. Targeted BEACOPP variants in patients with newly diagnosed advanced stage classical Hodgkin lymphoma: Final analysis of a randomized phase II study. *Blood* 2015;**126**(23):580.

Borchmann P, Engert A. Reply to H.J.A. Adams et al, E.A. Hawkes et al, and C.F. Hess et al. *Journal of Clinical Oncology* 2017;**35**(3):375-6.

Borchmann P, Goergen H, Kobe C, Eichenauer D, Greil R, Lohri A, et al. EBEACOPP with or without rituximab in interim-PET-positive advanced-stage Hodgkin lymphoma: updated results of the international, randomized phase 3 GHSG HD18 trial. *Hematological Oncology* 2017;**35**(Suppl. 2):65.

Borchmann P, Goergen H, Kobe C, Fuchs M, Greil R, Zijlstra JM, et al. Treatment reduction in patients with advanced-stage Hodgkin lymphoma and negative interim pet: final results of the international, randomized phase 3 trial HD18 by the German Hodgkin study group. *Haematologica* 2017;**102**(S150):24-5.

Borchmann P, Goergen H, Kobe C, Lohri A, Greil R, Eichenauer DA, et al. Early interim PET in patients with advanced-stage Hodgkin's lymphoma treated within the phase 3 GHSG HD18 study. *Blood* 2017;**130**(Suppl. 1):653.

Borchmann P, Goergen H, Kobe C, Lohri A, Greil R, Eichenauer DA, et al. PET-guided treatment in patients with advanced-stage Hodgkin's lymphoma (HD18): final results of an open-label, international, randomised phase 3 trial by the German Hodgkin Study Group. *Lancet* 2018;**390**(10114):2790-802.

Borchmann P, Haverkamp H, Lohri A, Kreissl S, Greil R, Markova J, et al. Addition of rituximab to BEACOPPescalated to improve the outcome of early interim PET positive advanced stage hodgkin lymphoma patients: Second planned interim analysis of the HD18 study. *Blood* 2014;**124**(21):500.

Borchmann P, Haverkamp H, Lohri A, Mey U, Kreissl S, Greil R, et al. Progression-free survival of early interim PET-positive patients with advanced stage Hodgkin's lymphoma treated with BEACOPPescalated alone or in combination with rituximab (HD18): an open-label, international, randomised phase 3 study by the German Hodgkin Study Group. *Lancet Oncology* 2017;**18**(4):454-63.

Crump M. Predicting outcomes after a positive interim FDG-PET scan in advanced Hodgkin's lymphoma. *Lancet Oncology* 2017;**18**(4):416-8.

* Kobe C, Goergen H, Baues C, Kuhnert G, Voltin CA, Zijlstra J, et al. Outcome-based interpretation of early interim PET in advanced-stage Hodgkin lymphoma. *Blood* 2018;**132**(21):2273-9.

Kobe C, Goergen H, Fuchs M, Eich HT, Baues C, Diehl V, et al. Treatment reduction in patients with advanced-stage Hodgkin lymphoma and negative interim FDG-PET: final results of the international, randomized, phase 3 HD18 trial by the German Hodgkin Study Group. *European Journal of Nuclear Medicine and Molecular Imaging* 2017;**44**(Suppl. 1):S312.

Kreissl S, Goergen H, Kobe C, Fuchs M, Greil R, Huttmann A, et al. Treatment reduction in patients with advanced-stage Hodgkin-lymphoma and negative interim PET: final results of the international randomized phase 3 trial HD18 by the German Hodgkin Study Goup. *Oncology Research and Treatment* 2017;**40**(Suppl. 3):201.

Markova 2012 *{published data only}*

* Markova J, Kahraman D, Kobe C, Skopalova M, Mocikova H, Klaskova K, et al. Role of [18F]-fluoro-2-deoxy-D-glucose positron emission tomography in early and late therapy assessment of patients with advanced Hodgkin lymphoma treated with bleomycin, etoposide, adriamycin, cyclophosphamide, vincristine, procarbazine and prednisone. *Leukemia & Lymphoma* 2012;**53**:64-70.

Markova J, Kobe C, Skopalova M, Dedeckova K, Mocikova H, Klaskova K, et al. Response assessment after 4 cycles of BEACOPP using FDG-PET in patients with advanced-stage Hodgkin lymphoma. *Annals of Oncology* 2011;**22**(Suppl. 4):160.

Markova J, Kobe C, Skopalova M, Klaskova K, Dedeckova K, Plutschow A, et al. FDG-PET for assessment of early treatment response after four cycles of chemotherapy in patients with advanced-stage Hodgkin's lymphoma has a high negative predictive value. *Annals of Oncology* 2009;**20**:1270-4.

Markova J, Kobe C, Skopalova M, Zikavska L, Vernerova Z, Klaskova K, et al. Early and late response assessment with FDG-PET after BEACOPP-based chemotherapy in advanced-stage Hodgkin lymphoma patients has a high negative predictive value. In: Haematologica. Vol. 94. 2009:33.

Mesguich 2016 *{published data only}*

Adams HJ, Kwee TC. Hodgkin lymphoma: is there really a need for interim and end-of-treatment FDG-PET evaluations? *British Journal of Haematology* 2018;**181**(1):122-3.

Mesguich C, Cazeau AL, Bouabdallah K, Hindie E. Hodgkin lymphoma: is there really a need for interim and end-of-treatment FDG-PET evaluations? - Response to Adams & Kwee. *British Journal of Haematology* 2018;**181**(1):124-5.

* Mesguich C, Cazeau AL, Bouabdallah K, Soubeyran P, Guyot M, Milpied N, et al. Hodgkin lymphoma: a negative interim-PET cannot circumvent the need for end-of-treatment-PET evaluation. *British Journal of Haematology* 2016;**175**:652-60.

Mesguich C, Cazeau AL, Bouabdallah K, Soubeyran P, Guyot M, Milpied N, et al. Hodgkin's lymphoma: Interim FDG-PET result with a score <= 2 on the 5-point scale may obviate the need for end-of-treatment FDG-PET evaluation. *Journal of Nuclear Medicine* 2015;**56**(Suppl. 3):595.

Oki 2014 *{published data only}*

Oki Y, Chuang H, Chasen B, Jessop A, Pan T, Fanale M, et al. The prognostic value of interim positron emission tomography scan in patients with classical Hodgkin lymphoma. *British Journal of Haematology* 2014;**165**:112-6.

Okosun 2012 *{published data only}*

Okosun J, Shaw K, Montoto S, Marcus R, Fields P, Virchis A, et al. Interim FDG-PET scanning in patients with Hodgkin lymphoma and HIV predict response to ABVD chemotherapy. *Haematologica* 2011;**96**(Suppl. 2):322-3.

* Okosun J, Warbey V, Shaw K, Montoto S, Fields P, Marcus R, et al. Interim fluoro-2-deoxy-D-glucose-PET predicts response and progression-free survival in patients with Hodgkin lymphoma and HIV infection. *AIDS* 2012;**26**:861-5.

Orlacchio 2012 *{published data only}*

Orlacchio A, Schillaci O, Gaspari E, Della Gatta F, Danieli R, Bolacchi F, et al. Role of [18F]-FDG-PET/MDCT in evaluating early response in patients with Hodgkin's lymphoma. *La Radiologia Medica* 2012;**117**:1250-63.

Rossi 2014 *{published data only}*

Rossi C, Kanoun S, Berriolo-Riedinger A, Dygai-Cochet I, Humbert O, Legouge C, et al. Interim 18F-FDG PET SUVmax reduction is superior to visual analysis in predicting outcome early in Hodgkin lymphoma patients. *Journal of Nuclear Medicine* 2014;**55**:569-73.

Simon 2016 *{published data only}*

Barna S, Miltenyi Z, Simon Z, Jona A, Magyari F, Nagy Z, et al. Prognostical value of Interim and restaging PET/CT on Hodgkin lymphoma with CHEAP (Chemotherapy Effectiveness Assessed by PET/CT) long term observation. *European Journal of Nuclear Medicine and Molecular Imaging* 2014;**41**(Suppl. 2):S511.

Illes A, Magyari F, Barna S, Simon Z, Payer E, Garai I, et al. Experiences of the first two years with interim PET/CT in Hodgkin lymphoma - the Hungarian CHEAP study. *Haematologica* 2010;**95**(Suppl. 4):S45.

Miltenyi Z, Barna S, Garai I, Simon Z, Jona A, Magyari F, et al. Prognostic value of interim and restaging PET/CT in Hodgkin lymphoma. Results of the CHEAP (Chemotherapy Effectiveness Assessment by PET/CT) study - long term observation. *Neoplasma* 2015;**62**:627-34.

Miltenyi Z, Simon Z, Jona A, Magyari F, Barna S, Garai I, et al. Interim PET/CT in Hodgkin lymphoma - final results of the Hungarian CHEAP study (2007-2011). *Haematologica* 2013;**98**(Suppl. 2):42.

* Simon Z, Barna S, Miltenyi Z, Husi K, Magyari F, Jona A, et al. Combined prognostic value of absolute lymphocyte/monocyte ratio in peripheral blood and interim PET/CT results in Hodgkin lymphoma. *International Journal of Hematology* 2016;**103**(1):63-9.

Straus 2011 *{published data only}*

Kostakoglu L, Gandikota N, Hutchings M, Cotter R, Lamonica D, Nanni C, et al. Deauville criteria and post One-Cycle SUVmax decrease seem to predict progression free survival (PFS) better than metabolic tumor measurements in classical Hodgkin lymphoma (cHL). *European Journal of Nuclear Medicine and Molecular Imaging* 2012;**39**(Suppl. 2):S222.

Kostakoglu L, Schoder H, Hall N, Straus DJ, Johnson JL, Schwartz L, et al. Interim FDG PET imaging in CALGB 50203 trial of stage I/II non-bulky Hodgkin lymphoma: Would using combined PET and CT criteria better predict response than each test alone? *Blood* 2011;**118**(21):3644.

Kostakoglu L, Schoder H, Johnson JL, Hall NC, Schwartz LH, Straus DJ, et al. Interim [(18)F]fluorodeoxyglucose positron emission tomography imaging in stage I-II non-bulky Hodgkin lymphoma: would using combined positron emission tomography and computed tomography criteria better predict response than each test alone? *Leukemia & Lymphoma* 2012;**53**:2143-50.

* Straus DJ, Johnson JL, LaCasce AS, Bartlett NL, Kostakoglu L, Hsi ED, et al. Doxorubicin, vinblastine, and gemcitabine (CALGB 50203) for stage I/II nonbulky Hodgkin lymphoma: pretreatment prognostic factors and interim PET. *Blood* 2011;**117**:5314-20.

Touati 2014 *{published data only}*

Hutchings M, Kamper P. New clues to the prognostic challenge of Hodgkin Lymphoma. *Leukemia & Lymphoma* 2015;**56**(2):277-8.

* Touati M, Delage-Corre M, Monteil J, Abraham J, Moreau S, Remenieras L, et al. CD68-positive tumor-associated macrophages predict unfavorable treatment outcomes in classical Hodgkin lymphoma in correlation with interim fluorodeoxyglucose-positron emission tomography assessment. *Leukemia and Lymphoma* 2015;**56**:332-41.

Ying 2014 *{published data only}*

Ying Z, Wang X, Song Y, Zheng W, Wang X, Xie Y, et al. Prognostic value of 18F-FDG PET-CT in Hodgkin lymphoma. *Chinese Journal Of Hematology* 2014;**35**:325-7.

Zaucha 2017 *{published data only}*

Adams HJ, Kwee TC. The predictive value of interim FDG-PET in early-stage Hodgkin lymphoma is not well established. *Annals of Oncology* 2018;**29**(2):510-2.

Zaucha JM, Chauvie S, Malkowski B, Warszewska A, Biggi A, Kobylecka M, et al. The prognostic role of interim PET after first chemotherapy cycle in ABVD-treated Hodgkin lymphoma (HL) patients - Polish lymphoma research group (PLRG) observational study. *Haematologica* 2013;**98**(Suppl. 2):38.

* Zaucha JM, Malkowski B, Chauvie S, Subocz E, Tajer J, Kulikowski W, et al. The predictive role of interim PET after the first chemotherapy cycle and sequential evaluation of response to ABVD in Hodgkin lymphoma patients - the Polish Lymphoma Research Group (PLRG) Observational Study. *Annals of Oncology* 2017;**28**(12):3051-7.

Zaucha JM, Malkowski B, Subocz E, Chauvie S, Tajer J, Kulikowski W, et al. The prognostic role of interim PET after first chemotherapy cycle and PET sequential evaluation of response to ABVD in <hodgkin lymphoma patients - The Polish Lymphoma Research Group (PLRG) observational study. *Blood* 2015;**126**(23):3943.

Zinzani 2012 *{published data only}*

Adams HJ, Kwee TC. Does interim 18F-FDG-PET response-adapted therapy really benefit advanced-stage Hodgkin lymphoma patients? *Nuclear Medicine Communications* 2016;**37**(12):1333-4.

Puccini B, Rigacci L, Zinzani PL, Broccoli A, Gallamini A, Merli F, et al. Early positive FDG-PET scan do not confirm its prognostic impact in localized bulky disease Hodgkin lymphoma patients. *Haematologica* 2011;**96**(Suppl. 3):26-7.

Rigacci L, Puccini B, Gallamini A, Merli F, Stelitano C, Balzarotti M, et al. Early FDG-PET scan confirms its prognostic impact also in localized stage, ABVD treated Hodgkin lymphoma patients. *Haematologica* 2009;**94**(Suppl. 2):34.

Rigacci L, Puccini B, Zinzani PL, Kovalchuk S, Broccoli A, Evangelista A, et al. Clinical characteristics of patients with negative interim-PET and positive final PET: Data from the prospective PET-oriented HD0801 study by Fondazione Italiana linfomi (FIL). *Hematological Oncology* 2017;**35**(Suppl. 2):38.

Rigacci L, Zinzani PL, Puccini B, Broccoli A, Gallamini A, Merli F, et al. Early FDG-PET scan confirms its prognostic impact also in localized stage, ABVD treated Hodgkin lymphoma patients. *Haematologica* 2010;**95**(Suppl. 2):474.

Rigacci L, Zinzani PL, Puccini B, Broccoli A, Gallamini A, Merli F, et al. Early FDG-PET scan confirms its prognostic impact also in localized stage, ABVD treated Hodgkin lymphoma patients. In: Haematologica. Vol. 96. 2010:S13.

Rigacci L, Zinzani PL, Puccini B, Broccoli A, Gallamini A, Merli F, et al. Early positive FDG-PET scan do not confirm its prognostic impact in bulky disease Hodgkin lymphoma patients. *Haematologica* 2011;**96**(Suppl. 2):320-1.

Rigacci L, Zinzani PL, Puccini B, Broccoli A, Gallamini A, Merli F, et al. IHP interpretation criteria of interim-PET scan confirms prognostic impact in early stage Hodgkin lymphoma patients without bulky disease. *Blood* 2010;**116**(21):3890.

Stefoni V, Broccoli A, Alinari L, Ambrosini V, Derenzini E, Fanti S, et al. Predictive role of early interim FDG-PET in Hodgkin lymphoma. *Blood* 2009;**114**(22):1659.

Stefoni V, Pellegrini C, Rigacci L, Broccoli A, Puccini B, Fanti S, et al. Interim PET 2: Is it a real prognostic factor in selecting two different subsets of Hodgkin disease patients? *Haematologica* 2011;**96**(Suppl. 3):123.

* Zinzani PL, Rigacci L, Stefoni V, Broccoli A, Puccini B, Castagnoli A, et al. Early interim 18F-FDG PET in Hodgkin's lymphoma: evaluation on 304 patients. *European Journal of Nuclear Medicine and Molecular Imaging* 2012;**39**:4-12.

Zinzani PL, Tani M, Fanti S, Alinari L, Musuraca G, Marchi E, et al. Early positron emission tomography (PET) restaging: a predictive final response in Hodgkin's disease patients. *Annals of Oncology* 2006;**17**:1296-300.

References to studies excluded from this review

Adams 2016 *{published data only}*

Adams HJ, Kwee TC. RAPID trial demonstrates low positive predictive value of interim FDG-PET in early-stage Hodgkin lymphoma after three cycles of ABVD. *Journal of Pediatric Hematology/Oncology* 2016;**38**(2):165.

Adams 2017 *{published data only}*

Adams HJ, Kwee TC. Predictive value of interim [18F]fluorodeoxyglucose-positron emission tomography in advanced-stage Hodgkin lymphoma is not well established. Journal of Clinical Oncology 2017;**35**(3):370-1.

Adams 2018 *{published data only}*

Adams HJ, Kwee TC. Interim FDG-PET has no value in selecting patients who require treatment modification in both early- and advanced-stage Hodgkin lymphoma. British Journal of Haematology 2018;**183**(1):129-31.

Adams 2018a *{published data only}*

Adams HJ, Kwee TC. No evidence to promote interim FDG-PET adapted therapy in the NCCN guidelines for Hodgkin lymphoma. Journal of the National Comprehensive Cancer Network 2018;**16**(3):226-7.

Adams 2018b *{published data only}*

Adams HJ, Kwee TC. Strikingly heterogeneous results among studies on interim fluorodeoxyglucose-positron emission tomography-adapted treatment in advanced-stage Hodgkin lymphoma. Journal of Clinical Oncology 2018;**36**(20):2123-4.

Adams 2019 *{published data only}*

Adams HJ, Kwee TC. Post-ABVD biopsy results, and not post-ABVD FDG-PET results, predict outcome in early-stage Hodgkin lymphoma. British Journal of Haematology 2019;**184**(2):290-2.

Advani 2007 *{published data only}*

Advani R, Maeda L, Lavori P, Quon A, Hoppe R, Breslin S, et al. Impact of positive positron emission tomography on prediction of freedom from progression after Stanford V chemotherapy in Hodgkin's disease. *Journal of Clinical Oncology* 2007;**25**:3902-7.

Afanasyev 2017 *{published data only}*

Afanasyev BV, Moiseev IS, Alekseev SM, Mikhailova NB, Kondakova EV, Ilyin NV, et al. Multicenter prospective escalation-deescalation PET-guided clinical study in classical type Hodgkin disease in the north-west of Russian federation (RNWOHG-HD1): Rationale and design. Cellular Therapy and Transplantation 2017;**6**(4):76-81.

Albano 2017 *{published data only}*

Albano D, Patti C, Lagalla R, Midiri M, Galia M. Whole-body MRI, FDG-PET/CT, and bone marrow biopsy, for the assessment of bone marrow involvement in patients with newly diagnosed lymphoma. *Journal of Magnetic Resonance Imaging* 2017;**45**(4):1082-9.

Albano 2018 *{published data only}*

Albano D, Patti C, Matranga D, Lagalla R, Midiri M, Galia M. Whole-body diffusion-weighted MR and FDG-PET/CT in Hodgkin Lymphoma: Predictive role before treatment and early assessment after two courses of ABVD. *European Journal of Radiology* 2018;**103**:90-8.

Altamirano 2008 *{published data only}*

Altamirano J, Esparza JR, de la Garza Salazar J, Calvo PS, Vera SR, Chalapud Revelo JR, et al. Staging, response to therapy, and restaging of lymphomas with18F-FDG PET. *Archives of Medical Research* 2008;**39**(1):69-77.

Ansell 2016 *{published data only}*

Ansell SM. Hodgkin lymphoma: 2016 update on diagnosis, risk-stratification, and management. *American Journal of Hematology* 2016;**91**(4):434-42.

Awan 2013 *{published data only}*

Awan UE, Siddiqui N, SaadUllah M, Bashir H, Farooqui ZS, Muzaffar N, et al. FDG-PET scan in assessing lymphomas and the application of Deauville Criteria. *Journal of the Pakistan Medical Association* 2013;**63**(6):725-30.

Barrington 2011a *{published data only}*

Barrington SF, O'Doherty MJ, Pike L, Franceschetto A, Cucca M, Brun E, et al. Standardised PET-CT reporting for an international multicentre trial in lymphoma (RATHL). *European Journal of Nuclear Medicine and Molecular Imaging* 2011;**38 (SUPPL. 2)**:S115.

Barrington 2017 *{published data only}*

Barrington SF, Johnson PW. <18F-FDG PET/CT in lymphoma: Has imaging-directed personalized medicine become a reality? *Journal of Nuclear Medicine* 2017;**58**(10):1539-44.

Bar-Shalom 2003 *{published data only}*

Bar-Shalom R, Yefremov N, Haim N, Dann EJ, Epelbaum R, Keidar Z, et al. Camera-based FDG pet and 67Ga SPECT in evaluation of lymphoma: Comparative study. *Radiology* 2003;**227**(2):353-60.

Basu 2009 *{published data only}*

Basu S. Early FDG-PET response-adapted risk stratification and further therapeutic decision-making in lymphoma: Will this replace the established prognostic indices and be the standard-of-care in clinical management? *European Journal of Nuclear Medicine and Molecular Imaging* 2009;**36**(12):2089-90.

Becherer 2002 *{published data only}*

Becherer A, Mitterbauer M, Jaeger U, Kalhs P, Greinix HT, Karanikas G, et al. Positron emission tomography with 2-fluoro-D-2-deoxyglucose (FDG-PET) predicts relapse of malignant lymphoma after high-dose therapy with stem cell transplantation. *Leukemia* 2002;**16**(2):260-7.

Bednaruk-Mlynski 2015 *{published data only}*

Bednaruk-Mlynski E, Pienkowska J, Skorzak A, Malkowski B, Kulikowski W, Subocz E, et al. Comparison of positron emission tomography/computed tomography with classical contrast-enhanced computed tomography in the initial staging of Hodgkin lymphoma. *Leukemia and Lymphoma* 2015;**56**(2):377-82.

Biggi 2012 *{published data only}*

Biggi A, Kostakoglu L, Barrington S, Chauvie S, Gregianin M, Meignan M, et al. How the threshold of positive results influence progression free survival (PFS) in advanced Hodgkin's lymphoma treated with ABVD: Experience from the International Validation Study. *European Journal of Nuclear Medicine and Molecular Imaging* 2012;**39**(2):S222.

Biggi 2017 *{published data only}*

Biggi A, Bergesio F, Menga M, Bianchi A, Fallanca F, Chauvie S, et al. Diagnostic accuracy of FDG PET/CT at the of treatment of Hodgkin lymphoma in the HD0607 trial. *Clinical and Translational Imaging* 2017;**5**:S44.

Bishop 2015 *{published data only}*

Bishop G. PET-directed therapy for Hodgkin's lymphoma. *New England Journal of Medicine* 2015;**373**(4):392.

Bjurberg 2006 *{published data only}*

Bjurberg M, Gustavsson A, Ohlsson T, Brun E. FDG-PET in the detection of residual disease and relapse in patients with Hodgkin's lymphoma. Experience from a Swedish centre. *Acta Oncologica* 2006;**45**(6):743-9.

Blum 2002 *{published data only}*

Blum R, Prince HM, Hicks RJ, Patrikeos A, Seymour J. Discordant response to chemotherapy detected by PET scanning: Unveiling of a second primary cancer. *American Journal of Clinical Oncology: Cancer Clinical Trials* 2002;**25**(4):368-70.

Bodet-Milin 2008 *{published data only}*

Bodet-Milin C, Kraeber-Bodere F, Moreau P, Campion L, Dupas B, Le Gouill S. Investigation of FDG-PET/CT imaging to guide biopsies in the detection of histological transformation of indolent lymphoma. *Haematologica* 2008;**93**(3):471-2.

Bodet-Milin 2009 *{published data only}*

Bodet-Milin C, Salaun PY, Crespin C, Vuillez JP, Kraeber-Bodere F. FDG-PET scanning in managing patients with lymphoma. *Medecine Nucleaire* 2009;**33**(8):486-90.

Boisson 2007 *{published data only}*

Boisson N, Cachin F, Kelly A, De Freitas D, Isnardi V, Mestas D, et al. PET-CT in the Hodgkin's disease. *Medecine Nucleaire* 2007;**31**(10):562-7.

Borchmann 2016 *{published data only}*

Borchmann S, von Tresckow B, Engert A. Current developments in the treatment of early-stage classical Hodgkin lymphoma. *Current Opinion in Oncology* 2016;**28**(5):377-83.

Bucerius 2006 *{published data only}*

Bucerius J, Herkel C, Joe AY, Altehoefer C, Finke J, Moser E, et al. ¹⁸F-FDG PET and conventional imaging for assessment of Hodgkin's disease and non Hodgkin's lymphoma: An analysis of 193 patient studies. *Nuklearmedizin* 2006;**45**(3):105-10.

Carras 2018 *{published data only}*

Carras S, Dubois B, Senecal D, Jais JP, Peoc'h M, Quittet P, et al. Interim PET response-adapted strategy in untreated advanced stage Hodgkin lymphoma: Results of GOELAMS LH 2007 phase 2 multicentric trial. *Clinical Lymphoma, Myeloma & Leukemia* 2018;**18**(3):191-8.

Ciammella 2016 *{published data only}*

Ciammella P, Filippi AR, Simontacchi G, Buglione M, Botto B, Mangoni M, et al. Post-ABVD/pre-radiotherapy 18F-FDG-PET provides additional prognostic information for early-stage Hodgkin lymphoma: A retrospective analysis on 165 patients. *British Journal of Radiology* 2016;**89**(1061):20150983.

Cremerius 1999 *{published data only}*

Cremerius U, Fabry U, Kroll U, Zimny M, Neuerburg J, Osieka R, et al. [Clinical value of FDG PET for therapy monitoring of malignant lymphoma--results of a retrospective study in 72 patients]. *Clinic for Nuclear Medicine (Stuttgart)* 1999;**38**(1):24-30.

Cremerius 2001 *{published data only}*

Cremerius U, Fabry U, Neuerburg J, Zimny M, Bares R, Osieka R, et al. Prognostic significance of positron emission tomography using fluorine-18-fluorodeoxyglucose in patients treated for malignant lymphoma. *Clinic for Nuclear Medicine (Stuttgart)* 2001;**40**(1):23-30.

Cuccaro 2016 *{published data only}*

Cuccaro A, Annunziata S, Cupelli E, Martini M, Calcagni ML, Rufini V, et al. CD68+ cell count, early evaluation with PET and plasma TARC levels predict response in Hodgkin lymphoma. *Cancer Medicine* 2016;**5**(3):398-406.

D'Urso 2018 *{published data only}*

D'Urso D, Stefano A, Romano A, Russo G, Cosentino S, Fallanca F, et al. Analysis of metabolic parameters coming from basal and interim PET in Hodgkin lymphoma. *Current Medical Imaging Reviews* 2018;**14**(4):533-44.

Damlaj 2017 *{published data only}*

Damlaj M, Al-Zahrani M, Syed G, Gmati G, Pasha T, Abuelgasim K, et al. Escalation from ABVD following positive interim functional imaging improves progression free survival but not overall survival in advanced classical Hodgkin lymphoma-a real world analysis. *Blood* 2017;**130**(Suppl. 1):2799.

Damlaj 2019 *{published data only}*

Damlaj M, Al-Zahrani M, Syed G, Gmati G, Alahmari B, Pasha T, et al. Interim functional imaging is an independent predictor of progression-free survival in advanced classical Hodgkin lymphoma - A real-world analysis. *Clinical Lymphoma, Myeloma and Leukemia* 2019;**19**(1):e71-9.

Danilov 2017 *{published data only}*

Danilov AV, Li H, Press OW, Shapira I, Swinnen LJ, Noy A, et al. Feasibility of interim positron emission tomography (PET)-adapted therapy in HIV-positive patients with advanced Hodgkin lymphoma (HL): a sub-analysis of SWOG S0816 Phase 2 trial. *Leukemia & Lymphoma* 2017;**58**(2):461-5.

Dann 2009 *{published data only}*

Dann EJ, Bar-Shalom R, Tamir A, Ben-Shachar M, Avivi I, Zuckerman T, et al. For standard and high-risk patients with Hodgkin lymphoma six cycles of tailored BEACOPP, based on interim scintigraphy, are effective and female fertility is preserved. *Blood* 2009;**114**(22):1552.

Dann 2010 *{published data only}*

Dann EJ, Bar-Shalom R, Tamir A, Epelbaum R, Avivi I, Ben-Shachar M, et al. A functional dynamic scoring model to elucidate the significance of post-induction interim fluorine-18-fluorodeoxyglucose positron emission tomography findings in patients with Hodgkin's lymphoma. *Haematologica* 2010;**95**(7):1198-206.

Dann 2010a *{published data only}*

Dann EJ, Bairey O, Bar-Shalom R, Izak M, Korenberg A, Akria L, et al. Tailored therapy in Hodgkin lymphoma, based on predefined risk factors and early interim PET/CT, can lead to modification and safe reduction in therapy: Results of 134 patients on the Israel National Hodgkin Study. *Blood* 2010;**116**(21):2809.

Dann 2012 *{published data only}*

Dann EJ, Bairey O, Bar-Shalom R, Mashiach T, Izak M, Korenberg A, et al. Early Hodgkin lymphoma therapy, based on predefined risk factors and early interim PET/CT. Israeli H2 protocol: Preliminary report. *Haematologica* 2012;**97**(Suppl. 1):85.

Dann 2013 *{published data only}*

Dann EJ, Bairey O, Bar-Shalom R, Izak M, Korenberg A, Akria L, et al. Tailored therapy in Hodgkin lymphoma, based on predefined risk factors and early interim PET/CT, Israeli H2 protocol: Preliminary report on 317 patients. *Haematologica* 2013;**98**(Suppl. 2):37.

Dann 2016 *{published data only}*

Dann EJ, Bairey O, Bar-Shalom R, Mashiach T, Barzilai E, Korenberg A, et al. Adjustment of therapy for Hodgkin lymphoma based on interim PET is beneficial and radiotherapy may be substituted with chemotherapy in patients with negative interim study: Final results of H2 trial. European Haematology Association Open Access Library 2016;(S107):4-5.

Dann 2017 *{published data only}*

Dann EJ, Bairey O, Bar-Shalom R, Mashiach T, Barzilai E, Kornberg A, et al. Modification of initial therapy in early and advanced Hodgkin lymphoma, based on interim PET/CT is beneficial: a prospective multicentre trial of 355 patients. *British Journal of Haematology* 2017;**178**:709-718.

Dann 2018 *{published data only}*

Dann EJ, Paltiel O. Interim FDG-PET has no value in selecting patients who require treatment modification in both early- and advanced-stage Hodgkin lymphoma: Response to Adams and Kwee. *British Journal of Haematology* 2018;**183**(1):131-3.

deAndres-Galiana 2015 *{published data only}*

deAndres-Galiana EJ, Fernandez-Martinez JL, Luaces O, del Coz JJ, Fernandez R, Solano J, et al. On the prediction of Hodgkin lymphoma treatment response. *Clinical and Translational Oncology* 2015;**17**(8):612-9.

Diehl 2007 *{published data only}*

Diehl V, Kobe C, Haverkamp H, Dietlein M, Engert A. FDG-PET for assessment of residual tissue after completion of chemotherapy in Hodgkin lymphoma report on the 2nd interim analysis of the PET investigation in the trial HD15 of the GHSG. In: Blood. The American Society of Hematology. 2007.

El-Galaly 2012 *{published data only}*

El-Galaly TC, Mylam KJ, Brown P, Specht L, Christiansen I, Munksgaard L, et al. Positron emission tomography/computed tomography surveillance in patients with Hodgkin lymphoma in first remission has a low positive predictive value and high costs. *Haematologica* 2012;**97**(6):931-6.

Evens 2014 *{published data only}*

Evens AM, Kostakoglu L. The role of FDG-PET in defining prognosis of Hodgkin lymphoma for early-stage disease. *Blood* 2014;**124**(23):3356-64.

Fanti 2008 *{published data only}*

Fanti S, Castellucci P, Stefoni V, Nanni C, Tani M, Rubello D, et al. Early relapse in a patient with Hodgkin's disease and negative interim FDG-PET. *Annals of Nuclear Medicine* 2008;**22**(5):429-32.

Filmont 2003 *{published data only}*

Filmont J E, Czernin J, Yap C, Silverman DH, Quon A, Phelps ME, et al. Value of F-18 fluorodeoxyglucose positron emission tomography for predicting the clinical outcome of patients with aggressive lymphoma prior to and after autologous stem-cell transplantation. *Chest* 2003;**124**(2):608-13.

Fornecker 2017 *{published data only}*

Fornecker LM, Lazarovici J, Aurer I, Casasnovas RO, Gac AC, Bonnet C, et al. Pet-based response after 2 cycles of brentuximab vedotin in combination with avd for first-line treatment of unfavorable early-stage Hodgkin Lymphoma: first analysis of the primary endpoint of breach, a randomized phase II trial of LYSA-FIL-EORTC intergroup. *Blood* 2017;**130**(Suppl. 1):736.

Freudenberg 2004 *{published data only}*

Freudenberg LS, Antoch G, Schutt P, Beyer T, Jentzen W, Muller SP, et al. FDG-PET/CT in re-staging of patients with lymphoma. *European Journal of Nuclear Medicine and Molecular Imaging* 2004;**31**(3):325-9.

Friedberg 2002 *{published data only}*

Friedberg JW, Fischman A, Neuberg D, Kim H, Takvorian T, Mauch PM, et al. FDG-PET is superior to gallium scintigraphy in the staging and follow-up of patients with de novo Hodgkin's disease: a prospective, blinded comparison. Leukemia & Lymphoma 2004;**45**(1):85-92.

Friedberg 2004 *{published data only}*

Friedberg JW, Fischman A, Neuberg D, Kim H, Takvorian T, Ng AK, et al. FDG-PET is superior to gallium scintigraphy in staging and more sensitive in the follow-up of patients with de novo Hodgkin lymphoma: a blinded comparison. *Leukemia and Lymphoma* 2004;**45**(1):85-92.

Front 1999 *{published data only}*

Front D, Bar-Shalom R, Mor M, Haim N, Epelbaum R, Frenkel A, et al. Hodgkin disease: prediction of outcome with 67Ga scintigraphy after one cycle of chemotherapy. *Radiology* 1999;**210**(2):487-91.

Fruchart 2006 *{published data only}*

Fruchart C, Reman O, Le Stang N, Musafiri D, Cheze S, Macro M, et al. Prognostic value of early 18 fluorodeoxyglucose positron emission tomography and gallium-67 scintigraphy in aggressive lymphoma: A prospective comparative study. *Leukemia and Lymphoma* 2006;**47**(12):2547-57.

Gallamini 2008 *{published data only}*

Gallamini A, Hutchings M, Avigdor A, Polliack A. Early interim PET scan in Hodgkin lymphoma: Where do we stand? *Leukemia and Lymphoma* 2008;**49**(4):659-62.

Gallamini 2017 *{published data only}*

Gallamini A, Tarella C, Viviani S, Rossi A, Patti C, Mulé A, et al. Early chemotherapy intensification with escalated BEACOPP in advanced-stage Hodgkin lymphoma with a positive interim PET-CT after 2 ABVD cycles: Long-term results of the GITIL/FIL HD 0607 trial. *Journal of Clinical Oncology* 2018;**36**(5):454-462.

Gallamini 2018 *{published data only}*

Gallamini A, Tarella C, Viviani S, Rossi A, Patti C, Mule A, et al. Early chemotherapy intensification with escalated BEACOPP in patients with advanced-stage Hodgkin lymphoma with a positive interim positron emission tomography/computed tomography scan after two ABVD cycles: Long-term results of the GITIL/FIL HD 0607 trial. *Journal of Clinical Oncology* 2018;**36**(5):454-62.

Gallamini 2018a *{published data only}*

Gallamini A, Viviani S, Pavoni C, Rambaldi A. Reply to H.J.A. Adams et al and C. Mesguich et al. *Journal of Clinical Oncology* 2018;**36**(20):2127-8.

Gallowitsch 2008 *{published data only}*

Gallowitsch HJ, Igerc I, Kohlfurst S, Lind P. The incremental value of F-18 FDG PET and PET/CT in malignant lymphoma. *Imaging Decisions MRI* 2008;**12**(4):2-6.

Goldschmidt 2011 *{published data only}*

Goldschmidt N, Or O, Klein M, Savitsky B, Paltiel O. The role of routine imaging procedures in the detection of relapse of patients with Hodgkin lymphoma and aggressive non-Hodgkin lymphoma. *Annals of Hematology* 2011;**90**(2):165-71.

Greil 2018 *{published data only}*

Greil R. Treatment optimisation trial in the first-line treatment of advanced stage Hodgkin lymphoma; comparison of 4-6 cycles of escalated BEACOPP with 4-6 cycles of BrECADD. *Memo - Magazine of European Medical Oncology.* 2018;**11**(1):1-28.

Guidez 2016 *{published data only}*

Guidez S, Delwail V, Brice P. PET-scan in management of Hodgkin's lymphoma. *Hematologie* 2016;**22**(6):389-91.

Hagtvedt 2015 *{published data only}*

Hagtvedt T, Seierstad T, Lund KV, Løndalen AM, Bogsrud TV, Smith HJ, et al. Diffusion-weighted MRI compared to FDG PET/CT for assessment of early treatment response in lymphoma. *Acta Radiologica* 2015;**56**(2):152-8.

Haioun 2005 *{published data only}*

Haioun C, Itti E, Rahmouni A, Brice P, Rain JD, Belhadj K, et al. [18F]fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) in aggressive lymphoma: an early prognostic tool for predicting patient outcome. *Blood* 2005;**106**(4):1376-81.

Hartmann 2012 *{published data only}*

Hartmann S, Agostinelli C, Diener J, Doring C, Fanti S, Zinzani PL, et al. GLUT1 expression patterns in different Hodgkin lymphoma subtypes and progressively transformed germinal centers. *BMC Cancer* 2012;**12**:586.

Hartridge-Lambert 2013 *{published data only}*

Hartridge-Lambert SK, Schoder H, Lim RC, Maragulia JC, Portlock CS. ABVD alone and a PET scan complete remission negates the need for radiologic surveillance in early-stage, nonbulky Hodgkin lymphoma. *Cancer* 2013;**119**(6):1203-9.

Honda 2014 *{published data only}*

Honda A, Nakamura F, Nannya Y, Shintani Y, Fukayama M, Ichikawa M, et al. Pulmonary lymphocyte-rich classical Hodgkin lymphoma with early response to ABVD therapy. *Annals of Hematology* 2014;**93**(6):1073-4.

Hueltenschmidt 2001 *{published data only}*

Hueltenschmidt B, Sautter-Bihl ML, Lang O, Maul FD, Fischer J, Mergenthaler HG, et al. Whole body positron emission tomography in the treatment of Hodgkin disease. *Cancer* 2001;**91**(2):302-10.

Huic 2006 *{published data only}*

Huic D, Mutvar A, Radman I, Grosev D, Labar B, Zuvic M, et al. The value of F-18 FDG triple-head coincidence PET in the posttreatment evaluation of patients with lymphoma. *Clinical Nuclear Medicine* 2006;**31**(5):275-8.

Hutchings 2007 *{published data only}*

Hutchings M, Loft A, Hansen M, Berthelsen AK, Specht L. Clinical impact of FDG-PET/CT in the planning of radiotherapy for early-stage Hodgkin lymphoma. *European Journal of Haematology* 2007;**78**(3):206-12.

Iagaru 2008 *{published data only}*

Iagaru A, Wang Y, Mari C, Quon A, Goris ML, Horning S, et al. (18)F-FDG-PET/CT evaluation of response to treatment in lymphoma: when is the optimal time for the first re-evaluation scan? *Hellenic Society of Nuclear Medicine* 2008;**11**(3):153-6.

Illidge 2015 *{published data only}*

Illidge T. Personalised approach to treating early Hodgkin's lymphoma. *BMJ* 2015;**350**:h2927.

Jerusalem 2003 *{published data only}*

Jerusalem G, Beguin Y, Fassotte MF, Belhocine T, Hustinx R, Rigo P, et al. Early detection of relapse by whole-body positron emission tomography in the follow-up of patients with Hodgkin's disease. *Annals of Oncology* 2003;**14**(1):123-30.

Johnson 2015 *{published data only}*

Johnson PW, Federico M, Fossa A, O'Doherty M, Roberts T, Stevens L, et al. Response-adapted therapy based on interim FDG-PET scans in advanced Hodgkin lymphoma: first analysis of the safety of deescalation and efficacy of escalation in the international RATHL study (CRUK/07/033). *Clinical Advances in Hematology and Oncology* 2015;**13**(8 Suppl. 9):6-7.

Johnson 2016 *{published data only}*

Johnson P, Federico M, Kirkwood A, Fossa A, Berkahn L, Carella A, et al. Adapted treatment guided by interim PET-CT scan in advanced Hodgkin's lymphoma. *New England Journal of Medicine* 2016;**374**(25):2419-29.

Kamran 2016 *{published data only}*

Kamran SC, Jacene HA, Chen YH, Mauch PM, Ng AK. Clinical outcome of patients with early stage favorable Hodgkin lymphoma treated with ABVDX2 cycles followed by PET/CT restaging and 20 Gy of involved-site radiotherapy. *Haematologica* 2016;**101**(Suppl. 5):14.

Kamran 2018 *{published data only}*

Kamran SC, Jacene HA, Chen YH, Mauch PM, Ng AK. Clinical outcome of patients with early stage favorable Hodgkin lymphoma treated with ABVD x two cycles followed by FDG-PET/CT restaging and 20 Gy of involved-site radiotherapy. *Leukemia and Lymphoma* 2018;**59**(6):1384-90.

Kobe 2008 *{published data only}*

Kobe C, Dietlein M, Franklin J, Markova J, Lohri A, Amthauer H, et al. Positron emission tomography has a high negative predictive value for progression or early relapse for patients with residual disease after first-line chemotherapy in advanced-stage Hodgkin lymphoma. *Blood* 2008;**112**(10):3989-94.

Kobe 2014 *{published data only}*

Kobe C, Kuhnert G, Kahraman D, Haverkamp H, Eich HT, Franke M, et al. Assessment of tumor size reduction improves outcome prediction of positron emission tomography/computed tomography after chemotherapy in advanced-stage Hodgkin lymphoma. *Journal of Clinical Oncology* 2014;**32**(17):1776-81.

Kostakoglu 2006 *{published data only}*

Kostakoglu L, Goldsmith SJ, Leonard JP, Christos P, Furman RR, Atasever T, et al. FDG-PET after 1 cycle of therapy predicts outcome in diffuse large cell lymphoma and classic Hodgkin disease. *Cancer* 2006;**107**:2678-87.

Li 2013 *{published data only}*

Li YJ, Li ZM, Xia XY, Huang HQ, Xia ZJ, Lin TY, et al. Prognostic value of interim and posttherapy 18F-FDG PET/CT in patients with mature T-cell and natural killer cell lymphomas. *Journal of Nuclear Medicine* 2013;**54**(4):507-15.

Lowe 2002 *{published data only}*

Lowe VJ, Wiseman GA. Assessment of Lymphoma Therapy Using (18)F-FDG PET. *Journal of Nuclear Medicine* 2002;**43**(8):1028-30.

Milgrom 2017 *{published data only}*

Milgrom SA, Pinnix CC, Chuang H, Oki Y, Akhtari M, Mawlawi O, et al. Early-stage Hodgkin lymphoma outcomes after combined modality therapy according to the post-chemotherapy 5-point score: can residual pet-positive disease be cured with radiotherapy alone? *British Journal of Haematology* 2017;**179**(3):488-96.

Mocikova 2010 *{published data only}*

Mocikova H, Obrtlikova P, Vackova B, Trneny M. Positron emission tomography at the end of first-line therapy and during follow-up in patients with Hodgkin lymphoma: a retrospective study. *Annals of Oncology* 2010;**21**(6):1222-7.

Mocikova 2011 *{published data only}*

Mocikova H, Pytlik R, Markova J, Steinerova K, Kral Z, Belada D, et al. Pre-transplant positron emission tomography in patients with relapsed Hodgkin lymphoma. *Leukemia & Lymphoma* 2011;**52**(9):1668-74.

Molnar 2010 *{published data only}*

Molnar Z, Simon Z, Borbenyi Z, Deak B, Galuska L, Keresztes K, et al. Prognostic value of FDG-PET in Hodgkin lymphoma for posttreatment evaluation. Long term follow-up results. *Neoplasma* 2010;**57**(4):349-54.

Moskowitz 2015 *{published data only}*

Moskowitz CH. Early FDG-PET adapted treatment improves the outcome of early FDG-PET-positive patients with stages I/II hodgkin lymphoma (HL): Final results of the randomized intergroup EORTC/LYSA/FIL H10 trial. *Clinical advances in Hematology & Oncology: H&O* 2015;**13**(8 Supplement 9):16-7.

Naumann 2001 *{published data only}*

Naumann R, Vaic A, Beuthien-Baumann B, Bredow J, Kropp J, Kittner T, et al. Prognostic value of positron emission tomography in the evaluation of post-treatment residual mass in patients with Hodgkin's disease and non-Hodgkin's lymphoma. *British Journal of Haematology* 2001;**115**(4):793-800.

NCT00784537 *{published data only}*

NCT00784537. High-dose chemotherapy and stem cell transplantation, in patients PET-2 positive, after 2 courses of ABVD and comparison of RT versus no RT in PET-2 negative patients. clinicaltrials.gov/show/nct00784537 (first received 4 November 2008).

NCT00795613 *{published data only}*

NCT00795613. Positron Emission Tomography (PET)-adapted chemotherapy In advanced Hodgkin Lymphoma (HL). clinicaltrials.gov/show/nct00795613 (first received 21 November 2008).

NCT01358747 *{published data only}*

NCT01358747. Study of a treatment driven by early PET response to a treatment not monitored by early PET in patients with AA stage 3-4 or 2B HL. clinicaltrials.gov/show/nct01358747 (first received 24 May 2011).

NCT01652261 *{published data only}*

NCT01652261. Very early FDG-PET/CT-response adapted therapy for advanced Hodgkin Lymphoma (H11). clinicaltrials.gov/show/nct01652261 (first received 30 July 2012).

NCT02292979 *{published data only}*

NCT02292979. Brentuximab vedotin associated with chemotherapy in untreated patients with Hodgkin Lymphoma. clinicaltrials.gov/show/nct02292979 (first received 18 November 2014).

Nguyen 2017 *{published data only}*

Nguyen VT, Pophali PA, Tsai J P, Jagadeesh D, Dean RM, Pohlman B, et al. Early stage, bulky Hodgkin lymphoma patients have a favorable outcome when treated with or without consolidative radiotherapy: potential role of PET scan in treatment planning. *British Journal of Haematology* 2017;**179**(4):674-6.

Panizo 2004 *{published data only}*

Panizo C, Perez-Salazar M, Bendandi M, Rodriguez-Calvillo M, Boan JF, Garcia-Velloso MJ, et al. Positron emission tomography using 18F-fluorodeoxyglucose for the evaluation of residual Hodgkin's disease mediastinal masses. *Leukemia and Lymphoma* 2004;**45**(9):1829-33.

Paolini 2007 *{published data only}*

Paolini R, Rampin L, Rodella E, Ramazzina E, Banti E, Al-Nahhas A, et al. The prognostic value of 18F-FDG PET-CT in the management of Hodgkin's lymphoma: Preliminary results of a prospective study. *Nuclear Medicine Review* 2007;**10**(2):87-90.

Pavlovsky 2019 *{published data only}*

Pavlovsky A, Fernandez I, Kurgansky N, Prates V, Zoppegno L, Negri P, et al. PET-adapted therapy after three cycles of ABVD for all stages of Hodgkin lymphoma: results of the GATLA LH-05 trial. *British Journal of Haematology* 2019;**185**(5):865-73.

Pichler 2000 *{published data only}*

Pichler R, Maschek W, Hatzl-Griesenhofer M, Huber H, Wimmer G, Wahl G, et al. Clinical value of FDG hybrid-PET in staging and restaging of malignant lymphoma - compared with conventional diagnostic methods. *NuklearMedizin* 2000;**39**(6):166-73.

Reinhardt 2005 *{published data only}*

Reinhardt MJ, Herkel C, Altehoefer C, Finke J, Moser E. Computed tomography and 18F-FDG positron emission tomography for therapy control of Hodgkin's and non-Hodgkin's lymphoma patients: when do we really need FDG-PET? *Annals of Oncology* 2005;**16**(9):1524-9.

Rigacci 2002 *{published data only}*

Rigacci L, Castagnoli A, Carpaneto A, Carrai V, Vaggelli L, Matteini M. Can (18)F-FDG PET after first cycle chemotherapy predict the efficacy of therapy in Hodgkin's disease? *Haematologica* 2002;**87**(5):ELT24.

Rigacci 2017 *{published data only}*

Rigacci L, Puccini B, Zinzani P, Kovalchuk S, Broccoli A, Evangelista A, et al. Clinical characteristics of patients with negative interim-pet and positive final PET: data from the prospective PET-oriented HD0801 study by Fondazione Italiana linfomi (FIL). *Hematological Oncology* 2017;**35**(Suppl. 2):38.

Rubello 2015 *{published data only}*

Rubello D, Gordien P, Morliere C, Guyot M, Bordenave L, Colletti PM, et al. Variability of hepatic 18F-FDG uptake at interim PET in patients with Hodgkin lymphoma. *Clinical Nuclear Medicine* 2015;**40**:e405-10.

Sakr 2017 *{published data only}*

Sakr R, Massoud M, Kerbage F, Rached L, Zeghondy J, Akoury E, et al. Real-life Experience for Integration of PET-CT in the Treatment of Hodgkin Lymphoma in Lebanon. *Clinical Lymphoma, Myeloma & Leukemia* 2017;**17S**:S92-5.

Schot 2007 *{published data only}*

Schot BW, Zijlstra JM, Sluiter WJ, van Imhoff GW, Pruim J, Vaalburg W, et al. Early FDG-PET assessment in combination with clinical risk scores determines prognosis in recurring lymphoma. *Blood* 2007;**109**(2):486-91.

Simontacchi 2015 *{published data only}*

Simontacchi G, Filippi AR, Ciammella P, Buglione M, Saieva C, Magrini SM, et al. Interim PET after two ABVD cycles in early-stage Hodgkin lymphoma: Outcomes following the continuation of chemotherapy plus radiotherapy. *International Journal of Radiation Oncology Biology Physics* 2015;**92**:1077-83.

Slaby 2002 *{published data only}*

Slaby J, Belohlavek O, Taborska K, Prochazka M, Trneny M, Klener P. [Predictive features of positron emission tomography after two cycles of induction therapy in malignant lymphoma]. *Casopis Lékaru Ceských [Journal of Czech Physicians]* 2002;**141**(10):312-5.

Spaepen 2001 *{published data only}*

Spaepen K, Stroobants S, Dupont P, Thomas J, Vandenberghe P, Balzarini J, et al. Can positron emission tomography with [(18)F]-fluorodeoxyglucose after first-line treatment distinguish Hodgkin's disease patients who need additional therapy from others in whom additional therapy would mean avoidable toxicity? *British Journal of Haematology* 2001;**115**:272-8.

Specht 2007 *{published data only}*

Specht L. FDG-PET scan and treatment planning for early stage Hodgkin lymphoma. *Radiotherapy and Oncology* 2007;**85**(2):176-7.

Spinner 2018 *{published data only}*

Spinner MA, Advani RH. Risk-adapted therapy for advanced-stage Hodgkin lymphoma. *Hematology* 2018;**1**:200-6.

Straus 2018 *{published data only}*

Straus DJ, Jung SH, Pitcher B, Kostakoglu L, Grecula JC, Hsi ED, et al. CALGB 50604: risk-adapted treatment of nonbulky early-stage Hodgkin lymphoma based on interim PET. *Blood* 2018;**132**(10):1013-21.

Strigari 2016 *{published data only}*

Strigari L, Attili A, Duggento A, Chiaravalloti A, Schillaci O, Guerrisi MG. Quantitative analysis of basal and interim PET/CT images for predicting tumor recurrence in patients with Hodgkin's lymphoma. *Nuclear Medicine Communications* 2016;**37**:16-22.

Sucak 2011 *{published data only}*

Sucak GT, Ozkurt ZN, Suyani E, Yasar DG, Akdemir OU, Aki Z, et al. Early post-transplantation positron emission tomography in patients with Hodgkin lymphoma is an independent prognostic factor with an impact on overall survival. *Annals of Hematology* 2011;**90**(11):1329-36.

Tirelli 2015 *{published data only}*

Tirelli U, Spina M. PET-adapted salvage therapy in Hodgkin's lymphoma. *Lancet Oncology* 2015;**16**(3):239-40.

Tomita 2015 *{published data only}*

Tomita N, Hattori Y, Fujisawa S, Hashimoto C, Taguchi J, Takasaki H, et al. Post-therapy 18F-fluorodeoxyglucose positron emission tomography for predicting outcome in patients with peripheral T cell lymphoma. *Annals of Hematology* 2015;**94**(3):431-6.

Torizuka 2004 *{published data only}*

Torizuka T, Nakamura F, Kanno T, Futatsubashi M, Yoshikawa E, Okada H, et al. Early therapy monitoring with FDG-PET in aggressive non-Hodgkin's lymphoma and Hodgkin's lymphoma. *European Journal of Nuclear Medicine and Molecular Imaging* 2004;**31**(1):22-8.

Trotman 2017 *{published data only}*

Trotman J, Fossa A, Federico M, Stevens L, Kirkwood A, Clifton-Hadley L, et al. Response-adjusted therapy for advanced Hodgkin lymphoma (RATHL) trial: longer follow up confirms efficacy of de-escalation after a negative interim PET scan (CRUK/07/033). *Hematological Oncology* 2017;**35**:65-7.

Tseng 2012 *{published data only}*

Tseng D, Rachakonda LP, Su Z, Advani R, Horning S, Hoppe RT, et al. Interim-treatment quantitative PET parameters predict progression and death among patients with Hodgkin's disease. *Radiation Oncology* 2012;**7**:5.

Villa 2018 *{published data only}*

Villa D, Sehn LH, Aquino-Parsons C, Tonseth P, Scott D W, Gerrie AS, et al. Interim PET-directed therapy in limited stage Hodgkin lymphoma initially treated with ABVD. *Haematologica* 2018;**12**:12.

Weidmann 1999 *{published data only}*

Weidmann E, Baican B, Hertel A, Baum RP, Chow KU, Knupp B, et al. Positron emission tomography (PET) for staging and evaluation of response to treatment in patients with Hodgkin's disease. *Leukemia and Lymphoma* 1999;**34**(5-6):545-51.

Wilson 2018 *{published data only}*

Wilson D, Benard F, Gascoyne RD, Slack GW, Farinha P, Morris J, et al. Interim PET-directed therapy in limited-stage Hodgkin lymphoma initially treated with ABVD. *Haematologica* 2018;**103**(12):e590-3.

Xie 2018 *{published data only}*

Xie W, Jiang X F, Zhao W L, Wang L. Prognostic evaluation of different PET/CT reading methods in Hodgkin lymphoma and diffused large B-cell lymphoma. *Journal of Shanghai Jiaotong University* 2018;**38**(8):954-9.

Yasgur 2015 *{published data only}*

Yasgur BS. Interim PET results guide ongoing therapy in Hodgkin lymphoma. *Oncology Report* 2015;**11**(8):available at: https://www.mdedge.com/hematologynews/nhlhub/article/101136/indolent-lymphoma/interim-pet-results-guide-ongoing-therapy.

Yoshimi 2008 *{published data only}*

Yoshimi A, Izutsu K, Takahashi M, Kako S, Oshima K, Kanda Y, et al. Conventional allogeneic hematopoietic stem cell transplantation for lymphoma may overcome the poor prognosis associated with a positive FDG-PET scan before transplantation. *American Journal of Hematology* 2008;**83**(6):477-81.

Zabrocka 2016 *{published data only}*

Zabrocka E, Sierko E, Wojtukiewicz MZ. Positron emission tomography scanning in the management of Hodgkin lymphoma patients: a single-institution experience. *Advances in Clinical and Experimental Medicine* 2016;**25**(6):1185-92.

Zaucha 2009 *{published data only}*

Zaucha J, Danielewicz I, Malkowski B, Zaucha R, Lesniewski-Kmak K. The role of PET for interim response assessment in patients with Hodgkin's lymphoma. *Wspolczesna Onkologia* 2009;**13**(4):161-6.

Zinzani 1999 *{published data only}*

Zinzani PL, Magagnoli M, Chierichetti F, Zompatori M, Garraffa G, Bendandi M, et al. The role of positron emission tomography (PET) in the management of lymphoma patients. *Annals of Oncology* 1999;**10**(10):1181-4.

Zinzani 2002 *{published data only}*

Zinzani PL, Chierichetti F, Zompatori M, Tani M, Stefoni V, Garraffa G, et al. Advantages of positron emission tomography (PET) with respect to computed tomography in the follow-up of lymphoma patients with abdominal presentation. *Leukemia and Lymphoma* 2002;**43**(6):1239-43.

Zinzani 2016 *{published data only}*

Zinzani PL, Broccoli A, Gioia DM, Castagnoli A, Ciccone G, Evangelista A, et al. Interim positron emission tomography response-adapted therapy in advanced-stage hodgkin lymphoma: final results of the phase II part of the HD0801 study. Journal of Clinical Oncology 2016;**34**(12):1376-85.

References to studies awaiting assessment

Abramson 2010 *{published data only}*

Abramson JS, Barnes JA, LaCasce AS, Zukotynski K, Israel D, Feng Y, et al. End of treatment but not interim PET scan predicts outcome in non-bulky limited stage Hodgkin lymphoma. *Haematologica* 2010;**95**(Suppl. 4):S16.

Algrin 2010 *{published data only}*

Algrin C, Berenger N, Chevret S, Vercellino L, De Bazelaire C, Brice P, et al. Interim-positron emission tomography with [18F]fluorodeoxyglucose (interim-PET) evaluation in mediastinal lymphoma including Hodgkin lymphoma (HL) and primary mediastinal large B-cell lymphoma (PMBL). *Blood* 2010;**116**(21):2860.

Arce-Calisaya 2013 *{published data only}*

Arce-Calisaya P, Scarsbrook A, Thygesen H, Chowdhury F, Patel C. Interim FDG PET/CT in Hodgkin's lymphoma - Does binary response assessment criteria have any prognostic value? *European Journal of Nuclear Medicine and Molecular Imaging* 2013;**40**(Suppl. 2):S476.

Baratto 2015 *{published data only}*

Baratto L, Guerra L, Elisei F, Crivellaro C, De Ponti E, Bolis S, et al. Interim-PET in Hodgkin lymphoma: Deauville criteria and metabolic parameters as prognostic factors. *Clinical and Translational Imaging* 2015;**3**(Suppl. 1):S16-7.

Barna 2011 *{published data only}*

Barna S, Fedinecz N, Magyari F, Varga J, Illes A, Garai I. Prognostic value of interim 18FDG-PET/CT in patients with Hodgkin's lymphoma using different 5-point visual scales for interpretation. *European Journal of Nuclear Medicine and Molecular Imaging* 2011;**38**(Suppl. 2):S377.

Barrington 2011 *{published data only}*

Barrington SF, Kostakoglu L, Hutchings M, Meignan M, Biggi A, Gregianin M, et al. Are the Deauville criteria a reliable tool for assessment of interim PET in Hodgkin lymphoma? In: European Journal of Nuclear Medicine and Molecular Imaging. Vol. 38. 2011:S164.

Bentur 2017 *{unpublished data only}*

Bentur OS, Eldad DJ, Paran E, Lavie D, Nachmias B, Dally N, et al. The predictive value of interim PET-CT in elderly patients with Hodgkin lymphoma. In: Haematologica. Conference: 22th Congress of the European Hematology Association. Spain. Vol. 102. 2017:460-1.

Berenger 2010 *{published data only}*

Berenger N, Vercellino L, Algrin C, Groheux D, Hindie E, Lussato D, et al. Prognostic value of interim 18F-FDG PET/CT in mediastinal bulky Hodgkin lymphoma. *European Journal of Nuclear Medicine and Molecular Imaging* 2010;**37**(Suppl. 2):S436.

Bhatwadekar 2017 *{published data only}*

Bhatwadekar S, Deshpande S, Khadse S, Shah B, Desai D, Kachchhi U, et al. Excellent outcome in Hodgkin lymphoma with ABVD and CMT: A single-centre retrospective analysis. *Hematological Oncology* 2017;**35**(Suppl. 2):316-7.

Cimino 2014 *{published data only}*

Cimino G, Zaucha JM, Cirillo S, Saviolo C, Hutchings M, El-Galaly TC, et al. The complementary prognostic role of baseline and interim PET in predicting treatment outcome in advanced-stage Hodgkin lymphoma. *Blood* 2014;**124**(21):4405.

Cocorocchio 2009 *{published data only}*

Cocorocchio E, Vanazzi A, Botteri E, Alietti A, Negri M, Bassi S, et al. Prognostic role of interim 18FDG-PET in Hodgkin lymphoma: A single-center experience. *Journal of Clinical Oncology* 2009;**27**(15 suppl. 1):e19520.

Cocorocchio 2011 *{published data only}*

Cocorocchio E, Botteri E, Gigli F, Bassi S, Bertazzoni P, Sammassimo S, et al. Evaluation of interim 18FDG-PET in advanced Hodgkin lymphoma (HL) patients (PTS) treated with ChlVPP/ABVVP regimen. In: Annals of Oncology. Vol. 22. 2011:216.

Copeland 2010 *{published data only}*

Copeland A, Fanale M, Chuang H, Macapinlac H, Faria SC, Siegmund B, et al. Single institution experience with interim PET evaluation in newly diagnosed CHL receiving ABVD chemotherapy: Need for standardization. *Haematologica* 2010;**95**(Suppl. 4):S50.

Cuzzocrea 2015 *{published data only}*

Cuzzocrea M, Guerra L, Elisei F, Crivellaro C, De Ponti E, Bolis S, et al. The Deauville criteria and metabolic parameters as prognostic factors in interim PET in Hodgkin lymphoma: A single centre experience. *European Journal of Nuclear Medicine and Molecular Imaging* 2015;**42**(1 suppl. 1):S152-3.

De Rueda 2013 *{published data only}*

De Rueda B, Costilla L, Catalina S, Grasa J, Rubio D, Giraldo P. Prognostic value of 18F-FDG PET/CT in Hodgkin lymphoma. *Haematologica* 2013;**98**(Suppl. 1):572-3.

Fabbri 2011 *{published data only}*

Fabbri A, Rigacci L, Lazzi S, Di Lollo S, Pietrini A, Puccini B, et al. 'Early FDG-PET' predicts clinical course of Hodgkin's lymphoma although does not correlate with macrophages infiltration in diagnostic specimens. *Haematologica* 2011;**96**(Suppl. 2):321-2.

Fiore 2010 *{published data only}*

Fiore F, Viviani S, Luminari S, Levis A, Di Raimondo F, Merli F, et al. Early interim FDG-PET during intensified BEACOPP therapy for advanced-stage Hodgkin disease shows a lower positive predictive value than during ABVD. *Haematologica* 2010;**95**(Suppl. 4):S19.

Gallegos 2012 *{published data only}*

Gallegos C, De Rueda B, Grasa JM, Banso A, Giraldo P. The importance of PET/CT as method of evaluation of early response to treatment in HL. *Haematologica* 2012;**97**(Suppl. 1):566.

Hohaus 2015 *{published data only}*

Hohaus S, Cuccaro A, Annunziata S, Martini M, D'Alo F, Calcagni ML, et al. The risk of progression of Hodgkin lymphoma in patients with negative interim PET: A role for the number of tumor-infiltrating macrophages (CD68+ cell counts) and B symptoms. *Haematologica* 2015;**100**(Suppl. 3):7.

Hutchings 2010 *{published data only}*

Hutchings M, Kostakoglu L, Loft A, Coleman M, Specht L. Correlation of FDG-PET results after one cycle and after two cycles of chemotherapy in Hodgkin lymphoma. *Journal of Clinical Oncology* 2010;**28**(15):8061.

Knight-Greenfield 2013 *{published data only}*

Knight-Greenfield A, Marshall RA, Hutchings M, Doucette J, Stern J, Coleman M, et al. Interim FDG PET/CT to predict progression-free survival (PFS) better than clinical and baseline metabolic measurements in Hodgkin lymphoma (cHL). *Journal of Clinical Oncology* 2013;**31**(15):8555.

Leontjeva 2016 *{published data only}*

Leontjeva A, Demina E, Ryabukhina J, Tumyan G, Trofimova O, Sotnikov V, et al. Significance of early interim PET results in advanced Hodgkin lymphoma treated intensive program EACOPP-14. *British Journal of Haematology* 2016;**173**(Suppl. 1):107.

Luminari 2010 *{published data only}*

Luminari S, Cesaretti M, Tomasello C, Guida A, Bagni B, Merli F, et al. The use of FDG positron emission tomography (FDG-PET) in patients with Hodgkin lymphoma (HL) in the "real world": A population based study from northern Italy. *Haematologica* 2010;**95**(Suppl. 4):S42-3.

Luminari 2011 *{published data only}*10.3109/10428194.2011.580475

Luminari S, Cesaretti M, Tomasello C, Guida A, Bagni B, Merli F, et al. Use of 2-[18F]fluoro-2-deoxy-D-glucose positron emission tomography in patients with Hodgkin lymphoma in daily practice: a population-based study from Northern Italy. *Leukemia and Lymphoma* 2011;**52**(9):1689-96.

Medvedovskaya 2016 *{published data only}*

Medvedovskaya E, Leontjeva A, Demina E, Ryabukhina J, Tumyan G, Kokosadze N, et al. The impact of outcome of

interim PET/CT on advanced Hodgkin lymphoma treated with EACOPP-14. *Haematologica* 2016;**101**(Suppl. 5):26.

Molnar 2011 *{published data only}*

Molnar Z, Deak B, Kajary K, Lengyel Z, Molnar P, Rosta A, et al. The value of interim 18F-FDG PET/CT in Hodgkin lymphoma. *European Journal of Nuclear Medicine and Molecular Imaging* 2011;**38**(Suppl. 2):S165.

Molnar 2011a *{published data only}*

Molnar Z, Deak B, Kajary K, Lengyel Z, Molnar P, Rosta A, et al. Interim FDG PET/CT examinations in advanced stage Hodgkin lymphoma. *Nuclear Medicine Review* 2011;**14**(Suppl. A):A4.

Moreira 2013 *{published data only}*

Moreira C, Fidalgo P, Queiroz L, Domingues N, Moreira I, Oliveira I, et al. Prognostic value of interim vs. end-of-treatment PET scan in Hodgkin's lymphoma. *Hematological Oncology* 2013;**31**(Suppl. 1):257.

Perrone 2009 *{published data only}*

Perrone T, Gaudio F, Giordano A, De Risi C, Spina A, Curci P, et al. Role of positron emission tomography (PET) after 2 and 4 courses of chemotherapy in patients with Hodgkin's lymphoma: A single center experience. *Haematologica* 2009;**94**(Suppl. 4):216.

Pophali 2014 *{published data only}*

Pophali PA, Rybicki LA, Fenner KB, Jagadeesh D, Dean RM, Pohlman B, et al. Bulky disease does not adversely affect overall survival in early stage Hodgkin lymphoma: Role of interim PET and possible omission of radiotherapy in select patients. *Blood* 2014;**124**:21.

Rusconi 2010 *{published data only}*

Rusconi C, Ravelli E, Gabutti C, Zilioli V, Meli E, Nichelatti M, et al. Baseline and dynamic prognostic factors in newly diagnosed classical Hodgkin's lymphoma. *Haematologica* 2010;**95**(Suppl. 4):S43.

Spallino 2017 *{published data only}*

Spallino M, Guerra L, Cuzzocrea M, Elisei F, Crivellaro C, De Ponti E, et al. The Deauville criteria and QPET as prognostic factors in interim PET in adult Hodgkin lymphoma: A single centre experience. *Clinical and Translational Imaging* 2017;**5**(Suppl. 1):S43.

Yaghmour 2012 *{published data only}*

Yaghmour G, Farhat M, Valdivieso BS, Janakiraman N. PET-negative at 2, 3 or 4 cycles of ABVD in Hodgkin's lymphoma is still good. *Journal of General Internal Medicine* 2012;**27**(Suppl. 2):S258-9.

Zanoni 2011 *{published data only}*

Zanoni L, Agostinelli C, Gallamini A, Rigacci L, Sista M, Piccaluga P, et al. The predictive value of interim PET and immunohistochemical markers in Hodgkin lymphoma (HL). *European Journal of Nuclear Medicine and Molecular Imaging* 2011;**38**(Suppl. 2):S165.

References to ongoing studies

NCT00736320 *{unpublished data only}*

Kobe C, Dietlein M, Kuhnert G, Holstein A, Kahraman D, Haverkamp H, et al. Recruitment and PET interpretation in the HD16 trial for early stage Hodgkin lymphoma. Treatment stratification by FDG-PET. In: NuklearMedizin. Vol. 50. Jahrestagung der Deutschen Gesellschaft für Nuklearmedizin. Bremen, Germany. 2012:A39-40.

NCT00736320. HD16 for early stages - treatment optimization trial in the first-line treatment of early stage Hodgkin lymphoma; treatment stratification by means of FDG-PET. clinicaltrials.gov/ct2/show/nct00736320 (first received 15 August 2008).

Additional references

Adams 2015a

Adams HJ, Nievelstein RA, Kwee TC. Prognostic value of interim FDG-PET in Hodgkin lymphoma: systematic review and meta-analysis. *British Journal of Haematology* 2015;**170**(3):356-66.

Adams 2016a

Adams HJ, Kwee TC. Controversies on the prognostic value of interim FDG-PET in advanced-stage Hodgkin lymphoma. *European Journal of Haematology* 2016;**97**(6):491-8.

Altman 1999

Altman DG. Practical Statistics for Medical Research. 1 edition. London: Chapman & Hall, 1991.

Altman 2012

Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. *PLOS Medicine* 2012;**9**(5):e1001216. [PMID: 22675273]

Amitai 2018

Amitai I, Gurion R, Vidal L, Dann EJ, Raanani P, Gafter-Gvili A. PET-adapted therapy for advanced Hodgkin lymphoma – systematic review. *Acta Oncologica* 2018;**57**(6):765-72.

Barrington 2014

Barrington SF, Mikhaeel NG, Kostakoglu L, Meignan M, Hutchings M, Mueller SP, et al. Role of imaging in the staging and response assessment of lymphoma: consensus of the International Conference on Malignant Lymphomas Imaging Working Group. *Journal of Clinical Oncology* 2014;**32**(27):3048-58.

Barrington 2017a

Barrington SF, Kluge R. FDG PET for therapy monitoring in Hodgkin and non-Hodgkin lymphomas. *European Journal of Nuclear Medicine and Molecular Imaging* 2017;**44**(1):97-110.

Berriolo-Riedinger 2018

Berriolo-Riedinger A, Becker S, Casasnovas O, Vander Borght T, Edeline V. Role of FDG PET-CT in the treatment management of Hodgkin lymphoma. *Cancer Radiotherapy* 2018;**22**(5):393-400.

Blank 2017

Blank O, von Tresckow B, Monsef I, Specht L, Engert A, Skoetz N. Chemotherapy alone versus chemotherapy plus radiotherapy for adults with early stage Hodgkin lymphoma. *Cochrane Database of Systematic Reviews* 2017, Issue 4. Art. No: CD007110. [DOI: 10.1002/14651858.CD007110.pub3]

Boellaard 2010

Boellaard R, O'Doherty MJ, Weber WA, Mottaghy FM, Lonsdale MN, Stroobants SG, et al. FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. *European Journal of Nuclear Medicine and Molecular Imaging* 2010;**37**(1):181-200. [PMID: 19915839]

Borchmann 2011

Borchmann P, Haverkamp H, Diehl V, Cerny T, Markova J, Ho AD, et al. Eight cycles of escalated-dose BEACOPP compared with four cycles of escalated-dose BEACOPP followed by four cycles of baseline-dose BEACOPP with or without radiotherapy in patients with advanced-stage Hodgkin's lymphoma: final analysis of the HD12 trial of the German Hodgkin Study Group. *Journal of Clinical Oncology* 2011;**29**(32):4234-42. [PMID: 21990399]

Bouwmeester 2012

Bouwmeester W, Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and methods in clinical prediction research: a systematic review. *PLOS Medicine* 2012;**9**(5):1-12.

Bröckelmann 2018

Bröckelmann PJ, Eichenauer DA, Jakob T, Follmann M, Engert A, Skoetz N. Clinical practice guideline: Hodgkin lymphoma in adults—diagnosis, treatment, and follow-up. *Deutsches Ärzteblatt International* 2018;**115**:535-40.

Canellos 1992

Canellos GP, Anderson JR, Propert KJ, Nissen N, Cooper MR, Henderson ES, et al. Chemotherapy of advanced Hodgkin's disease with MOPP, ABVD, or MOPP alternating with ABVD. *New England Journal of Medicine* 1992;**327**(21):1478-84. [PMID: 1383821]

Cheson 2014

Cheson BD, Fisher RI, Barrington SF, Cavalli F, Schwartz LH, Zucca E, et al. Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification. *Journal of Clinical Oncology* 2014;**32**(27):3059-68. [PMID: 25113753]

Covidence [Computer program]

Veritas Health Innovation Covidence. Version accessed 4 March 2019. Melbourne, Australia.: Veritas Health Innovation. Available at www.covidence.org.

Cuccaro 2014

Cuccaro A, Bartolomei F, Cupelli E, Galli E, Giachelia M, Hohaus S. Prognostic factors in hodgkin lymphoma. *Mediterranean Journal of Hematology and Infectious Diseases* 2014;**6**(1):e2014053.

Debray 2017

Debray T, Damen J, Snell K, Ensor J, Hooft L, Reitsma J, et al. A guide to systematic review and meta-analysis of prediction model performance. *BMJ* 2017;**356**:i6460.

Debray 2018

Debray T, Moons K, Riley R. Detecting small-study effects and funnel plot asymmetry in meta-analysis of survival data: A comparison of new and existing tests. *Research Synthesis Methods* 2018;**9**(1):41-50.

Deeks 2011

Deeks JJ, Higgins JP, Altman DG. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (editors), Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Engert 2003

Engert A, Schiller P, Josting A, Herrmann R, Koch P, Sieber M, et al. Involved-field radiotherapy is equally effective and less toxic compared with extended-field radiotherapy after four cycles of chemotherapy in patients with early-stage unfavorable Hodgkin's lymphoma: results of the HD8 trial of the German Hodgkin's Lymphoma Study Group. *Journal of Clinical Oncology* 2003;**21**(19):3601-8.

Engert 2007

Engert A, Franklin J, Eich HT, Brillant C, Sehlen S, Cartoni C, et al. Two cycles of doxorubicin, bleomycin, vinblastine, and dacarbazine plus extended-field radiotherapy is superior to radiotherapy alone in early favorable Hodgkin's lymphoma: final results of the GHSG HD7 trial. *Journal of Clinical Oncology* 2007;**25**(23):3495-502.

Engert 2010

Engert A, Plutschow A, Eich HT, Lohri A, Dorken B, Borchmann P, et al. Reduced treatment intensity in patients with early-stage Hodgkin's lymphoma. *New England Journal of Medicine* 2010;**363**(7):640-52. [PMID: 20818855]

Engert 2012

Engert A, Haverkamp H, Kobe C, Markova J, Renner C, Ho A, et al. Reduced-intensity chemotherapy and PET-guided radiotherapy in patients with advanced stage Hodgkin's lymphoma (HD15 trial): a randomised, open-label, phase 3 non-inferiority trial. *Lancet* 2012;**379**(9828):1791-9. [PMID: 22480758]

Fitzgerald 2019

Fitzgerald TJ. Optimisation of adaptive therapy for advanced Hodgkin lymphoma. *Lancet Oncology* 2019;**20**(2):167-8.

Franklin 2005

Franklin JG, Paus MD, Pluetschow A, Specht L. Chemotherapy, radiotherapy and combined modality for Hodgkin's disease, with emphasis on second cancer risk. *Cochrane Database of Systematic Reviews* 2005, Issue 4. Art. No: CD003187. [DOI: 10.1002/14651858.CD003187.pub2] [PMID: 16235316]

Gallamini 2007

Gallamini A, Hutchings M, Rigacci L, Specht L, Merli F, Hansen M, et al. Early interim 2-[18F]fluoro-2-deoxy-D-glucose positron emission tomography is prognostically superior to international prognostic score in advanced-stage Hodgkin's lymphoma: a report from a joint Italian-Danish study. *Journal of Clinical Oncology* 2007;**25**(24):3746-52. [PMID: 17646666]

Geersing 2012

Geersing GJ, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons KG. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. *PLOS One* 2012;**7**(7):e32844.

Hayden 2013

Hayden JA, van der Windt DA, Cartwright JL, Cote P, Bombardier C. Assessing bias in studies of prognostic factors. *Annals of Internal Medicine* 2013;**158**(4):280-6. [PMID: 23420236]

Higgins 2011

Higgins JP, Deeks JJ (editors). Chapter 7: Selecting studies and collecting data. In: Higgins JP, Green S (editors), Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Higgins 2011b

Higgins JP, Deeks JJ, Altman DG. Chapter 16: Special topics in statistics. In: Higgins JP, Green S, editors(s). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011.

Howlader 2015

Howlader N, Noone AM, Krapcho M, Garshell J, Neyman N, Altekruse SF, et al. SEER Cancer Statistics Review, 1975-2010. National Cancer Institute. Bethesda, MD. http://seer.cancer.gov/csr/1975_2010/ based on November 2012 SEER data submission, posted to the SEER web site, April 2013 (accessed 23 June 2015).

Iorio 2015

Iorio A, Spencer FA, Falavigna M, Alba C, Lang E, Burnand B, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. *BMJ* 2015;**350**:1-8.

Josting 2010

Josting A. Prognostic factors in Hodgkin lymphoma. *Expert Review of Hematology* 2010;**3**(5):583-92.

Juweid 2007

Juweid ME, Stroobants S, Hoekstra OS, Mottaghy FM, Dietlein M, Guermazi A, et al. Use of positron emission tomography for response assessment of lymphoma: consensus of the Imaging Subcommittee of International Harmonization Project in Lymphoma. *Journal of Clinical Oncology* 2007;**25**:571-8.

Klimm 2005

Klimm B, Engert A, Diehl V. First-line treatment of Hodgkin's lymphoma. *Current Hematology Reports* 2005;**4**(1):15-22.

Kobe 2008a

Kobe C, Dietlein M, Franklin J, Markova J, Lohri A, Amthauer H, et al. Positron emission tomography has a high negative predictive value for progression or early relapse for patients with residual disease after first-line chemotherapy in advanced-stage Hodgkin lymphoma. *Blood* 2008;**112**(10):3989-94. [PMID: 18757777]

Kobe 2010

Kobe C, Dietlein M, Fuchs M. Interpretation and validation of interim positron emission tomography in Hodgkin lymphoma. Leukemia and Lymphoma 2010;**51**(3):552-3. [PMID: 20141443]

Kobe 2010a

Kobe C, Dietlein M, Kriz J, Furth C, Fuchs M, Borchmann P, et al. The role of PET in Hodgkin's lymphoma and its impact on radiation oncology. *Expert Review of Anticancer Therapy* 2010;**10**(9):1419-28. [PMID: 20836677]

Kyzas 2005

Kyzas PA, Loizou KT, Ioannidis JP. Selective reporting biases in cancer prognostic factor studies. *Journal of the National Cancer Institute* 2005;**97**(14):1043-55. [PMID: 16030302]

Kyzas 2007

Kyzas PA, Denaxa-Kyza D, Ioannidis JP. Quality of reporting of cancer prognostic marker studies: association with reported prognostic effect. *Journal of the National Cancer Institute* 2007;**99**(3):236-43. [PMID: 17284718]

Kılıçkap 2013

Kılıçkap S, Barışta I, Ulger S, Celik I, Selek U, Yıldız F, et al. Clinical Features and Prognostic Factors of Hodgkin's Lymphoma: A Single Center Experience. *Balkan Medical Journal* 2013;**30**(2):178-85.

Lefebvre 2011

Lefebvre C, Manheimer E, Glanville J. Chapter 6: Searching for studies. In: Higgins JP, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Lister 1989

Lister TA, Crowther D, Sutcliffe SB, Glatstein E, Canellos GP, Young RC, et al. Report of a committee convened to discuss the evaluation and staging of patients with Hodgkin's disease: Cotswolds meeting. *Journal of Clinical Oncology* 1989;**7**(11):1630-6.

Macaskill 2010

Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y. Chapter 10 Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C, editors(s). Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Version 1.0. The Cochrane Collaboration, 2010. Available from: http://srdta.cochrane.org/.

Mallett 2010

Mallett S, Timmer A, Sauerbrei W, Altman DG. Reporting of prognostic studies of tumour markers: a review of published

articles in relation to REMARK guidelines. *British Journal of Cancer* 2010;**102**(1):173-80. [PMID: 19997101]

Markova 2009

Markova J, Kobe C, Skopalova M, Klaskova K, Dedeckova K, Plutschow A, et al. FDG-PET for assessment of early treatment response after four cycles of chemotherapy in patients with advanced-stage Hodgkin's lymphoma has a high negative predictive value. *Annals of Oncology* 2009;**20**(7):1270-4. [PMID: 19228806]

McShane 2005

McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumour MARKer prognostic studies (REMARK). *British Journal of Cancer* 2005;**93**(4):387-91.

Meignan 2009

Meignan M, Gallamini A, Meignan M, Gallamini A, Haioun C. Report on the First International Workshop on Interim-PET-Scan in Lymphoma. *Leukemia and Lymphoma* 2009;**50**(8):1257-60.

Meignan 2009a

Meignan M, Itti E, Bardet S, Lumbroso J, Edeline V, Olivier P, et al. Development and application of a real-time on-line blinded independent central review of interim PET scans to determine treatment allocation in lymphoma trials. *Journal of Clinical Oncology* 2009;**27**(16):2739-41.

Moher 2009

Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Journal of Clinical Epidemiology* 2009;**62**(10):1006-12.

Moons 2009

Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ* 2009;**338**:b375.

Moons 2014

Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Medicine* 2014;**11**:e1001744.

Moskowitz 2018

Moskowitz CH, Walewski J, Nademanee A, Masszi T, Agura E, Holowiecki J, et al. Five-year PFS from the AETHERA trial of brentuximab vedotin for Hodgkin lymphoma at high risk of progression or relapse. *Blood Journal* 2018;**132**(25):2639.

Parmar 1998

Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Statistics in Medicine* 1998;**17**(24):2815-34.

Peat 2014

Peat G, Riley RD, Croft P, Morley KI, Kyzas PA, Moons KG, et al. Improving the transparency of prognosis research: the role of reporting, data sharing, registration, and protocols. *PLOS Medicine* 2014;**11**(7):e1001671.

Radford 2015

Radford J, Illidge T, Counsell N, Hancock B, Pettengell R, Johnson P, et al. Results of a trial of PET-directed therapy for early-stage Hodgkin's lymphoma. *New England Journal of Medicine* 2015;**372**(1):1598-607.

Rancea 2013

Rancea M, Monsef I, von Tresckow B, Engert A, Skoetz N. High-dose chemotherapy followed by autologous stem cell transplantation for patients with relapsed/refractory Hodgkin lymphoma. *Cochrane Database of Systematic Reviews* 2013, Issue 6. Art. No: CD009411. [DOI: 10.1002/14651858.CD009411.pub2] [PMID: 23784872]

Rancea 2013a

Rancea M, Engert A, von Tresckow B, Halbsguth T, Behringer K, Skoetz N. Hodgkin's lymphoma in adults: diagnosis, treatment and follow-up. *Deutsches Arzteblatt International* 2013;**110**(11):177-83, 183e1-3. [PMID: 23555321]

Re 2005

Re D, Thomas RK, Behringer K, Diehl V. From Hodgkin disease to Hodgkin lymphoma: biologic insights and therapeutic potential. *Blood* 2005;**105**(12):4553-60. [PMID: 15728122]

Review Manager 2014 [Computer program]

Nordic Cochrane Centre, The Cochrane Collaboration Review Manager 5 (RevMan 5). Version 5.3. Copenhagen: Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Riley 2013

Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLOS Medicine* 2013;**10**(2):e1001380. [PMID: 23393429]

Riley 2019

Riley RD, Moons KG, Snell KE, Ensor J, Hooft L, Altman DG, et al. A guide to systematic review and meta-analysis of prognostic factor studies. *BMJ* 2019;**364**:k4597.

Riley 2019b

Riley RD, van der Windt D, Croft P, Moons KGM. Prognosis Research in Healthcare. Concepts, Methods, and Impact. Oxford University Press, 2019.

Sauerbrei 2005

Sauerbrei W. Prognostic factors. Confusion caused by bad quality design, analysis and reporting of many studies. *Advances in Oto-Rhino-Laryngology* 2005;**62**:184-200. [PMID: 15608428]

Sickinger 2015

Sickinger MT, von Tresckow B, Kobe C, Engert A, Borchmann P, Skoetz N. Positron emission tomography-adapted therapy for first-line treatment in individuals with Hodgkin lymphoma. *Cochrane Database of Systematic Reviews* 2015, Issue 1. Art. No: CD010533. [DOI: 10.1002/14651858.CD010533.pub2]

Skoetz 2013

Skoetz N, Trelle S, Rancea M, Haverkamp H, Diehl V, Engert A, et al. Effect of initial treatment strategy on survival of patients with advanced-stage Hodgkin's lymphoma: a systematic review and network meta-analysis. *Lancet. Oncology* 2013;**14**(10):943-52. [PMID: 23948348]

Skoetz 2017a

Skoetz N, Will A, Monsef I, Brillant C, Engert A, von Tresckow B. Comparison of first-line chemotherapy including escalated BEACOPP versus chemotherapy including ABVD for people with early unfavourable or advanced stage Hodgkin lymphoma. *Cochrane Database of Systematic Reviews* 2017, Issue 5. Art. No: CD007941. [DOI: 10.1002/14651858.CD007941.pub3]

Steyerberg 2013

Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLOS Medicine* 2013;**10**(2):e1001381.

Swerdlow 2008

Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H, et al. WHO Classification of Tumours of the Haematopoietic and Lymphoid Tissues. Fourth edition. Vol. **2**. Lyon: WHO Press, 1211 Geneva 27, Switzerland, 2008. [ISBN-13: 9789283224310]

Thomas 2002

Thomas RK, Re D, Zander T, Wolf J, Diehl V. Epidemiology and etiology of Hodgkin's lymphoma. *Annals of Oncology* 2002;**13**:147-52. [PMID: 12401681]

Tierney 2007

Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials* 2007;**8**:16.

Trivella 2006

Trivella M. Systematic Reviews of Prognostic Factor Studies (Section: Estimating the Hazard Ratio) [DPhil]. University of Oxford (Altman DG, thesis advisor) June 10th 2006:219.

von Tresckow 2012

von Tresckow B, Plutschow A, Fuchs M, Klimm B, Markova J, Lohri A, et al. Dose-intensification in early unfavorable Hodgkin's lymphoma: final analysis of the German Hodgkin study group HD14 trial. *Journal of Clinical Oncology* 2012;**30**(9):907-13. [PMID: 22271480]

Weigler-Sagie 2010

Weiler-Sagie M, Bushelev O, Epelbaum R, Dann EJ, Haim N, Avivi I, et al. (18)F-FDG avidity in lymphoma readdressed: a study of 766 patients. *Journal of Nuclear Medicine* 2010;**51**(1):25-30.

References to other published versions of this review

Skoetz 2017

Skoetz N, Collins G, Moons K, Estcourt LJ, Engert A, Kobe C, et al. Interim PET for prognosis in adults with Hodgkin lymphoma: a prognostic factor exemplar review. *Cochrane Database of Systematic Reviews* 2017, Issue 4. Art. No: CD012643. [DOI: 10.1002/14651858.CD012643]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Andre 2017

Study characteristics

Methods	Secondary citation(s) • Raemarkers 2014, Cottereau 2018 Language of publication • English Study design • Prospective, multi-centre, phase III randomised trial Study centre(s) • Various Countries • Belgium, Croatia, Denmark, France, Italy, the Netherlands, Slovakia, Switzerland

Andre 2017 *(Continued)*

	Median follow-up time (range) • 55 months
Participants	Number of included participants • Total: 1925 • Randomised to standard treatment without change in protocol because of interim PET: 954 Inclusion criteria • Previously untreated • Classic supradiaphragmic stage I and II HL • Age 15 to 70 years Exclusion criteria • Previous laparotomy • Concomitant or previous cancer other than basal-cell carcinoma of the skin or in situ carcinoma of the cervix • Concomitant severe illness that would reduce life expectancy • Social circumstances not allowing for proper treatment and follow-up • Positivity for the human immunodeficiency virus (exclusion criteria reported in Fermé 2007[1]) Consent • Yes; written informed consent Recruitment period • November 2006 to June 2011 Age (range, in years) • Favourable, standard treatment group median: 31 (15-49) • Unfavourable, standard treatment group median: 32 (15-70) Ethnic group(s) • Not reported Stages of disease • Early stages (I and II) Comorbidities • Not reported, except for the exclusion criteria Therapy regimen • ABVD and radiotherapy depending on treatment arm, favourable/unfavourable disease, and early PET (ePET) positivity
Prognostic factor(s)	Prognostic factor(s) • Early PET (ePET) Definition of prognostic factor(s) • Not reported

Andre 2017 *(Continued)*

	Timing of prognostic factor measurement • After 2 ABVD cycles Method for measurement (use of specific scale and cut-off) • International Harmonization Project criteria. According to these criteria: PET-negative corresponds to Deauville score 1 (no uptake) and score 2 (uptake ≤ mediastinum) • Central review performed online (up to 6 experts, and one local expert) Was the same definition and method for measurement used in all participants? • Central review started later for 2 centres in Italy due to technical difficulties, only 75% of ePET were centrally reviewed Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Not reported
Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), defined as time from random assignment to date of progression (as experiencing relapse after previous complete remission, progressive disease, or death from any cause) Secondary outcome(s) and definition(s) • Overall survival (OS), not defined Timing of outcome measurement • At 5 years Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • Not applicable
Analysis	**Univariable analysis: Yes** Total number of participants included inunivariateanalysis for each outcome • PFS: all • OS: all Statistical method • Kaplan-Meier method • HR (95% CI) • Randomised arms were compared using the log-rank test stratified by Ann Arbor stage and availability of a baseline FDG-PET scan How was the prognostic factor treated? • Binary

Andre 2017 *(Continued)*

	Multivariable analysis: No
Risk of bias (QUIPS)	Study participation • Low risk • Clear description of participants and study characteristics. Study attrition • Low risk • Length of follow-up reported. Exclusion of participants due to safety amendment during the study. Prognostic factor measurement • Moderate risk • Adequate measurement and description. Central review only for 75% of scans and delayed in the case of 2 centres due to technical difficulties. **Outcome: Overall survival** *Not reported* **Outcome: Progression-free survival** Outcome measurement • Low risk • No definition of outcome. Outcome measured the same way for all participants. 'Other prognostic factors (covariates)' • Low risk Statistical analysis and reporting • Low risk • Statistical method appropriate for the data. **Outcome: Adverse events** *Not reported*
Notes	Conflict of interest • Casasnovas O: honoraria received from Genentech, Takeda, Gilead Sciences, Sanofi; consulting or advisory role at Genentech, Takeda, Gilead Sciences; research funding received from Genentech; travel, accomodation, expenses received from Genentech, Takeda, Gilead Sciences • Brice P: research funding received from Merck Sharp & Dohme Oncology, Takeda; travel, accomodation, expenses received from Takeda • Specht L: consulting or advisory role at Takeda; research funding received from Varian Medical Systems; travel, accomodation, expenses received from Takeda • Delarue R: honoraria received from Servier, Gilead Sciences, Roche, Celgene, Takeda; consulting or advisory role at Gilead Sciences, Roche; Speakers' Bureau at Karyopharm Therapeutics; travel, accomodation, expenses received from Roche, Takeda, Celgene • Hutchings M: consulting or advisory role at Takeda, Genentech, Celgene, Bayer; research funding received from Takeda, Janssen-Cilag, Genentech, Celgene; travel, accomodation, expenses received from Takeda, Bristol-Myers, Squibb, Janssen-Cilag Funding • Supported by European Organisation for Research and Treatment of Cancer (Belgium), LYmphoma Study Association (France), Fondazione Italiana Limfomi (Italy), Fondation Belge Contre le Cancer (Belgium), Dutch Cancer Society (the Netherlands), Institut National du Cancer (France), Assistance

Andre 2017 *(Continued)*

Publique des Hopitaux de Paris (France), Societe Française de Medecine Nucleaire et Imagerie Moleculaire (France), Associazone Angela Serra (Italy), van Vlissingen Lymfoom Fonds (the Netherlands), and Chugai Pharmaceutical (Japan).

[1] Fermé C, Eghbali H, Meerwaldt JH, Rieux C, Bosq J, Berger F, et al. Chemotherapy plus involved-field radiation in early-stage Hodgkin's disease. New England Journal of Medicine 2007;357:1916-1927

Annunziata 2016

Study characteristics

Methods	Secondary citation(s) • NA Language of publication • English Study design • Retrospective, single-centre study Study centre(s) • Not reported Country • Italy Median follow-up time (range) • Not reported
Participants	Number of included participants • 68 Inclusion criteria • HL diagnosis Exclusion criteria • Not reported Consent • Not reported Recruitment period • January 2007 to December 2014 Age (range, in years) • 39 (16-72) Ethnic group(s) • Not reported

Annunziata 2016 *(Continued)*

	Stages of disease • All stages Comorbidities • Not reported Therapy regimen • ABVD according to the presence of risk factors defined by the European Organisation for Research and Treatment of Cancer (EORTC) • Favourable group (age < 50 years with ≤ 3 involved nodal areas, absence of mediastinal bulk (mediastinum-to-thorax ratio < 0.35), and erythrocyte sedimentation rate (ESR) < 50 mm without B symptoms or ESR < 30 mm with B symptoms): 3 cycles ABVD followed by radiotherapy, or 4 cycles ABVD without radiotherapy • Unfavourable group (age ≥ 50 years, > 4 involved nodal areas, presence of mediastinal bulk (mediastinum to-thorax ratio ≥ 0.35), or ESR ≥ 50 mm without B symptoms or ESR ≥ 30 mm with B symptoms): 4 cycles ABVD followed by radiotherapy, or 6 cycles ABVD without radiotherapy (therapy regimen reported in Raemaekers 2014[1])
Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Half-body PET scan (base of the skull to mid-thigh) Timing of prognostic factor measurement • Around day 25 (mean, range 22-27) after cycle 1 of ABVD Method for measurement (use of specific scale and cut-off) • Deauville 5-point scoring system • Scores of 1-3 considered negative, scores of 4-5 considered positive • 2 nuclear medicine physicians interpreted all scans Was the same definition and method for measurement used in all participants? • Yes Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Not reported
Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), with progression during treatment, lack of complete remission at the end of first-line treatment, and relapse counted as adverse events (AE) Secondary outcome(s) and definition(s) • None Timing of outcome measurement • At 2 years Was the same definition and method for measurement used in all participants? • Yes

Annunziata 2016 *(Continued)*

	Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • Not applicable
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • PFS: all Statistical method • Receiver operating characteristic (ROC) approach • Kaplan-Meier (survival analysis) • Log-rank (differences between groups) • Cox proportional hazards model How was the prognostic factor treated? • Binary **Multivariable analysis: No**
Risk of bias (QUIPS)	Study participation • Unclear • Clear description of participants and study characteristics. No inclusion and exclusion criteria provided. Study attrition • Unclear risk • No loss to follow-up reported. No length of follow-up reported. Prognostic factor measurement • Low risk • Adequate measurement and description. Prognostic factor measured the same way for all participants. **Outcome: Overall survival** *Not reported* **Outcome: Progression-free survival** Outcome measurement • Low risk • Clear definition. Outcome measured the same way for all participants. 'Other prognostic factors (covariates)' • High risk • Stated in methods section that multiple factors were taken into account for analysis, but unclear which variables and how adjustment was conducted. Disease stage not accounted for.

Annunziata 2016 *(Continued)*

	Statistical analysis and reporting • High risk • Poorly reported. Unclear whether multivariable analysis was reported. **Outcome: Adverse events** *Not reported*
Notes	Conflict of interest • The authors declare that they have no conflict of interest. Funding • Not reported [1] Raemaekers JM, André MP, Federico M, Girinsky T, Oumedaly R, Brusamolino E, et al. Omitting radiotherapy in early positron emission tomography–negative stage I/II Hodgkin lymphoma is associated with an increased risk of early relapse: clinical results of the preplanned interim analysis of the randomised EORTC/LYSA/FIL H10 trial. Journal of Clinical Oncology 2014;32(12):1188-1194

Barnes 2011

Study characteristics	
Methods	Secondary citation(s) • Sher 2009 Language of publication • English Study design • Retrospective, multi-centre study (2 centres) Study centre(s) • Massachusetts General Hospital Cancer Center and Dana-Farber Cancer Institute, Massachusetts, USA Country • USA Median follow-up time (range) • 46 months
Participants	Number of included participants • 96 Inclusion criteria • Diagnosed with classic, histology-proven HL • Adults • Limited-stage non-bulky disease (mass < 10 cm) • ABVD chemotherapy • Availability of interim PET and end-of-treatment PET

Barnes 2011 *(Continued)*

Exclusion criteria

- Nodular lymphocyte predominant HL

Consent

- Not reported

Recruitment period

- January 2000 to December 2008

Age (range, in years)

- 34 (18-77)

Ethnic group(s)

- Not reported

Stages of disease

- Early stages (I to IIB)

Comorbidities

- Not reported

Therapy regimen

- 4 or 6 cycles of ABVD with or without IFRT

Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Whole-body PET scan (base of the skull to mid-thighs) Timing of prognostic factor measurement • After 2 to 4 treatment cycles Method for measurement (use of specific scale and cut-off) • 2 nuclear medicine physicians interpreted all scans, final result based on consensus • Grading on a 4-point scale with scores 0 or 1 considered negative and scores 2 to 4 considered positive Was the same definition and method for measurement used in all participants? • Yes Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Not reported
Outcome(s)	Primary outcome(s) and definition(s) • Overall survival (OS), defined as the time from initial pathological diagnosis to death from any cause • Progression-free survival (PFS), defined as time from diagnosis to progression or death from any cause Secondary outcome(s) and definition(s) • Overall response rate (ORR), defined as number of subjects with either complete response (CR) or partial response (PR)

Barnes 2011 *(Continued)*

	• Primary refractory disease, defined as progressive disease on treatment or relapse within 3 months of completing therapy Timing of outcome measurement • Unclear: 4 years reported in text, 10 years reported in figure Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • Not reported If yes, how were missing data handled? • Not applicable
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • OS: all • PFS: all Statistical method • Kaplan-Meier (survival analysis) • Log-rank test • Fisher's exact test (CR) How was the prognostic factor treated? • Binary **Multivariable analysis: No**
Risk of bias (QUIPS)	Study participation • Unclear risk • Description of participants provided. Missing interim and end-of-treatment PET was part of the exclusion criteria. No comparison of baseline study sample (n = 155) and participants (n = 96) included. No reasons for missing scans provided. Study attrition • Low risk • No loss to follow-up. Prognostic factor measurement • Moderate risk • Adequate measurement and description. No standardised criteria, but description of scoring system used. Prognostic factor measured the same way for all participants. Blinding not reported. **Outcome: Overall survival** Outcome measurement

Barnes 2011 *(Continued)*

- High risk
- Clear definition. Reporting of timing inconsistent (4 vs. 10 years).

'Other prognostic factors (covariates)'

- Low risk

Statistical analysis and reporting

- High risk
- Statistical method appropriate for the data, but discrepancies between text and graphs detected.

Outcome: Progression-free survival

Outcome measurement

- High risk
- Clear definition. Reporting of timing inconsistent (4 vs. 10 years).

'Other prognostic factors (covariates)'

- Low risk

Statistical analysis and reporting

- High risk
- Statistical method appropriate for the data, but discrepancies between text and graphs detected.

Outcome: Adverse events

Not reported

Notes	Conflict of interest • There are no relevant conflicts of interests to disclose. Funding • Not reported

Casasnovas 2019

Study characteristics	
Methods	Secondary citation(s) • Casasnovas 2018 Language of publication • English Study design • Open-label, randomised phase 3 trial Study centre(s) • Multicentre (90 centres) Countries • Belgium, France

Casasnovas 2019 *(Continued)*

	Median follow-up time (range) • 50.4 months (IQR: 42.9-59.3) for all participants, not reported separately for standard treatment group
Participants	Number of included participants • 823 in total • 413 in standard treatment group Inclusion criteria • Age 16-60 years • Newly diagnosed HL • ECOG performance status score < 3 • Minimum life expectancy of 3 months • Ann Arbor disease stage III, IV, or IIB with a mediastinum-to-thorax ratio of 0.33 or greater or extranodal localisation • No previous treatment for HL • Baseline PET (PET0) with at least one hypermetabolic lesion • Negative HIV, hepatitis C virus, and human T-lymphotropic serology • Normal liver, renal, and haematological functions except for abnormalities related to HL Exclusion criteria • Nodular lymphocyte predominant subtype • Severe cardiopulmonary or metabolic disease Consent • Written, informed consent Recruitment period • 19 May 2011 to 29 April, 2014 Age (range, in years) • 31 (IQR; ranges 23 - 41) Ethnic group(s) • Not reported Stages of disease • II to IV, with B symptoms Comorbidities • Not reported Therapy regimen • Standard treatment group: 4 cycles of $BEACOPP_{escalated}$, irrespective of PET2 result. After PET4: If PET4-negative: 2 further cycles of $BEACOPP_{escalated}$, if PET4-positive: salvage therapy.
Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Whole-body PET scan (groin to head)

Casasnovas 2019 *(Continued)*

Timing of prognostic factor measurement

- 2 to 4 weeks after completion of cycles 2 and 4 of chemotherapy

Method for measurement (use of specific scale and cut-off)

- Deauville criteria, with scores 1 to 3 considered negative, and scores 4 or 5 considered positive; Independent central review by 3 expert reviewers, final decision was based on at least two concordant responses

Was the same definition and method for measurement used in all participants?

- Yes; participants were scanned on the same camera for all PET scans

Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)?

- Presumably yes, but not explicitly mentioned

Outcome(s)

Primary outcome(s) and definition(s)

- Progression-free survival (PFS), defined as the time from randomisation to first progression, relapse, or death from any cause or last follow-up

Secondary outcome(s) and definition(s)

- Safety, not defined
- Overall response, not defined
- Event-free survival, defined as the time from randomisation to the first documented disease progression, relapse, start of a new anti-lymphoma therapy, death from any cause, or last follow-up
- Disease-free survival, defined as the time that complete response was recorded to the date of first documented disease progression, relapse or death related to lymphoma, toxicity from the study treatment (including treatment-related secondary cancer), unknown cause or last follow-up
- Overall survival, defined as the time from randomisation to death from any cause or last follow-up

Timing of outcome measurement

- PFS: at 5 years

Was the same definition and method for measurement used in all participants?

- Yes

Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)?

- Not reported

Missing data

Participants with any missing value?

- Yes; N = 11 stopped treatment before PET2, and further N = 14 stopped treatment before PET4

If yes, how were missing data handled?

- All 413 participants included in ITT analysis, N = 412 included in safety analysis, N = 372 included in per-protocol analysis

Analysis

Univariable analysis: Yes

Total number of participants included in univariable analysis for each outcome

- PFS, OS: N = 413 in ITT analysis, N = 372 in per-protocol analysis

Statistical method

- Kaplan-Meier (survival analysis)

Casasnovas 2019 *(Continued)*

- Log-rank test
- Cox proportional hazard regression models

How was the prognostic factor treated?

- Binary

Multivariable analysis: Yes

Total number of participants included in multivariable analysis for each outcome

- PFS: 759 (all participants that had reviewed PET2 and PET4 scans; not reported separately for standard group after PET2 without treatment modification)
- OS: not reported

Statistical method

- Cox proportional hazards regression model

How was the prognostic factor treated?

- Binary

Number of candidate covariates

- 8

List of all candidate covariates

- PET assessment (PET2 and PET4)
- Sex
- Age
- Eastern Cooperative Oncology Group score
- B symptoms
- Ann Arbor disease stage
- Bulky disease
- International Prognosis Score

Risk of bias (QUIPS)

Study participation

- Low risk
- Adequate description of study population and recruitment. Detailed inclusion criteria.

Study attrition

- Low risk
- Reasons for loss to follow-up provided for most participants with missing data.

Prognostic factor measurement

- Low risk
- Adequate measurement and description. Prognostic factor measured the same way for all participants.

Outcome: Overall survival

Outcome measurement

- Low risk
- Clear definition. Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

Casasnovas 2019 *(Continued)*

- Low risk
- Only advanced stages included.

Statistical analysis and reporting

- Low risk
- Statistical methods appropriate and analysis fully reported.

Outcome: Progression-free survival

Outcome measurement

- Low risk
- Clear definition. Outcome determined based on investigator assessment.

'Other prognostic factors (covariates)'

- Low risk
- Only advanced stages included. Multivariable analysis conducted.

Statistical analysis and reporting

- Low risk
- Statistical methods appropriate and analysis fully reported.

Outcome: Adverse events

Not reported

Notes	Conflict of interest • R-OC has received grants, personal fees and non-financial support from Gilead, Roche, and Takeda, personal fees and non-financial support from Bristol-Myers Squibb, Celegne, and Merck Sharpe & Dohme, and personal fees from Abbvie. PB has received personal fees from Bristol-Myers Squibb, Merck Sharpe & Dohme, and Takeda, grants from Takeda Millenium, and non-financial support from Roche. AS has received personal fees from Takeda. EN-V has received personal fees from Keocyt and Sanofi. FM has received personal fees from Celegne, Gilead, Janssen, and Roche/Genentech. RD has received personal fees from Bristol-Myers Squibb, Celegne, Gilead, Janssen, Karyopharm, Roche, Sanofi, and Takeda. MM has received personal fees from Roche China. The other authors declare no competing interests. Funding • Programme Hospitalier de Recherche Clinique

Cerci 2010

Study characteristics	
Methods	Secondary citation(s) • NA Language of publication • English Study design • Prospective, single-centre study

Cerci 2010 *(Continued)*

	Study centre(s) • São Paulo University Clinics Hospital, Brazil Country • Brazil Median follow-up time (range) • 36 months (32-40)
Participants	Number of included participants • 104 Inclusion criteria • Newly diagnosed, biopsy-proven, classic HL Exclusion criteria • Pregnancy Consent • Yes; written Recruitment period • August 2005 to December 2007 Age (range, in years) • 28 (13-82) Ethnic group(s) • Not reported Stages of disease • All stages Comorbidities • Not reported Therapy regimen • ABVD 4-6 cycles (stage I and II), 6-8 cycles (stage III), 8 cycles (stage IV) • Radiation therapy (stage I or II with no adverse risk factors and treated with 4 cycles ABVD; participants with bulky disease)
Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Whole-body PET scan Timing of prognostic factor measurement • After 2 cycles of ABVD, as late as possible within the week before start of cycle 3 Method for measurement (use of specific scale and cut-off)

Cerci 2010 *(Continued)*

	• No specific scale indicated • 2 board-certified nuclear medicine physicians interpreted all scans • PET-negative defined as no pathologic 18F-FDG uptake at any site; PET-positive defined as presence of focal 18F-FDG uptake not attributed to physiologic biodistribution Was the same definition and method for measurement used in all participants? • Yes Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Not reported
Outcome(s)	Primary outcome(s) and definition(s) • 3-year event-free survival (EFS), defined as the time from diagnosis to treatment failure (incomplete response after first-line treatment, progression during therapy, relapse or death) or last follow-up Secondary outcome(s) and definition(s) • 3-year overall survival (OS) Timing of outcome measurement • At 3 years Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • Not applicable
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • EFS: all • OS: all Statistical method • Log-rank (probability of treatment failure) • Kaplan-Meier (survival curves) How was the prognostic factor treated? • Binary **Multivariable analysis: No**
Risk of bias (QUIPS)	Study participation • Low risk

Cerci 2010 *(Continued)*

- Clear description of participants and study characteristics, consecutive sampling and no participants excluded based on interim-PET availability.

Study attrition

- Low risk
- Loss to follow-up reported.

Prognostic factor measurement

- Low risk
- Adequate measurement and description. Prognostic factor measured the same way for all participants.

Outcome: Overall survival

Outcome measurement

- Low risk
- No definition. Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

- High risk
- Univariable analysis for multiple prognostic factors showed significance of factor of interest, but no multivariable analysis performed. Disease stage not accounted for.

Statistical analysis and reporting

- High risk
- Statistical method in univariable analysis appropriate for the data, but no figures, only table with prognostic values, sensitivity and specificity. Discrepancies detected between text and graphs.

Outcome: Event-free survival

Outcome measurement

- Low risk
- Clear definition. Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

- High risk
- Univariable analysis for multiple prognostic factors showed significance of factor of interest, but no multivariable analysis performed. Disease stage not accounted for.

Statistical analysis and reporting

- High risk
- Statistical method in univariable analysis appropriate for the data, but no figures, only table with prognostic values, sensitivity and specificity. Discrepancies detected between text and graphs.

Outcome: Adverse events

Not reported

Notes	Conflict of interest • Not reported Funding • Not reported

Gallamini 2014

Study characteristics	
Methods	Secondary citation(s) • Agostinelli 2016, Biggi 2013, Gallamini 2006, Gallamini 2007 Language of publication • English Study design • Retrospective, international, multi-centre study (17 centres) Study centre(s) • 17 academic institutions worldwide Countries • Various Median follow-up time (range) • 37 months (2-110)
Participants	Number of included participants • 260 Inclusion criteria • HL participants with early stage unfavourable disease (IIA with adverse prognostic factors) or advanced stage disease (IIB – IVB) • Staging with PET-CT at baseline and after 2 courses of ABVD • No change of treatment according to PET2 • Minimum follow-up of 1 year after completion of first treatment Exclusion criteria • Missing CT data, baseline PET, interim PET, CT or PET slices; poor quality PET images; miscellaneous reasons (n=9) Consent • No; due to retrospective study design Recruitment period • January 2002 to December 2009 Age (range, in years) • 37.3 (14-82) Ethnic group(s) • Not reported Stages of disease • Early stage unfavourable (IIA HL with adverse prognostic factors) • Advanced stages (IIB – IVB)

Gallamini 2014 *(Continued)*

	Comorbidities • Not reported Therapy regimen • 4-8 cycles ABVD with or without involved-field radiotherapy or consolidation radiotherapy
Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Not reported Timing of prognostic factor measurement • A median of 12.3 days (range, 7-22) after cycle 2 of ABVD Method for measurement (use of specific scale and cut-off) • Deauville 5-point scoring system • PET negative defined as scores 1-3, PET positive defined as scores 4 or 5 • 6 reviewers interpreted all scans independently Was the same definition and method for measurement used in all participants? • Yes Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Yes
Outcome(s)	Primary outcome(s) and definition(s) • Disease progression, defined as new disease within 6 months of first-line treatment • Relapse, defined as disease occurring 6 months or longer after achieving complete remission • Progression-free survival (PFS), defined as time from diagnosis to either disease progression or relapse, or to death as a result of any cause, whichever occurred first • Overall survival (OS), defined as the time from diagnosis to death from any cause Secondary outcome(s) and definition(s) • Inter-observer agreement using the 5-PS for PET2 interpretation Timing of outcome measurement • At 3 years Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Yes
Missing data	Participants with any missing value? • No If yes, how were missing data handled?

Gallamini 2014 *(Continued)*

- Not applicable

Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • PFS: 260 Statistical method • Kaplan Meier survival curves with Mantel-Haenszel, log-rank, Wilcoxon and Tarone-Ware tests • Univariable regression analyses How was the prognostic factor treated? • Binary **Multivariable analysis: Yes** Total number of participants included in multivariable analysis for each outcome • PFS: 260 Statistical method • Cox proportional hazards regression model How was the prognostic factor treated? • Binary Number of candidate covariates • 9 List of all candidate covariates • Bulky disease • Lymphocyte • Albumin • White blood cells • IPS (0-2 vs. ≥3) • Continued complete remission (CR) vs. no CR • Lactate dehydrogenase • Bone marrow involvement • PET2
Risk of bias (QUIPS)	Study participation • Low risk • Clear description of participants and study characteristics. Study attrition • Low risk • Length of follow-up reported. Prognostic factor measurement • Low risk • Adequate measurement and description. Blinding not reported. **Outcome: Overall survival**

Gallamini 2014 *(Continued)*

Outcome measurement

- Low risk
- Clear definition. Outcome measured the same way for all participants. Blinding not reported.

'Other prognostic factors (covariates)'

- Low risk
- Only unfavourable and advances stages included.

Statistical analysis and reporting

- Low risk
- Statistical method appropriate for the data.

Outcome: Progression-free survival

Outcome measurement

- Low risk
- Clear definition. Outcome measured the same way for all participants. Blinding not reported.

'Other prognostic factors (covariates)'

- Low risk
- Only unfavourable and advances stages included.

Statistical analysis and reporting

- Low risk
- Statistical method appropriate for the data.

Outcome: Adverse events

Not reported

Notes	Conflict of interest • Not reported Funding • Reporting incomplete • The authors would like to thank: ... Keosys company for providing the Positoscope (R) network to distribute images to reviewers.

Gandikota 2015

Study characteristics	
Methods	Secondary citation(s) • NA Language of publication • English Study design • Retrospective study

Gandikota 2015 *(Continued)*

Study centre(s)

- Not reported

Country/Countries

- Not reported

Median follow-up time (range)

- 46 months (24-126)

Participants	Number of included participants • 78 Inclusion criteria • Biopsy-proven, early-stage (IA to IIB) classic HL of any subtype with or without bulky disease • Age > 18 years • Completion of planned ABVD and radiation therapy • At least 24 months of follow-up or until proven relapse if earlier Exclusion criteria • None Consent • No Recruitment period • January 2000 to December 2012 Age (range, in years) • 43 (median; 22-86) Ethnic group(s) • Not reported Stages of disease • Early stages (IA to IIB) Comorbidities • Not reported Therapy regimen • ABVD (number of cycles based on risk factors and institutional guidelines) followed by involved-field or extended-field radiotherapy
Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • PET-CT scan (from base of the skull to upper thigh) Timing of prognostic factor measurement

Gandikota 2015 *(Continued)*

	• After ABVD cycle 2 to 4 or at the end of chemotherapy Method for measurement (use of specific scale and cut-off) • 5-point scale • PET negative defined as a score ≤ 3 • Staff physicians who were unaware of patient outcomes reviewed all scans Was the same definition and method for measurement used in all participants? • Yes Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Yes
Outcome(s)	Primary outcome(s) and definition(s) • Outcomes relevant to this review were not explored in the study Secondary outcome(s) and definition(s) • Not applicable Timing of outcome measurement • Not applicable Was the same definition and method for measurement used in all participants? • Not applicable Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not applicable
Missing data	Participants with any missing value? • Yes: one patient without baseline PET due to pregnancy; one patient without detectable disease on the baseline scan (excision of single site disease) If yes, how were missing data handled? • One patient without detectable disease on the baseline scan did not receive follow-up PET since not considered necessary
Analysis	**Univariable analysis: No** **Multivariable analysis: No**
Risk of bias (QUIPS)	*No risk of bias assessment, since outcomes relevant to this review were not explored in the study.*
Notes	Conflict of interest • The authors made no disclosure. Funding • No specific funding was disclosed.

Hutchings 2005

Study characteristics	
Methods	Secondary citation(s) • NA Language of publication • English Study design • Not reported Study centre(s) • Guy's and St. Thomas' Hospital, London Country • UK Median follow-up time (range) • 40.2 months (6-125)
Participants	Number of included participants • 85 Inclusion criteria • Histologically-confirmed HL • Early interim FDG-PET scans Exclusion criteria • None Consent • Not reported Recruitment period • May 1993 to January 2004 Age (range, in years) • 36.7 (15-73) Ethnic group(s) • Not reported Stages of disease • All stages Comorbidities • Not reported Therapy regimen

Hutchings 2005 *(Continued)*

- According to departmental protocols: mainly ABVD, number of cycles not reported; additional radiotherapy depending on stage and site of HL
- Alternative therapy for participants without satisfactory remission during initial chemotherapy

Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Half-body PET scan (mid-brain to upper thigh) Timing of prognostic factor measurement • After 2 or 3 cycles of chemotherapy, within the second week of the interval between cycles or as late as possible before administration of the next cycle Method for measurement (use of specific scale and cut-off) • No specific scale indicated • 2 experienced nuclear medicine physicians interpreted all scans, differences decided by consensus • PET-negative defined as no evidence of disease; PET-positive defined as increased uptake suspicious for malignant disease; Minimal residual uptake (MRU) defined as low-grade uptake not likely to represent malignancy Was the same definition and method for measurement used in all participants? • Yes Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Not reported
Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), defined as the time from diagnosis to first evidence of progression or relapse, or to disease-related death • Overall survival (OS), defined as the time from diagnosis to death from any cause Secondary outcome(s) and definition(s) • None Timing of outcome measurement • At 2 and 5 years Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • NA
Analysis	**Univariable analysis: Yes**

Hutchings 2005 *(Continued)*

Total number of participants included in univariable analysis for each outcome

- PFS: all
- OS: not reported

Statistical method

- Kaplan-Meier (survival curves)
- Log-rank (differences between groups)
- Proportional hazards Cox regression analysis

How was the prognostic factor treated?

- Binary

Multivariable analysis: Yes

Total number of participants included in multivariable analysis for each outcome

- PFS: all
- OS: none

Statistical method

- Kaplan-Meier (survival curves)
- Log-rank (differences between groups)
- Proportional hazards Cox regression analysis

How was the prognostic factor treated?

- Binary

Number of candidate covariates

- 4

List of all candidate covariates

- Early interim PET
- Ann Arbor stage
- PET-MRU vs. PET-negative
- PET-positive vs. PET-negative

Risk of bias (QUIPS)

Study participation

- Unclear risk
- All eligible participants included. Clear description of participants and study characteristics. No inclusion / exclusion criteria provided. No comparison to baseline population, and no explanation of missing scans provided.

Study attrition

- Moderate risk
- Loss to follow-up (8 participants), but reasons not provided.

Prognostic factor measurement

- Low risk
- Adequate description. PET results separated into negative, positive and low MRU, which sometimes was considered negative (clearly stated in these cases). No clear cut-off in numbers.

Outcome: Overall survival

Hutchings 2005 *(Continued)*

Outcome measurement

- Low risk
- Clear definition. Outcome measured the same way for all participants. Participants lost to follow-up were still included in analysis.

'Other prognostic factors (covariates)'

- High risk
- Disease stage not accounted for.

Statistical analysis and reporting

- Low risk
- Statistical method in univariable analysis appropriate for the data.

Outcome: Progression-free survival

Outcome measurement

- Low risk
- Clear definition. Outcome measured the same way for all participants. Participants lost to follow-up were still included in analysis.

'Other prognostic factors (covariates)'

- Low risk
- Adjusted for disease stage.

Statistical analysis and reporting

- Low risk
- Statistical method appropriate for the data.

Outcome: Adverse events

Not reported

Notes	Conflict of interest • Not reported Funding • Not reported

Hutchings 2006

Study characteristics	
Methods	Secondary citation(s) • NA Language of publication • English Study design • Prospective, multi-centre study (4 centres)

Hutchings 2006 *(Continued)*

	Study centre(s) • Copenhagen University Hospital, Rigshospitalet, Herlev Hospital, Aarhus University Hospital Country • Denmark Median follow-up time (range) • 22.8 months (6.1-40.8)
Participants	Number of included participants • Total: 99 • With Interim-PET: 77 Inclusion criteria • Newly diagnosed HL • Adults (≥18 years of age) Exclusion criteria • Diabetes mellitus • Pregnancy Consent • Yes; written Recruitment period • November 2001 to June 2004 Age (range, in years) • 36.2 (18.6 – 74.0) Ethnic group(s) • Not reported Stages of disease • All stages Comorbidities • Not reported Therapy regimen • Various therapy regimens: ABVD (91%), ABV/MOPP (3%), ABVD/COPP (3%), BEACOPPesc. (3%), PVAG (1%)
Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Half-body PET scan (mid-brain to upper thigh) Timing of prognostic factor measurement

Hutchings 2006 *(Continued)*

- Within the last week before start of cycle 3 (PET2) and before cycle 5 (PET4)

Method for measurement (use of specific scale and cut-off)

- No specific scale indicated
- 2 experienced nuclear medicine physicians interpreted all scans, differences in interpretation decided by consensus
- Definitions for PET-positive and PET–negative not reported

Was the same definition and method for measurement used in all participants?

- Yes

Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)?

- Yes; nuclear medicine physicians were blinded from all clinical information except diagnosis

Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), defined as the time from diagnosis to first evidence of progression or relapse, or to disease-related death • Overall survival (OS), defined as the time from diagnosis to death from any cause Secondary outcome(s) and definition(s) • None Timing of outcome measurement • At 2 years Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Yes; clinicians were blinded from the results of PET
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • Not applicable
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • PFS: all • OS: not reported Statistical method • Kaplan-Meier (survival curves) • Log-rank (differences between groups) • Proportional hazards Cox regression analysis How was the prognostic factor treated? • Binary **Multivariable analysis: Yes**

Hutchings 2006 *(Continued)*

Total number of participants included in multivariable analysis for each outcome

- PFS: all
- OS: not reported

Statistical method

- Kaplan-Meier (survival curves)
- Log-rank (differences between groups)
- Proportional hazards Cox regression analysis

How was the prognostic factor treated?

- Binary

Number of candidate covariates

- 3

List of all candidate covariates

- Interim PET
- Clinical stage
- Extranodal disease

Risk of bias (QUIPS)	Study participation • High risk • Significant number of participants without PET (n = 22 out of total n = 99). Imbalance between groups with or without PET scan regarding stage of disease. Study attrition • Low risk • Lack of compliance in a small number of participants (n = 7 out of n = 99), but not in the subjects included in PET2 analysis. Prognostic factor measurement • Low risk • Adequate measurement and description. Significant number of participants without PET (n = 22 out of n = 99). **Outcome: Overall survival** Outcome measurement • Low risk • Clear definition of outcome. Outcome measured the same way for all participants. 'Other prognostic factors (covariates)' • High risk • Disease stage not accounted for. Statistical analysis and reporting • Low risk • Statistical method appropriate for the data. **Outcome: Progression-free survival** Outcome measurement

Hutchings 2006 *(Continued)*

- Low risk
- Clear definition of outcome. Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

- Low risk
- Adjusted for disease stage.

Statistical analysis and reporting

- Low risk
- Statistical method appropriate for the data.

Outcome: Adverse events

Not reported

Notes	Conflict of interest • The authors have no financial interests in products studied in this work. Funding • Not reported

Hutchings 2014

Study characteristics	
Methods	Secondary citation(s) • NA Language of publication • English Study design • Prospective, multi-centre study Study centre(s) • Not reported Countries • USA, Italy, Poland, Denmark Median follow-up time (range) • 29 months
Participants	Number of included participants • 126* *Potential overlap of Danish participants with those included in Hutchings 2006 Inclusion criteria • Newly diagnosed classic HL

Hutchings 2014 *(Continued)*

Exclusion criteria

- None

Consent

- Yes; written

Recruitment period

- Not reported

Age (range, in years)

- 34.1 (median, 16.8-76.7)

Ethnic group(s)

- Not reported

Stages of disease

- All stages

Comorbidities

- Not reported

Therapy regimen

- Early-stage disease: 2-4 cycles ABVD followed by radiotherapy, or 6 cycles ABVD
- Advanced-stage disease: 6-8 cycles ABVD with or without consolidation radiotherapy, with exceptions (5 Danish participants treated with BEACOPPesc)

Prognostic factor(s)

Prognostic factor(s)

- Interim PET

Definition of prognostic factor(s)

- Whole-body PET scan

Timing of prognostic factor measurement

- Within the last 5 days of cycle 1 (PET1) and cycle 2 (PET2) (US and Italian participants had PET2 only if PET1 was positive)

Method for measurement (use of specific scale and cut-off)

- Deauville 5-point scoring system
- Scores of 1-3 considered negative, scores of 4-5 considered positive
- Baseline interpretation by an expert with access to clinical information, second interpretation by an independent expert from another country blinded to clinical information

Was the same definition and method for measurement used in all participants?

- No; not all participants received PET2

Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)?

- Yes; experts in both stages blinded to clinical outcome, baseline experts also blinded to clinical information

---

Outcome(s)

Primary outcome(s) and definition(s)

- Progression-free survival (PFS), not defined

Hutchings 2014 *(Continued)*

- Overall survival (OS), not defined

Secondary outcome(s) and definition(s)

- None

Timing of outcome measurement

- At 2 and 3 years

Was the same definition and method for measurement used in all participants?

- Not reported; unclear due to multi-national study design

Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)?

- Not reported

Missing data	Participants with any missing value? • No If yes, how were missing data handled? • Not applicable
Analysis	**Univariable analysis: Yes** Total number of participants included in univariate analysis for each outcome • PFS: all • OS: all Statistical method • Kaplan-Meier (survival analysis) • Log-rank (differences between groups) How was the prognostic factor treated? • Binary **Multivariable analysis: Yes** Total number of participants included in multivariable analysis for each outcome • PFS: all • OS: none Statistical method • Kaplan-Meier (survival analysis) • Log-rank (differences between groups) How was the prognostic factor treated? • Binary Number of candidate covariates • 3 List of all candidate covariates • Interim PET (positive or negative)

Hutchings 2014 *(Continued)*

- Extranodal involvement
- Disease stage (early or advanced stage)

Risk of bias (QUIPS)	Study participation • Low risk • Description of participants and study characteristics given. No inclusion and exclusion criteria. Consecutive sampling and no exclusion based on interim PET availability. Detailed description of treatment regimen. Study attrition • Low risk • No loss to follow-up. Prognostic factor measurement • Low risk • Adequate measurement and description. Prognostic factor measured the same way for all participants. **Outcome: Overall survival** Outcome measurement • Low risk • Adequate measurement and description. Prognostic factor measured the same way for all participants. 'Other prognostic factors (covariates)' • High risk • Disease stage not accounted for. Statistical analysis and reporting • Low risk • Statistical method in univariable analysis appropriate for the data. **Outcome: Progression-free survival** Outcome measurement • Low risk • No definition of outcome. Outcome measured the same way for all participants. 'Other prognostic factors (covariates)' • High risk • Disease stage not accounted for. Statistical analysis and reporting • Low risk • Statistical method in univariable analysis appropriate for the data. **Outcome: Adverse events** *Not reported*
Notes	Conflict of interest • The author(s) indicated no potential conflicts of interest.

Hutchings 2014 *(Continued)*

Funding

- Not reported

Kobe 2018

Study characteristics

Methods	Secondary citation(s) • Borchmann 2017 Language of publication • English Study design • Open-label, international, randomised phase 3 trial Study centre(s) • 301 hospitals and private practices in five European countries Countries • Germany, Switzerland, Austria, the Netherlands, Czech Republic Median follow-up time (range) • Not reported for entire study population
Participants	Number of included participants • Total: 2101 • Qualified for randomisation: 1945 Inclusion criteria • Histologically proven primary diagnosis of HL • Advanced stages: stage IIB with one or both of the risk factors large mediastinal mass and extranodal lesions, or stage III or IV • No previous treatment for HL • Age 18-60 years at inclusion • Normal organ function, except for HL-related impairments • Negative HIV test • Negative pregnancy test • Life expectancy > 3 months Exclusion criteria • Incomplete diagnosis of the disease stage • Prior or concurrent disease that prevents treatment according to protocol • HL as part of a composite lymphoma • Prior chemotherapy or radiation • Malignant disease within the last 5 years (exceptions: basalioma, carcinoma in situ of the cervix uteri, completely resected melanoma TNMpT1) • Pregnancy, lactation • Eastern Cooperative Oncology Group (ECOG) performance status > 2

Kobe 2018 *(Continued)*

- Long-term ingestion of corticosteroids or antineoplastic drugs
- Patient's lack of accountability, inability to appreciate the nature, meaning and consequences of the trial and to formulate his/her own wishes correspondingly
- Noncompliance: refusal of blood products during treatment, epilepsy, drug dependency, change of residence to abroad, prior cerebral injury or similar circumstances that appear to make protocol treatment or long-term follow-up impossible
- Antiepileptic treatment
- General intolerance of any protocol medication
- Unsafe contraceptive methods
- Relationship of dependence or employer-employee relationship to the sponsor or the investigator
- Commitment to an institution on judicial or official order
- Participation in another interventional trial that could interact with this trial

Consent

- Yes; written, including consent to participate in the trial and to storage of data and tissue samples

Recruitment period

- 14 May, 2008 to 18 July 2014

Age (range, in years)

- Not reported for entire study population (Borchmann 2017)

Ethnic group(s)

- Not reported

Stages of disease

- Advanced stages: stage III-IV, or stage II with B symptoms and one or both risk factors of large mediastinal mass

Comorbidities

- None, due to exclusion criteria

Therapy regimen

- 6 or 8 cycles of eBEACOPP (standard arm)
- 4 cycles of eBEACOPP or 8 cycles of eBEACOPP with rituximab (experimental arm)

Prognostic factor(s)

Prognostic factor(s)

- Interim PET

Definition of prognostic factor(s)

- Not reported

Timing of prognostic factor measurement

- Between day 17 and day 21 of cycle 2 of chemotherapy

Method for measurement (use of specific scale and cut-off)

- Deauville 5-point scoring system
- PET negative defined as scores 1 or 2, PET positive defined as scores 3 to 5
- A multidisciplinary panel of experts centrally interpreted all scans

Was the same definition and method for measurement used in all participants?

Kobe 2018 *(Continued)*

- Yes

Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)?

- No; assessors who were masked to local findings, centrally reviewed PET-2 and CT scans as well as x-rays and clinical information (Borchmann 2017)

Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), defined as the time from completion of staging until progression, relapse, or death from any cause Secondary outcome(s) and definition(s) • Overall survival (OS), defined as time from completion of staging until death from any cause Timing of outcome measurement • At 3 years Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • Participants with progressive disease, denoted by DS5 (Deauville score 5), were taken off protocol • 505 participants treated before the protocol amendment in June 2011 were excluded from survival analysis If yes, how were missing data handled? • Participants with missing data were excluded from analysis
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • OS: 722 • PFS: 722 Statistical method • Kaplan-Meier (survival analysis) • Cox regression analysis (hazard ratios) How was the prognostic factor treated? • Binary **Multivariable analysis: Yes** Total number of participants included in multivariable analysis for each outcome • OS: 722 • PFS: 722 Statistical method • Kaplan-Meier (survival analysis) • Cox regression analysis (hazard ratios)

Kobe 2018 *(Continued)*

How was the prognostic factor treated?

- Binary

Number of candidate covariates

- 9

List of all candidate covariates

- Clinical stage
- B symptoms
- Large mediastinal mass
- Extra-nodal involvement
- Involvement of 3 or more nodal areas
- Elevated erythrocyte sedimentation rate
- International Prognosis Score
- HL subtype
- PET positivity (DS4 vs. 1-3)

Risk of bias (QUIPS)

Study participation

- Low risk
- Clear description of participants and study characteristics.

Study attrition

- Low risk
- Length of follow-up reported. Exclusion of participants due to safety amendment during the study.

Prognostic factor measurement

- Low risk
- Adequate measurement and description.

Outcome: Overall survival

Outcome measurement

- Low risk
- Clear definition. Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

- Low risk
- Only advanced stages included.

Statistical analysis and reporting

- Low risk
- Statistical method appropriate for the data.

Outcome: Progression-free survival

Outcome measurement

- Low risk
- Clear definition. Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

- Low risk

Kobe 2018 *(Continued)*

- Only advanced stages included.

Statistical analysis and reporting

- Low risk
- Statistical method appropriate for the data.

Outcome: Adverse events

Not reported

Notes

Conflict of interest

- We declare no competing interests.

Funding

- The HD18 trial was funded by the Deutsche Krebshilfe (No. 107957 and 110617) and the Swiss State Secretariat for Education, Research and Innovation (SERI), and supported by Roche Pharma AG (No. ML-21683).

Markova 2012

Study characteristics

Methods

Secondary citation(s)

- Markova 2009

Language of publication

- English

Study design

- Retrospective, single-centre study

Study centre(s)

- Prague, institution not reported

Country

- Czech Republic

Median follow-up time (range)

- 52 months

Participants

Number of included participants

- 69

Inclusion criteria

- Newly diagnosed, histologically proven HL
- Clinical stage IIB with large mediastinal mass and/or extranodal disease, stage III or IV
- Age 18-60 years

Exclusion criteria

- Presence of any concurrent disease precluding protocol treatment

Markova 2012 *(Continued)*

- Composite lymphoma
- Previous malignancy
- Previous chemo- or radiotherapy
- Pregnancy or lactation
- Diabetes mellitus and elevated fasting blood sugar level >130 mg/dl (exclusion from PET)

Consent

- Not reported

Recruitment period

- January 2004 to February 2008

Age (range, in years)

- 30.7 (± 8.4)

Ethnic group(s)

- Not reported

Stages of disease

- IIB to IVB

Comorbidities

- None, due to exclusion criteria

Therapy regimen

- Treatment according to the HD15 trial of the German Hodgkin Study Group (GHSG) randomly assigned to either 8 cycles of BEACOPPescalated, 6 cycles of BEACOPPescalated or 8 cycles of time-condensed BEACOPP14baseline
- Local radiotherapy for participants with partial remission with residual mass ≥2.5cm and positive PET scan after chemotherapy

---

Prognostic factor(s)

Prognostic factor(s)

- Interim PET

Definition of prognostic factor(s)

- Not reported

Timing of prognostic factor measurement

- After cycle 4 of chemotherapy, as close as possible to cycle 5

Method for measurement (use of specific scale and cut-off)

- A local nuclear medicine physician interpreted all interim-PET scans
- PET-positive defined as focal or diffuse uptake above background in a location incompatible with normal anatomy or physiology, without a specific standardised uptake cut-off value; PET-negative defined as no uptake, or increased uptake at the site of residual mass with an intensity lower or equal to the mediastinal blood pool

Was the same definition and method for measurement used in all participants?

- Yes

Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)?

Markova 2012 *(Continued)*

	• Not reported
Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), defined as the time from diagnosis to the first evidence of progression or relapse, or death from any cause Secondary outcome(s) and definition(s) • None Timing of outcome measurement • After cycle 4, 6/8 and 3 months after completion of chemotherapy Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • Not applicable
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • PFS: all Statistical method • Kaplan-Meier (survival analysis) • Log-rank test (comparison between groups) How was the prognostic factor treated? • Binary **Multivariable analysis: No**
Risk of bias (QUIPS)	Study participation • Low risk • All eligible participants included. Clear description of participants and study characteristics. Consecutive sampling. Inclusion and exclusion criteria provided. Study attrition • Low risk • No loss to follow-up. Prognostic factor measurement • Moderate risk • Prognostic factor measured differently: PET4 scans reviewed locally (at the centre) by one physician, whereas PET6/8 assessment included central review.

Markova 2012 *(Continued)*

Outcome: Overall survival

Not reported

Outcome: Progression-free survival

Outcome measurement

- Low risk
- Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

- Low risk
- Only advanced stages included.

Statistical analysis and reporting

- Low risk
- Statistical method in univariable analysis appropriate for the data.

Outcome: Adverse events

Not reported

Notes	Conflict of interest • Not reported Funding • Not reported

Mesguich 2016

Study characteristics	
Methods	Secondary citation(s) • NA Language of publication • English Study design • Retrospective, multi-centre study (2 centres) Study centre(s) • Haut-Lévêque Hospital and Bergonié Institute, Bordeaux, France Country • France Median follow-up time (range) • 58.9 months
Participants	Number of included participants

Mesguich 2016 *(Continued)*

- 76

Inclusion criteria

- Biopsy-proven, classic HL
- Availability of baseline, interim and end-of-treatment PET-CT

Exclusion criteria

- Treatment with chemotherapy different than ABVD
- Planned treatment modification following int-PET results
- End-PET performance > 6 months after end of treatment

Consent

- No; waived because of retrospective design

Recruitment period

- December 2005 to April 2011

Age (range, in years)

- 37 (median; 14-67)

Ethnic group(s)

- Not reported

Stages of disease

- All stages

Comorbidities

- Not reported

Therapy regimen

- Various therapy regimens: 3, 4, 6 or 8 cycles of ABVD with or without radiotherapy

Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Not reported Timing of prognostic factor measurement • After 2, 3 or 4 treatment cycles Method for measurement (use of specific scale and cut-off) • Deauville 5-point scoring system • Consensual reading of two nuclear medicine physicians • Two cut-offs for interim PET positivity tested and compared: either scores 4 to 5 considered PET positive, or scores 3 to 5 considered PET positive Was the same definition and method for measurement used in all participants? • Yes Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)?

Mesguich 2016 *(Continued)*

	• Yes
Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), defined as the time from diagnosis to either failure of first-line treatment, relapse or death Secondary outcome(s) and definition(s) • None Timing of outcome measurement • At 5 years Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • NA
Analysis	**Univariable analysis: Yes** Total number of participants included in univariate analysis for each outcome • PFS: all Statistical method • Kaplan Meier analysis curve • Log-rank test How was the prognostic factor treated? • Binary **Multivariable analysis: Yes** Total number of participants included in multivariable analysis for each outcome • PFS: all Statistical method • Cox proportional hazard models How was the prognostic factor treated? • Binary Number of candidate covariates • 3 List of all candidate covariates

Mesguich 2016 *(Continued)*

- Interim PET
- Disease stage*
- Bulky disease*

*2 separate models, each adjusted for one of the 2 covariates other than interim PET

Risk of bias (QUIPS)	Study participation • Low risk • Clear description of participants and study characteristics. Study attrition • Low risk • No loss to follow-up. Prognostic factor measurement • Low risk • Adequate measurement and description. Prognostic factor measured the same way for all participants. **Outcome: Overall survival** *Not reported* **Outcome: Progression-free survival** Outcome measurement • Low risk • Clear definition. Outcome measured the same way for all participants. Blinding not reported. 'Other prognostic factors (covariates)' • Low risk • Adjusted for disease stage. Statistical analysis and reporting • Low risk • Statistical method appropriate for the data. **Outcome: Adverse events** *Not reported*
Notes	Conflict of interest • None declared. Funding • No funding was sought or received for this study.

Oki 2014

Study characteristics	
Methods	Secondary citation(s)

Oki 2014 *(Continued)*

- NA

Language of publication

- English

Study design

- Retrospective, single-centre study

Study centre(s)

- MD Anderson Cancer Center, Houston, Texas, USA

Country

- USA

Median follow-up time (range)

- 45 months

Participants	Number of included participants • Total: 325 • 229 participants with PET2 analysed • 96 participants with PET3 excluded post-hoc Inclusion criteria • Classic HL • Treatment with ABVD • Availability of interim PET scan Exclusion criteria • Additional treatment (e.g. with brentuximab vedotin or rituximab) except for radiotherapy Consent • Not reported Recruitment period • January 2001 to May 2011 Age (range, in years) • Group I (early-stage non-bulky): 32 (median, 18-77) • Group II (stage II bulky): 36 (20-60) • Group III (advanced stage IPS ≤ 2): 30 (19-79) • Group IV (advanced stage IPS ≥ 3): 49 (19-84) Ethnic group(s) • Not reported Stages of disease • All stages Comorbidities • Not reported

Oki 2014 *(Continued)*

	Therapy regimen • ABVD with or without radiotherapy
Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Not reported Timing of prognostic factor measurement • After 2 or 3 cycles of ABVD Method for measurement (use of specific scale and cut-off) • Deauville 5-point scoring system • Scores of 1-3 considered negative, scores of 4-5 considered positive • Independent assessment by 3 nuclear medicine physicians Was the same definition and method for measurement used in all participants? • No, 10 participants had only PET without CT scan Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Yes
Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), defined as the time from diagnosis to disease progression, relapse or death from any cause Secondary outcome(s) and definition(s) • None Timing of outcome measurement • At 3 years Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • No
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • Not applicable
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • PFS: all Statistical method

Oki 2014 *(Continued)*

- Kaplan-Meier survival curves with log-rank test per subgroup
- Univariable Cox proportional hazard models

How was the prognostic factor treated?

- Binary

Multivariable analysis: No

Risk of bias (QUIPS)

Study participation

- Low risk
- Clear description of participants and study characteristics.

Study attrition

- Low risk
- No loss to follow-up.

Prognostic factor measurement

- Low risk
- Adequate measurement and description. Prognostic factor measured the same way for all participants.

Outcome: Overall survival

Not reported

Outcome: Progression-free survival

Outcome measurement

- Low risk
- Clear definition. Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

- High risk
- Disease stage not accounted for.

Statistical analysis and reporting

- High risk
- Exclusion of participants with PET3 during analysis due to lack of prognostic value. Stratification according to disease stage resulted in small sample sizes per subgroup.

Outcome: Adverse events

Not reported

Notes

Conflict of interest

- No conflict of interest to disclose for the study.

Funding

- Not reported

Okosun 2012

Study characteristics	
Methods	Secondary citation(s) • NA Language of publication • English Study design • Retrospective, multi-centre study (6 centres) Study centre(s) • 6 centres in London, UK Country • UK Median follow-up time (range) • 27 months
Participants	Number of included participants • 23 Inclusion criteria • Newly diagnosed, histologically confirmed classic HL • Advanced stage • HIV positivity Exclusion criteria • None Consent • Not reported Recruitment period • June 2007 to August 2010 Age (range, in years) • 42 (median, 32-60) Ethnic group(s) • Not reported Stages of disease • Advanced stages: stage III –IV or stage IIB with at least one adverse prognostic factor Comorbidities • HIV positive participants only Therapy regimen

Okosun 2012 *(Continued)*

- Treatment for HL: standard ABVD therapy
- Treatment for HIV: HAART (two NRTIs in combination with either a non-NRTI or a boosted protease inhibitor) antiretroviral therapy; G-CSF per centre protocol; prophylaxis for Pneumocystis jiroveci

Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Half-body PET-CT scan Timing of prognostic factor measurement • After 2-3 cycles of ABVD, within the week before start of the next cycle Method for measurement (use of specific scale and cut-off) • Deauville 5-point scoring system • Scores 1-3 considered negative, scores 4-5 considered positive • Assessed at 3 established PET centres by own nuclear medicine physician and central review by nuclear medicine expert Was the same definition and method for measurement used in all participants? • Yes Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Not reported
Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), defined as the time from diagnosis to disease progression or relapse or last follow-up Secondary outcome(s) and definition(s) • Overall survival (OS), defined as the time from diagnosis to death from any cause • Complete remission, defined as the disappearance of all disease manifestations at the end of therapy Timing of outcome measurement • At 2 years Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • Not applicable
Analysis	**Univariable analysis: Yes** Total number of participants included in univariate analysis for each outcome

Okosun 2012 *(Continued)*

- PFS: all
- OS: not applicable, since no participants died

Statistical method

- Kaplan-Meier survival curves with log-rank test

How was the prognostic factor treated?

- Binary

Multivariable analysis: No

Risk of bias (QUIPS)	Study participation • Low risk • Clear description of participants and study characteristics. Three participants did not have a staging PET, no reasons for missing PET provided. Study attrition • Low risk • No loss to follow-up. Length of follow-up reported. Participants without interim PET (n = 11) excluded. Prognostic factor measurement • Low risk • Adequate measurement and description. Prognostic factor measured the same way for all participants. **Outcome: Overall survival** *Not reported* **Outcome: Progression-free survival** Outcome measurement • Low risk • Clear definition. Outcome measured the same way for all participants. 'Other prognostic factors (covariates)' • Low risk • Only unfavourable and advanced stages included. Statistical analysis and reporting • High risk • Small sample size for some events (only two participants with positive interim PET result). **Outcome: Adverse events** *Not reported*
Notes	Conflict of interest • All authors have no conflicts of interest or disclaimers to declare. Funding • Not reported

Orlacchio 2012

Study characteristics	
Methods	Secondary citation(s) • NA Language of publication • English Study design • Retrospective, single-centre study Study centre(s) • Policlinico Universitario TorVergata, Rome, Italy Country • Italy Median follow-up time (range) • Not reported
Participants	Number of included participants • 132 Inclusion criteria • HL diagnosis based on biochemical tests and bone marrow biopsy • PET-MDCT staging examination, interim PET-MDCT and end of treatment PET-MDCT performed Exclusion criteria • None Consent • Not reported Recruitment period • January 2005 to June 2010 Age (range, in years) • 34 (mean, 16-74) Ethnic group(s) • Not reported Stages of disease • All stages Comorbidities • Not reported Therapy regimen

Orlacchio 2012 *(Continued)*

- ABVD dose dependent on disease stage: stages I-IIA 4x ABVD with radiotherapy; stages IIB-IV 6-8x ABVD with radiotherapy

Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • PET scan from pelvis to head Timing of prognostic factor measurement • At the end of the second ABVD cycle Method for measurement (use of specific scale and cut-off) • International Harmonization Project guidelines • Rated by a radiologist and nuclear medicine specialist, confirmation by semi-quantitative analysis Was the same definition and method for measurement used in all participants? • Yes Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Not reported
Outcome(s)	Primary outcome(s) and definition(s) • Complete remission, defined as the disappearance of symptoms and metabolic activity at any nodal or extranodal site with negative bone marrow biopsy • Partial remission, defined as persistence of significant metabolic activity at one site only, with at least 50% reduction in volume of the nodal masses or parenchymal nodular formations and persistence of disease at bone marrow level • Stable disease, defined as unchanged metabolic findings • Disease progression, defined as the appearance of new sites of pathological uptake and as a 50% increase in volume of nodal masses or previously detected parenchymal localisations Secondary outcome(s) and definition(s) • None Timing of outcome measurement • At the end of treatment Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • NA
Analysis	**Univariable analysis: Yes**

Orlacchio 2012 *(Continued)*

	Total number of participants included in univariate analysis for each outcome • Outcomes selected for univariable analysis unclear Statistical method • Sensitivity, specificity, PPV, NPV How was the prognostic factor treated? • Binary **Multivariable analysis: No**
Risk of bias (QUIPS)	*No risk of bias assessment, since outcomes relevant to this review were not explored in this study.*
Notes	Conflict of interest • None Funding • Not reported

Rossi 2014

Study characteristics	
Methods	Secondary citation(s) • NA Language of publication • English Study design • Retrospective, single-centre study Study centre(s) • Hospital of Dijon, France Country • France Median follow-up time (range) • 50 months (22-71)
Participants	Number of included participants • 59 Inclusion criteria • First diagnosis of classic HL Exclusion criteria • Positive serology for HIV

Rossi 2014 *(Continued)*

	Consent • Yes; written informed consent Recruitment period • January 2007 to January 2010 Age (range, in years) • 35.5 (16-76) Ethnic group(s) • Not reported Stages of disease • All stages Comorbidities • Not reported, except for exclusion of HIV positive participants Therapy regimen • Anthracycline-based chemotherapy dependent on disease stage: stages I-II 4-6x chemotherapy with radiotherapy; stages III-IV 8x chemotherapy
Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Whole-body PET-CT scan Timing of prognostic factor measurement • After 2 cycles of chemotherapy Method for measurement (use of specific scale and cut-off) • Deauville 5-point scoring system • Scores 1-3 considered negative, scores 4-5 considered positive • ΔSUVmax (PET0-PET2) dichotomized by applying the ROC approach • Independent review by 2 nuclear medicine physicians Was the same definition and method for measurement used in all participants? • Different scanner used for 4 participants Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Yes
Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), defined as the time from the beginning of treatment until progression, relapse, or death from any cause or the date of last follow-up • Time to progression (TTP), defined as time from the date of the first course of chemotherapy to any treatment failure, including progression, relapse, or death related to lymphoma, or the date of last follow-up (participants with death from other cause were censored at the time of death) Secondary outcome(s) and definition(s)

Rossi 2014 *(Continued)*

	• None Timing of outcome measurement • At 4 years Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • Not applicable
Analysis	**Univariable analysis: Yes** Total number of participants included in univariate analysis for each outcome • PFS: all Statistical method • Kaplan-Meier product limit method with log-rank test How was the prognostic factor treated? • Binary **Multivariable analysis: Yes** Total number of participants included in multivariable analysis for each outcome • PFS: all Statistical method • Cox proportional hazards regression models per outcome How was the prognostic factor treated? • Binary Number of candidate covariates • 2 List of all candidate covariates • ΔSUVmax (PET0-PET2) • International prognosis score (IPS)
Risk of bias (QUIPS)	Study participation • Low risk • Clear description of participants and study characteristics. Study attrition

Rossi 2014 *(Continued)*

- Low risk
- No loss to follow-up.

Prognostic factor measurement

- Low risk
- Adequate measurement and description. Prognostic factor measured the same way for all participants.

Outcome: Overall survival

Not reported

Outcome: Progression-free survival

Outcome measurement

- Low risk
- Clear definition. Outcome measured the same way for all participants. Blinding not reported.

'Other prognostic factors (covariates)'

- High risk
- Disease stage not accounted for.

Statistical analysis and reporting

- Low risk
- Statistical method appropriate for the data.

Outcome: Adverse events

Not reported

Notes	Conflict of interest • The costs of publication of this article were defrayed in part by the payment of page charges. Therefore, and solely to indicate this fact, this article is hereby marked “advertisement” in accordance with 18 USC section 1734. No potential conflict of interest relevant to this article was reported. Funding • Not reported

Simon 2016

Study characteristics	
Methods	Secondary citation(s) • Miltenyi 2015 Language of publication • English Study design • Retrospective study Study centre(s)

Simon 2016 *(Continued)*

- Not reported

Country

- Hungary

Median follow-up time (range)

- 47.52 months (11-80)

Participants	Number of included participants • 121 Inclusion criteria • Newly diagnosed HL • No previous treatment Exclusion criteria • Immunosuppressive medications • Immunodeficiency Consent • No Recruitment period • 2007 to 2013 Age (range, in years) • 36.7 (mean, 17-79) Ethnic group(s) • Not reported Stages of disease • All stages Comorbidities • None due to exclusion criteria Therapy regimen • ABVD dependent on disease stage: 6 or 8 cycles of ABVD, or 4 or 6 cycles of ABVD combined with radiotherapy
Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Not reported Timing of prognostic factor measurement • After cycle 2 of ABVD between days 11 and 14 Method for measurement (use of specific scale and cut-off)

Simon 2016 *(Continued)*

- Deauville 5-point scoring system
- Scores 1-3 considered negative, scores 4-5 considered positive
- Person(s) interpreting the scans not reported

Was the same definition and method for measurement used in all participants?

- Yes

Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)?

- Not reported

Outcome(s)	Primary outcome(s) and definition(s) • Overall survival (OS), not defined • Progression-free survival (PFS), not defined Secondary outcome(s) and definition(s) • None Timing of outcome measurement • At 5 years after diagnosis Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • NA
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • OS: all • PFS: all Statistical method • Kaplan-Meier (survival analysis) • Log-rank test (comparison between groups) • Cox proportional hazard model (effect of variants on survival) How was the prognostic factor treated? • Binary **Multivariable analysis: Yes** Total number of participants included in multivariable analysis for each outcome • OS: all • PFS: all

Simon 2016 *(Continued)*

Statistical method

- Kaplan-Meier (survival analysis)
- Log-rank test (comparison between groups)
- Cox proportional hazard model (effect of variants on survival

How was the prognostic factor treated?

- Binary

Number of candidate covariates

- 8

List of all candidate covariates

- Age
- Disease stage
- Gender
- B symptoms
- Bulky disease
- Treatment
- PET2 positivity
- Lymphocyte/monocyte ratio (LMR)

Risk of bias (QUIPS)

Study participation

- Unclear risk
- Description of participants provided, but no in- and exclusion criteria provided. Not clear how many participants were sampled and included from the baseline sample.

Study attrition

- Low risk
- No dropouts.

Prognostic factor measurement

- Low risk
- Adequate measurement and description. Prognostic factor measured the same way for all participants.

Outcome: Overall survival

Outcome measurement

- Low risk
- Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

- High risk
- Disease stage not accounted for.

Statistical analysis and reporting

- High risk
- Statistical analysis appropriate for the data. All primary outcomes reported, but discrepancies between text and graphs/tables detected.

Outcome: Progression-free survival

Simon 2016 *(Continued)*

Outcome measurement

- Low risk
- Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

- High risk
- Disease stage not accounted for.

Statistical analysis and reporting

- High risk
- Statistical analysis appropriate for the data and all primary outcomes reported. However, discrepancies between text and graphs/tables detected.

Outcome: Adverse events

Not reported

Notes

Conflict of interest

- None of the authors have any competing interest in the manuscript.

Funding

- Not reported

Straus 2011

Study characteristics

Methods

Secondary citation(s)

- Kostakoglu 2012

Language of publication

- English

Study design

- Prospective phase 2, multi-centre (29 centres), clinical trial

Study centre(s)

- 29 Cancer and Leukemia Group B (CALGB) institutions

Country/Countries

- Not reported

Median follow-up time (range)

- Not reported

Participants

Number of included participants

- Total: 99
- With interim-PET: 88

Inclusion criteria

Straus 2011 *(Continued)*

- Previously untreated, histologically confirmed, classic HL with clinical stages I or II, measurable through physical examination or imaging studies

Exclusion criteria

- Bulky disease

Consent

- Yes; written

Recruitment period

- 15 May 2004 to 29 September 2006

Age (range, in years)

- 37 (18-80)

Ethnic group(s)

- Not reported

Stages of disease

- Stages I - IIB

Comorbidities

- Not reported

Therapy regimen

- 6 cycles of AVG administered on days 1 and 15 per cycle

Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Not reported Timing of prognostic factor measurement • 1 to 2 weeks after completion of cycle 2 of AVG Method for measurement (use of specific scale and cut-off) • Visual assessment was performed using International Harmonization Project criteria • Central review by 2 independent reviewers and an adjudicator Was the same definition and method for measurement used in all participants? • Yes Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Yes
Outcome(s)	Primary outcome(s) and definition(s) • Complete response, defined as complete remission or complete remission unconfirmed after 6 cycles of chemotherapy Secondary outcome(s) and definition(s)

Straus 2011 *(Continued)*

- Progression-free survival (PFS), measured from study entry until relapse
- Adverse events (AEs), defined as toxicity including grade 3 or greater myelosuppression

Timing of outcome measurement

- At 3 years

Was the same definition and method for measurement used in all participants?

- Yes

Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)?

- Not reported

Missing data	Participants with any missing value? • No If yes, how were missing data handled? • None
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • Complete response: none • PFS: 88 Statistical method • Kaplan-Meier (survival analysis) • Log-rank test (comparison between groups) How was the prognostic factor treated? • Binary **Multivariable analysis: No**
Risk of bias (QUIPS)	Study participation • Low risk • Clear description of participants and study characteristics. Study attrition • Low risk • Loss to follow-up reported (n = 2). Prognostic factor measurement • Low risk • Adequate measurement and description. PET2 available for n = 88 out of a total of n = 99 participants. **Outcome: Overall survival** *Not reported* **Outcome: Progression-free survival** Outcome measurement • Low risk

Straus 2011 *(Continued)*

- Clear definition. Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

- Low risk
- Only stages I - IIB included.

Statistical analysis and reporting

- Low risk
- Statistical method in univariable analysis appropriate for the data.

Notes

Conflict of interest

- The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked ‘‘advertisement’’ in accordance with 18 USC section 1734.The authors declare no competing financial interests.

Funding

- This work was supported by the National Cancer Institute: CA77651 (D.J.S., H.S.), CA33601 (J.L.J.), CA32291 (A.S.L., G.P.C.), CA77440 (N.L.B.), CA04457 (L.K.), CA77658 (N.C.H., S.-H.J.), CA32291 (R.W.T.), CA47642 (M.E.J.), and CA77597 (B.D.C.). This work was supported in part by the Lymphoma Foundation, Adam Spector Fund for Hodgkin Research, the Ernest & Jeanette Dicker Charitable Foundation, and Mr Daniel Moon and Family (for D.J.S.). This work was also supported by CALGB (National Cancer Institute) with partial support by Eli Lilly and Company. The research for CALGB 50203 was supported in part by grants from the National Cancer Institute (CA31946) to the CALGB (Dr Monica M. Bertagnolli, Chair) and to the CALGB Statistical Center (Dr Stephen George, CA33601). The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute.

Touati 2014

Study characteristics

Methods

Secondary citation(s)

- NA

Language of publication

- English

Study design

- Retrospective, single-centre study

Study centre(s)

- University Hospital of Limoges, France

Country

- France

Median follow-up time (range)

- 65.8 months (2.2-194.5)

Participants

Number of included participants

Touati 2014 *(Continued)*

- Total: 158
- With interim-PET: 68

Inclusion criteria

- Histologically proven, classic HL

Exclusion criteria

- Nodular lymphocyte predominant HL

Consent

- Not reported

Recruitment period

- February 1995 to July 2011

Age (range, in years)

- 38 (16-85)

Ethnic group(s)

- Not reported

Stages of disease

- All stages

Comorbidities

- Not reported

Therapy regimen

- According to the standard of care at the time of diagnosis therapy regimens included ABVD, MOPP/ABV hybrid or BEACOPP; number of cycles not reported

Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Not reported Timing of prognostic factor measurement • After cycle 2 of chemotherapy Method for measurement (use of specific scale and cut-off) • Visual evaluation • PET-positive if focal or diffuse accumulation of FDG in lesions higher than in surrounding tissue • FDG-PET-CT data (2005 and later) retrospectively reinterpreted using the Deauville 5-point scoring system Was the same definition and method for measurement used in all participants? • Different PET imaging techniques over time (dual-head coincidence until 2005, then FDG-PET-CT), quality assurance and quality control program to ensure comparability of methods Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)?

Touati 2014 *(Continued)*

	• Not reported
Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), defined as time from date of diagnosis until relapse or death • Overall survival (OS), defined as time from first day of diagnosis until death from any cause Secondary outcome(s) and definition(s) • None Timing of outcome measurement • At 5 years Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • Not applicable
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • PFS: 68 • OS: 68 Statistical method • Kaplan-Meier (survival analysis) • Chi-squared test or t-test (differences between groups) • ANOVA (comparison of means) How was the prognostic factor treated? • Binary **Multivariable analysis: No**
Risk of bias (QUIPS)	Study participation • Unclear risk • Availability of interim PET as part of inclusion criteria, but not clear why less than 50% of participants had interim PET data. No comparison of baseline study sample (n = 357) with included participants (n = 158). Study attrition • Low risk • All participants with available interim PET included. Prognostic factor measurement

Touati 2014 *(Continued)*

- Moderate risk
- Retrospective reinterpretation of PET scans using the Deauville criteria. Method described, but unclear whether assessors were blinded to initial interpretation.

Outcome: Overall survival

Outcome measurement

- Low risk
- Clear definition. Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

- High risk
- Disease stage not accounted for.

Statistical analysis and reporting

- Low risk
- Statistical method appropriate for the data.

Outcome: Progression-free survival

Outcome measurement

- Low risk
- Clear definition. Outcome measured the same way for all participants.

'Other prognostic factors (covariates)'

- High risk
- Disease stage not accounted for.

Statistical analysis and reporting

- Low risk
- Statistical method appropriate for the data.

Outcome: Adverse events

Not reported

Notes	Conflict of interest • Not reported Funding • This work was supported by the University Hospital of Limoges, CHU Limoges, F-87042 France.

Ying 2014

Study characteristics	
Methods	Secondary citation(s) • NA Language of publication • Chinese, translated to English

Ying 2014 *(Continued)*

	Study design • Retrospective study Study centre(s) • Peking University Cancer Hospital Country • People's Republic of China Median follow-up time (range) • 29.4 months (12.2-52.4)* *For the whole population (n = 50), but only 35 participants underwent interim PET
Participants	Number of included participants • Total: 50 • With interim PET: 35 Inclusion criteria • Newly diagnosed HL according to the 2008 WHO Hematopoietic and Lymphoid Tissue Classification Exclusion criteria • Not reported Consent • Not reported Recruitment period • September 2009 to December 2012 Age (range, in years) • 33 (14-74) Ethnic group(s) • Not reported Stages of disease • All stages Comorbidities • Not reported Therapy regimen • ABVD or BEACOPP with or without radiotherapy
Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • From the top of the head to the middle thigh, the entire lower extremity was scanned if necessary

Ying 2014 *(Continued)*

	Timing of prognostic factor measurement • After 2 to 4 cycles of treatment Method for measurement (use of specific scale and cut-off) • Interpretation of scans by 2 experienced PET-CT physicians • Scale and cut-off not reported Was the same definition and method for measurement used in all participants? • Not reported Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)? • Not reported
Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), defined as the interval from diagnosis to first signs of tumour progression, patient death, or end of follow-up Secondary outcome(s) and definition(s) • None Timing of outcome measurement • At 3 years Was the same definition and method for measurement used in all participants? • Unclear, follow-up was conducted via telephone and/or outpatient visits Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Not reported
Missing data	Participants with any missing value? • Only 35/50 participants underwent interim PET If yes, how were missing data handled? • Not reported
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • PFS: 35 Statistical method • Kaplan-Meier curves and life tables (survival analysis) • Log-rank tests (comparison between groups) How was the prognostic factor treated? • Binary **Multivariable analysis: No**
Risk of bias (QUIPS)	Study participation • Low risk

Ying 2014 *(Continued)*

- Clear description of participants and study characteristics.

Study attrition

- Low risk
- No loss to follow-up. Length of follow-up reported.

Prognostic factor measurement

- Moderate risk
- Adequate measurement and description, but no standardised criteria for PET scan evaluation.

Outcome: Overall survival

Not reported

Outcome: Progression-free survival

Outcome measurement

- Low risk
- Clear definition. Outcome assessed differently for some participants (via telephone and/or outpatient visits).

'Other prognostic factors (covariates)'

- High risk
- Disease stage not accounted for.

Statistical analysis and reporting

- High risk
- Poor reporting of univariable analysis.

Outcome: Adverse events

Not reported

Notes	Translated from Chinese to English by Yu-Tian Xiao. Conflict of interest • Not reported Funding • This study was funded by Natural Science Foundation of China (NSFC) grant no. 81470328 and Youth Fund of NSFC grant no. 81600162, 81600130.

Zaucha 2017

Study characteristics	
Methods	Secondary citation(s) • NA Language of publication • English Study design

Zaucha 2017 *(Continued)*

- Prospective, observational, multi-centre study (11 centres)

Study centre(s)

- 11 haemato-oncology centres

Country

- Poland

Median follow-up time (range)

- 44.7 months (12.7–90.2)*

*Data for surviving participants only

Participants	Number of included participants • 310 registered participants, out of which 24 were excluded from analysis due to treatment intensification based on PET1 and/or clinical symptoms of active HL Inclusion criteria • Newly diagnosed with classic HL Exclusion criteria • Absent/poor-quality PET-CT images Consent • Yes; written informed consent Recruitment period • January 2008 to October 2014 Age (range, in years) • 30.8 (median, 18–80) Ethnic group(s) • Not reported Stages of disease • All stages Comorbidities • Not reported Therapy regimen • ABVD dependent on disease stage: stages I-IIA 2-4x ABVD with radiotherapy or 6x ABVD; stages IIB-IV 6-8x ABVD with or without radiotherapy
Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Whole-body scan (mandibular angle to one third upper femur) Timing of prognostic factor measurement

Zaucha 2017 *(Continued)*

- 11-13 days after end of ABVD cycle 1 (PET1)
- Additional scan after ABVD cycle 2 for participants with a PET1 score of 3-5 (PET2)

Method for measurement (use of specific scale and cut-off)

- Deauville 5-point scoring system
- Scores 1-3 considered negative, scores 4-5 considered positive
- 6 reviewers interpreted all scans using the blinded independent central review method, disagreements were resolved in a joint session

Was the same definition and method for measurement used in all participants?

- No; PET2 only administered to participants with a PET1 score of 3-5

Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)?

- Yes

Outcome(s)	Primary outcome(s) and definition(s) • Progression-free survival (PFS), not defined Secondary outcome(s) and definition(s) • Kinetics of response Timing of outcome measurement • At 3 years Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • Yes
Missing data	Participants with any missing value? • Yes; only 198 participants had PET2 scans If yes, how were missing data handled? • Not reported
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • 286 (PET1) / 198 (PET2) Statistical method • Kaplan-Meier (survival analysis) • Cox proportional hazard regression analysis (HR between treatment groups) How was the prognostic factor treated? • Binary **Multivariable analysis: No**
Risk of bias (QUIPS)	Study participation

Zaucha 2017 *(Continued)*

- Low risk
- Clear description of participants and study characteristics.

Study attrition

- Low risk

Prognostic factor measurement

- Moderate risk
- Adequate measurement and description. Prognostic factor measured the same way for all participants. However, while PET1 scans were available for all participants, the availability of PET2 scans was dependent on the result of PET1. No further scans were performed if PET1 was negative

Outcome: Overall survival

Not reported as a primary endpoint in the publication. IPD data were available and used to calculate the HR and SE for this outcome.

Outcome: Progression-free survival

Outcome measurement

- High risk
- No definition of outcome.

'Other prognostic factors (covariates)'

- High risk
- Disease stage not accounted for.

Statistical analysis and reporting

- High risk
- No detailed description of analysis.

Outcome: Adverse events

Not reported

Notes

Conflict of interest

- The authors have declared no conflicts of interest.

Funding

- No funders to report.

Zinzani 2012

Study characteristics

Methods

Secondary citation(s)

- Zinzani 2006

Language of publication

- English

Study design

Zinzani 2012 *(Continued)*

- Retrospective, multi-centre study (2 centres)

Study centre(s)

- Bologna and Florence, Italy

Country

- Italy

Median follow-up time (range)

- 45 months (6-100)

Participants	Number of included participants • 304 Inclusion criteria • Diagnosed with HL Exclusion criteria • Other treatment regimens than ABVD • Secondary lymphomas • Continuation of therapy during data analysis Consent • Yes; written informed consent Recruitment period • June 1997 to June 2009 Age (range, in years) • 32 (13-78) Ethnic group(s) • Not reported Stages of disease • All stages Comorbidities • Assessed, but not reported Therapy regimen • ABVD dependent on disease stage: early stages 6x ABVD or 4x ABVD with radiotherapy; advanced stages 6x ABVD
Prognostic factor(s)	Prognostic factor(s) • Interim PET Definition of prognostic factor(s) • Not reported Timing of prognostic factor measurement

Zinzani 2012 *(Continued)*

- After cycle 2 of ABVD

Method for measurement (use of specific scale and cut-off)

- Juweid criteria
- PET positive considered if focal FDG uptake that could not be attributed to physiological biodistribution, benign uptake or normal anatomy, with clearly increased activity relative to the background, excluding participants with minimal residual uptake
- 2 experienced board-certified nuclear medicine physicians interpreted all scans

Was the same definition and method for measurement used in all participants?

- Yes

Were prognostic factor(s) assessed blinded for outcome(s), and for each other (if relevant)?

- No

Outcome(s)	Primary outcome(s) and definition(s) • Response at the end of first-line treatment and at follow-up Secondary outcome(s) and definition(s) • Progression-free survival (PFS), defined as time from diagnosis to first observation of progressive disease or death from any cause • Overall survival (OS), defined as time from diagnosis to time of most recent visit or death Timing of outcome measurement • At 9 years (for PFS and OS) Was the same definition and method for measurement used in all participants? • Yes Was/were outcome(s) assessed blinded for prognostic factor(s), and for each other (if relevant)? • No
Missing data	Participants with any missing value? • No If yes, how were missing data handled? • Not applicable
Analysis	**Univariable analysis: Yes** Total number of participants included in univariable analysis for each outcome • PFS: all • OS: all Statistical method • Kaplan-Meier (survival analysis) • Log-rank test (comparison between groups) How was the prognostic factor treated? • Binary

Zinzani 2012 *(Continued)*

	Multivariable analysis: No
Risk of bias (QUIPS)	Study participation • Low risk • Clear description of participants and study characteristics. Study attrition • Low risk • No loss to follow-up. Prognostic factor measurement • Low risk • Adequate measurement and description. Prognostic factor measured the same way for all participants. No blinding of assessors. **Outcome: Overall survival** Outcome measurement • Low risk • Clear definition of outcome. Outcome measured the same way for all participants. 'Other prognostic factors (covariates)' • High risk • Disease stage not accounted for. Statistical analysis and reporting • Low risk • Statistical method appropriate for the data. **Outcome: Progression-free survival** Outcome measurement • Low risk • Clear definition of outcome. Outcome measured the same way for all participants. 'Other prognostic factors (covariates)' • High risk • Disease stage not accounted for. Statistical analysis and reporting • Low risk • Statistical method appropriate for the data. **Outcome: Adverse events** *Not reported*
Notes	Conflict of interest • None Funding • This work was partially supported by BolognAIL (Bologna, Italy).

ABVD: adriamycin/doxorubicin, bleomycin, vinblastine and dacarbazine; **BEACOPP:** bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine and prednisone; **ePET:** early positron emission tomography; **FDG:** [18F]-fluorodeoxy-D-glucose**; HL:** Hodgkin lymphoma; HR: hazard ratio; **IF-RT:** involved-field radiation therapy; **ITT:** intention-to-treat**; IQR:** interquartile range; **NPV:** negative predictive value**; OS:** overall survival; **PET:** positron emission tomography; **PET-CT:** positron emission tomography computed tomography; **PFS:** progression-free survival**; PPV:** positive predictive value.

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Adams 2016	Wrong publication type. Letter to the editor.
Adams 2017	Wrong publication type. Letter to the editor.
Adams 2018	Wrong publication type. Letter to the editor.
Adams 2018a	Wrong publication type. Letter to the editor.
Adams 2018b	Wrong publication type. Letter to the editor.
Adams 2019	Wrong publication type. Letter to the editor.
Advani 2007	Reported only end-of-chemotherapy PET scan results.
Afanasyev 2017	Wrong publication type. Protocol.
Albano 2017	PET-adapted outcomes.
Albano 2018	PET-adapted outcomes. Treatment was modified according to PET2 results.
Altamirano 2008	Wrong study population. Includes non-Hodgkin lymphoma patients.
Ansell 2016	Wrong publication type. Article.
Awan 2013	Wrong patient population. Patients with any lymphoma were included; no separate data for Hodgkin lymphoma patients.
Bar-Shalom 2003	Wrong study design. Comparison FDG PET and 67Ga scintigraphy.
Barrington 2011a	Wrong publication type. Meeting abstract.
Barrington 2017	Wrong publication type. Commentary.
Basu 2009	Wrong publication type. Commentary.
Becherer 2002	Wrong study design. End-of-chemotherapy PET. Includes non-Hodgkin lymphoma.
Bednaruk-Mlynski 2015	Wrong study design. Role of baseline PET/CT.
Biggi 2012	Wrong publication type. Conference abstract.
Biggi 2017	PET-adapted outcomes. Treatment was modified according to PET2 results.
Bishop 2015	Wrong publication type. Commentary.
Bjurberg 2006	Wrong treatment. Retrospective study of patients with residual tumour or suspected relapse after therapy.

Study	Reason for exclusion
Blum 2002	Wrong patient population. Non-Hodgkin lymphoma patients
Bodet-Milin 2008	Wrong patient population. Non-Hodgkin lymphoma patients.
Bodet-Milin 2009	Wrong publication type. Article.
Boisson 2007	Wrong publication type. Article.
Borchmann 2016	Wrong study design. Literature review.
Bucerius 2006	Wrong publication type. Conference abstract.
Carras 2018	PET-adapted outcomes. Treatment was modified according to PET2 results.
Ciammella 2016	PET-adapted outcomes.
Cremerius 1999	Wrong study design. Retrospective study to validate the clinical value of FDG-PET for therapy control.
Cremerius 2001	Wrong patient population. Patients with any lymphoma were included; no separate data for Hodgkin lymphoma patients.
Cuccaro 2016	PET-adapted outcomes. Positive interim PET results led to change in therapy in three patients; data from these patients was not reported separately from the study population in analysis.
D'Urso 2018	Wrong study design. Analysis of metabolic parameters.
Damlaj 2017	PET-adapted outcomes. Treatment was modified according to PET2 results.
Damlaj 2019	Pet-adapted outcomes.
Danilov 2017	PET-adapted outcomes. Treatment was modified according to interim PET results.
Dann 2009	PET-adapted outcomes.
Dann 2010	PET-adapted outcomes. Treatment was modified according to interim PET results.
Dann 2010a	PET-adapted outcomes.Treatment was modified according to PET2 results.
Dann 2012	PET-adapted outcomes.Treatment was modified according to PET2 results.
Dann 2013	PET-adapted outcomes.Treatment was modified according to PET2 results.
Dann 2016	PET-adapted outcomes. Treatment was modified according to PET2 results.
Dann 2017	PET-adapted outcomes. Treatment was modified according to PET2 results.
Dann 2018	Wrong publication type. Response to letter.
deAndres-Galiana 2015	Wrong study design. Prognostic factor identification study.
Diehl 2007	Wrong study design. The aim was to specify the negative predictive value of PET in patients with residual tumour mass after chemotherapy.

Study	Reason for exclusion
El-Galaly 2012	Wrong study design. The study evaluated the utility of PET scans for post-therapy routine surveillance imaging.
Evens 2014	Wrong publication type. Article.
Fanti 2008	Wrong study design. Case study.
Filmont 2003	Wrong patient population. Patients with aggressive lymphoma undergoing salvage therapy.
Fornecker 2017	PET-adapted outcomes. Treatment was modified according to interim PET results.
Freudenberg 2004	Wrong patient population. Patients with any lymphoma were included; no separate data for Hodgkin lymphoma patients.
Friedberg 2002	Wrong study design. The study intended to compare FDG-PET to gallium scintigraphy in the staging and follow-up of newly diagnosed patients with Hodgkin lymphoma.
Friedberg 2004	Wrong study design. The study intended to compare FDG-PET to gallium scintigraphy in the staging and follow-up of newly diagnosed patients with Hodgkin lymphoma.
Front 1999	Wrong treatment. The study investigated the utility of gallium scintigraphy performed early during treatment as a means to predict outcome and optimise treatment in Hodgkin lymphoma patients.
Fruchart 2006	Wrong patient population. Patients with B-cell lymphoma.
Gallamini 2008	Wrong publication type. Article.
Gallamini 2017	PET-adapted outcomes. Treatment was modified according to interim PET results.
Gallamini 2018	PET-adapted outcomes. Treatment was modified according to interim PET results.
Gallamini 2018a	Wrong publication type. Reply to letter.
Gallowitsch 2008	Wrong publication type. Commentary.
Goldschmidt 2011	Wrong patient population. Relapsed, aggressive non-Hodgkin lymphoma.
Greil 2018	PET-adapted outcomes. Treatment was modified according to interim PET results.
Guidez 2016	Wrong publication type. No abstract or full text.
Hagtvedt 2015	Wrong study design. Comparison between FDG-PET and diffusion-weighted magnetic resonance imaging for assessment of early treatment response in lymphoma.
Haioun 2005	Wrong patient population. Patients with any lymphoma were included; no separate data for Hodgkin lymphoma patients.
Hartmann 2012	Wrong study design. The study investigated protein expression patterns in different Hodgkin lymphoma subtypes.
Hartridge-Lambert 2013	Wrong study design. The study evaluated the risk of disease recurrence and the value of radiologic surveillance in patients treated with ABVD alone who achieved a complete remission according to post-treatment PET. PET was not treated as a prognostic factor.
Honda 2014	Wrong patient population. Letter to the editor, presenting the case of one patient with pulmonary Hodgkin lymphoma.

Study	Reason for exclusion
Hueltenschmidt 2001	Baseline and end-of-chemotherapy PET results.
Huic 2006	Wrong treatment. Patients within three months after completion of conventional initial therapy or salvage therapy with high-dose chemotherapy were included in the study population; no subgroup analysis was reported.
Hutchings 2007	End-of-chemotherapy PET.
Iagaru 2008	Wrong patient population. Patients with any lymphoma were included; no separate data for Hodgkin lymphoma patients.
Illidge 2015	PET-adapted outcomes. Commentary on a research news article about PET-adapted treatment in Hodgkin lymphoma patients.
Jerusalem 2003	End-of-chemotherapy PET.
Johnson 2015	PET-adapted outcomes.
Johnson 2016	PET-adapted outcomes.
Kamran 2016	PET-adapted outcomes.
Kamran 2018	PET-adapted outcomes. Treatment was modified according to PET2 results. Data was not reported separately for PET-positive and PET-negative patients.
Kobe 2008	Wrong study design. The study evaluated the negative predictive value of PET scans in advanced-stage Hodgkin lymphoma patients.
Kobe 2014	Wrong study design. The study evaluated how computed tomography might help improve the positive predictive value of PET in identifying potential high-risk patients.
Kostakoglu 2006	Wrong patient population. Patients with either diffuse large cell lymphoma or Hodgkin lymphoma were included; no separate data for Hodgkin lymphoma patients.
Li 2013	Wrong patient population. The study population consisted of patients with mature T-cell and natural killer cell lymphomas.
Lowe 2002	Wrong study design. Commentary.
Milgrom 2017	Wrong study design. The study population consisted mostly of PET-positive patients. The study compared data from PET-positive patients who received salvage chemotherapy or autologous stem cell transplantation with patients who received radiotherapy only.
Mocikova 2010	Wrong study design. The study evaluated the routine use of PET scans in Hodgkin lymphoma patients during follow-up and in cases of suspected relapse.
Mocikova 2011	Wrong treatment. The study evaluated the prognostic significance of pre-transplant PET scans after salvage chemotherapy before autologous stem cell transplant in patients with relapsed or refractory Hodgkin lymphoma.
Molnar 2010	End-of-chemotherapy PET.
Moskowitz 2015	PET-adapted outcomes. Treatment was modified according to interim PET results.
Naumann 2001	End-of-chemotherapy PET.

Study	Reason for exclusion
NCT00784537	PET-adapted outcomes. Treatment was modified according to interim PET results.
NCT00795613	PET-adapted outcomes. Treatment was modified according to interim PET results.
NCT01358747	PET-adapted outcomes. Treatment was modified according to interim PET results.
NCT01652261	PET-adapted outcomes. Study closed due to lack of recruitment.
NCT02292979	Wrong study design.
Nguyen 2017	PET-adapted outcomes. Treatment was modified according to interim PET results.
Panizo 2004	End of chemotherapy PET.
Paolini 2007	PET-adapted outcomes.
Pavlovsky 2019	PET-adapted outcomes.
Pichler 2000	Wrong study design. Comparison of FDG-Hybrid-PET scans.
Reinhardt 2005	Wrong study design. The study evaluated the accuracy of computed tomography and FDG-PET for prediction of progression-free survival of Hodgkin lymphoma and non-Hodgkin lymphoma patients after completion of therapy.
Rigacci 2002	Wrong study design. Letter.
Rigacci 2017	Wrong study design. Letter.
Rubello 2015	Wrong study design. The study evaluated the variability of FDG liver uptake in patients with Hodgkin lymphoma.
Sakr 2017	Wrong study design.
Schot 2007	Wrong treatment. The study population included patients with recurring lymphoma who were treated with second-line chemotherapy followed by autologous stem cell transplantation.
Simontacchi 2015	PET-adapted outcomes. Treatment was modified according to interim PET results.
Slaby 2002	Wrong patient population. Patients with any lymphoma were included; no separate data for Hodgkin lymphoma patients.
Spaepen 2001	Reported only end-of-chemotherapy PET scan results.
Specht 2007	Wrong publication type. Article.
Spinner 2018	Wrong publication type. Article.
Straus 2018	PET-adapted outcomes. Treatment was modified according to interim PET results.
Strigari 2016	Wrong study design. The aim of the study was to present a novel quantitative tool to refine the risk-class assessment of the Deauville criteria.
Sucak 2011	Wrong treatment. The study population included patients with relapsed or refractory lymphoma post-autologous stem cell transplantation.

Study	Reason for exclusion
Tirelli 2015	Wrong publication type. Article.
Tomita 2015	Wrong patient population. The study population consisted of patients with peripheral T cell lymphoma.
Torizuka 2004	PET-adapted outcomes. Includes non-Hodgkin lymphoma patients.
Trotman 2017	PET-adapted outcomes. Treatment was modified according to interim PET results.
Tseng 2012	Wrong patient population. The study population included relapsed patients.
Villa 2018	PET-adapted outcomes. Treatment was modified according to interim PET results.
Weidmann 1999	Wrong patient population. Includes relapsed patients.
Wilson 2018	Wrong publication type. Commentary.
Xie 2018	Wrong publication type. Review.
Yasgur 2015	Wrong publication type. Commentary.
Yoshimi 2008	Wrong treatment. The study population included lymphoma patients with a poor prognosis who had received FDG-PET scans within one month before allogeneic stem cell transplantation.
Zabrocka 2016	Wrong study design. The study evaluated the current usage of PET scans and its clinical usefulness at different points in Hodgkin lymphoma management based on a single-institution experience.
Zaucha 2009	Wrong publication type. Review.
Zinzani 1999	Wrong patient population. Includes non-Hodgkin lymphoma patients.
Zinzani 2002	Wrong patient population. Includes non-Hodgkin lymphoma patients.
Zinzani 2016	PET-adapted outcomes. Treatment was modified according to interim PET results.

ABVD: adriamycin/doxorubicin, bleomycin, vinblastine and dacarbazine; **FDG:** [18F]-fluorodeoxy-D-glucose**; PET:** positron emission tomography; **PET-CT:** positron emission tomography computed tomography.

Characteristics of studies awaiting classification *[ordered by study ID]*

Abramson 2010

Notes	Title: End of treatment but not interim PET scan predicts outcome in non-bulky limited stage Hodgkin lymphoma. (Conference abstract) Aim • To establish the prognostic value of interim PET scans in limited stage patients with non-bulky disease Study design • Retrospective Country/treatment centre(s)

Abramson 2010 *(Continued)*

- USA (Massachusetts General Hospital, Dana-Farber Cancer Institute, Harvard School of Public Health, Boston MA)

Number of included participants

- 96

Inclusion criteria

- Non-bulky limited stage cHL treated at the institutions between 2000 and 2008; Bulk was defined as a mass >=10 cm or >=1/3 of the intrathoracic diameter.

Exclusion criteria

- None

Treatment

- 4 to 6 cycles of ABVD with or without IFRT

Primary outcome measure(s)

- Overall survival
- Progression-free survival

Algrin 2010

Notes

Title: Interim-positron emission tomography with [18F]fluorodeoxyglucose (interim-PET) evaluation in mediastinal lymphoma including Hodgkin lymphoma (HL) and primary mediastinal large B-cell lymphoma (PMBL).

(Conference abstract)

Aim

- To investigate the prognostic value of qualitative and semi-quantitative evaluations of interim-PET in mediastinal lymphoma

Study design

- Retrospective

Country/treatment centre(s)

- Not reported

Number of included participants

- 48

Inclusion criteria

- Previously untreated, age under 60 at diagnosis and at least one interim-PET evaluation available

Exclusion criteria

- Individuals with sub-diaphragmatic or medullar localisations of lymphoma

Treatment

- Not reported

Primary outcome measure(s)

Algrin 2010 *(Continued)*

- Event-free survival

Arce-Calisaya 2013

Notes

Title: Interim FDG PET-CT in Hodgkin's lymphoma - Does binary response assessment criteria have any prognostic value?

(Conference abstract)

Aim

- To evaluate whether binary response assessment criteria (positive or negative) has any prognostic significance after 2 cycles of ABVD therapy

Study design

- Retrospective

Country/treatment centre(s)

- UK

Number of included participants

- 99

Inclusion criteria

- Newly diagnosed adults with advanced-stage HL undergoing baseline and interim (post-2 cycles ABVD) 18F-FDG PET-CT

Exclusion criteria

- None

Treatment

- ABVD

Primary outcome measure(s)

- Recurrence-free survival after 1 year

Baratto 2015

Notes

Title: Interim-PET in Hodgkin lymphoma: Deauville criteria and metabolic parameters as prognostic factors.

(Conference abstract)

Aim

- To explore the prognostic role of i-PET in individuals with HL

Study design

- Retrospective

Country/treatment centre(s)

- Italy

Baratto 2015 *(Continued)*

Number of included participants

- 83

Inclusion criteria

- Newly diagnosed HL, stage I-IV disease

Exclusion criteria

- None

Treatment

- Not reported

Primary outcome measure(s)

- Overall survival
- Disease-free survival

Barna 2011

Notes

Title: Prognostic value of interim 18FDG-PET-CT in patients with Hodgkin's lymphoma using different 5-point visual scales for interpretation.

(Conference abstract)

Aim

- To compare the effect on prognosis of the currently applied MRU definitions

Study design

- Prospective

Country/treatment centre(s)

- Hungary

Number of included participants

- 82

Inclusion criteria

- Newly-diagnosed HL

Exclusion criteria

- None

Treatment

- 6 courses of ABVB/EBVD, additional radiotherapy according to the protocol

Primary outcome measure(s)

- Overall survival
- Progression-free survival

Barrington 2011

Notes	Title: Are the Deauville criteria a reliable tool for assessment of interim PET in Hodgkin lymphoma? (Conference abstract) Aim • To measure agreement between experienced reporters reading interim PET-CT scans from an international cohort of patients according to the Deauville criteria • To measure progression-free survival in advanced HL according to interim PET Study design • Not reported Country/treatment centre(s) • International study Number of included participants • 262 Inclusion criteria • Individuals diagnosed with stage IIB-IV HL Exclusion criteria • None Treatment • ABVD Primary outcome measure(s) • Progression-free survival

Bentur 2017

Notes	Title: The predictive value of interim PET-CT in elderly patients with Hodgkin lymphoma. This is an abstract only and a lot of relevant information is missing. A full-text has not been published yet. It is particularly unclear whether participants have received treatment adaptation based on the interim PET result. Authors need to be contacted for more information. Aim • To evaluate the significance of iPET in elderly individuals with HL Study design • Retrospective study (1998 to 2016) Country/treatment centre(s) • Unclear, multicentre study (5 centres) Number of included participants • 95

Bentur 2017 *(Continued)*

Inclusion criteria

- Individuals diagnosed with HL between 1998 to 2016
- Older adults (>= 60 years)

Exclusion criteria

- Not reported

Treatment

- Fifty-nine participants received first-line treatment with ABVD, in 13 participants chemotherapy was followed by IVRT (treatment unclear for the remaining participants)

Primary outcome measure(s)

- Overall survival
- Progression-free survival
- Time frame: five years

Berenger 2010

Notes

Title: Prognostic value of interim 18F-FDG PET-CT in mediastinal bulky Hodgkin lymphoma.

(Conference abstract)

Aim

- To determine if Negative Predictive Value (NPV) remains high in individuals who present with mediastinal bulky disease

Study design

- Retrospective

Country/treatment centre(s)

- France

Number of included participants

- 38

Inclusion criteria

- Previously untreated individuals with HL, with localiSed mediastinal bulky disease

Exclusion criteria

- None

Treatment

- Chemotherapy with or without additional radiotherapy

Primary outcome measure(s)

- Progression-free survival
- NPV and PPV of iPET

Bhatwadekar 2017

Notes	Title: Excellent outcome in Hodgkin lymphoma with ABVD and CMT: A single-centre retrospective analysis. (Conference abstract) Aim • To evaluate the outcome of individuals with HL receiving ABVD alone or in combination with RT Study design • Retrospective Country/treatment centre(s) • India (Haemato Oncology Care Centre, Vadodara) Number of included participants • 63 Inclusion criteria • Not reported Exclusion criteria • Not reported Treatment • ABVD alone or in combination with RT Primary outcome measure(s) • Overall survival • Progression-free survival

Cimino 2014

Notes	Title: The complementary prognostic role of baseline and interim PET in predicting treatment outcome in advanced-stage Hodgkin lymphoma. (Conference abstract) Aim • To evaluate the contribution of PET combined with computed tomography (PET-CT) and contrast enhanced computed tomography (ceCT) in the staging and in the prognostication of untreated advanced HL Study design • Retrospective Country/treatment centre(s) • Italy, Poland, Denmark (multicentre) Number of included participants • 162

Cimino 2014 *(Continued)*

Inclusion criteria

- Not reported

Exclusion criteria

- Not reported

Treatment

- ABVD with or without RT

Primary outcome measure(s)

- Overall survival
- Event-free survival

Cocorocchio 2009

Notes

Title: Prognostic role of interim 18FDG-PET in Hodgkin lymphoma: A single-center experience.

(Conference abstract)

Aim

- Single-centre experience with using 18FDG-PET as a prognostic factor for long term complete remission (CR)

Study design

- Retrospective

Country/treatment centre(s)

- Italy

Number of included participants

- 65

Inclusion criteria

- Newly diagnosed with HL

Exclusion criteria

- Not reported

Treatment

- VBM or ChlVPP/ABVVP followed by IFRT

Primary outcome measure(s)

- Complete remission
- Freedom from treatment failure

Cocorocchio 2011

Notes	Title: Evaluation of interim 18FDG-PET in advanced Hodgkin lymphoma (HL) patients (PTS) treated with ChlVPP/ABVVP regimen. (Conference abstract) Aim • To evaluate the prognostic value of interim 18 FDG-PET in advanced HL patients treated with intensified ChlVPP/ABVVP Study design • Not reported Country/treatment centre(s) • Italy Number of included participants • 70 Inclusion criteria • Not reported Exclusion criteria • Not reported Treatment • 6 cycles of ChlVPP/ABVVP Primary outcome measure(s) • Overall survival • Freedom from treatment failure

Copeland 2010

Notes	Title: Single institution experience with interim PET evaluation in newly diagnosed CHL receiving ABVD chemotherapy: Need for standardization. (Conference abstract) Aim • To evaluate the use of interim PET for the identification of individuals with classic HL, who are at risk for relapse after first-line therapy Study design • Retrospective Country/treatment centre(s) • USA (MD Anderson Cancer Center, Houston TX) Number of included participants • 57

Copeland 2010 *(Continued)*

Inclusion criteria

- Newly diagnosed cHL

Exclusion criteria

- Not reported

Treatment

- ABVD

Primary outcome measure(s)

- Event-free survival

Cuzzocrea 2015

Notes

Title: The Deauville criteria and metabolic parameters as prognostic factors in interim PET in Hodgkin lymphoma: A single centre experience.

(Conference abstract)

Aim

- To explore the prognostic role of i-PET in individuals with HL

Study design

- Retrospective

Country/treatment centre(s)

- Italy

Number of included participants

- 83

Inclusion criteria

- Newly diagnosed HL, stage I-IV disease

Exclusion criteria

- Not reported

Treatment

- Not reported

Primary outcome measure(s)

- Overall survival
- Disease-free survival

De Rueda 2013

Notes

Title: Prognostic value of 18F-FDG PET-CT in Hodgkin lymphoma.

(Conference abstract)

De Rueda 2013 *(Continued)*

Aim

- To determine the value of 18F-FDG PET-CT after the second and sixth cycle of first line therapy with ABVD or BEACOPP in the outcome of individuals with HL

Study design

- Retrospective, January 2007 to December 2012

Country/treatment centre(s)

- Spain

Number of included participants

- 79

Inclusion criteria

- HL diagnosis

Exclusion criteria

- Not reported

Treatment

- ABVD or BEACOPP

Primary outcome measure(s)

- Progression-free survival

Fabbri 2011

Notes

Title: 'Early FDG-PET' predicts clinical course of Hodgkin's lymphoma although does not correlate with macrophages infiltration in diagnostic specimens.

(Conference abstract)

Aim

- To verify the prognostic role both of "early-FDG PET" and of macrophagic infiltration, and to test if "early-FDG PET" positivity could correlate with high macrophagic infiltration in diagnostic specimens

Study design

- Retrospective, February 2007 to July 2010

Country/treatment centre(s)

- Italy (Siena and Florence haematology departments)

Number of included participants

- 52

Inclusion criteria

- Diagnosed HL

Exclusion criteria

Fabbri 2011 *(Continued)*

- Not reported

Treatment

- 4 to 6 cycles of ABVD with or without IFRT

Primary outcome measure(s)

- Complete remission
- CD68 expression

Fiore 2010

Notes

Title: Early interim FDG-PET during intensified BEACOPP therapy for advanced-stage Hodgkin disease shows a lower positive predictive value than during ABVD.

(Conference abstract)

Aim

- To examine the predictive role on treatment outcome of early interim FDG-PET in individuals with HL, treated with BEACOPP (4 escalated + 4 baseline cycles)

Study design

- Retrospective

Country/treatment centre(s)

- Italy (8 haematological institutions)

Number of included participants

- 44

Inclusion criteria

- Diagnosed HL, advanced stage (IIB to IVB, or IIA with adverse prognostic factors)

Exclusion criteria

- Not reported

Treatment

- BEACOPP

Primary outcome measure(s)

- Complete remission
- Failure-free survival

Gallegos 2012

Notes

Title: The importance of PET-CT as method of evaluation of early response to treatment in HL.

(Conference abstract)

Aim

Gallegos 2012 *(Continued)*

- To assess the importance of PET-CT as method of evaluation of early response to treatment in HL

Study design

- Retrospective, 2002 to 2011

Country/treatment centre(s)

- Spain (The Miguel Servet's Hospital, Zaragoza)

Number of included participants

- 61

Inclusion criteria

- Diagnosed HL, first-line therapy

Exclusion criteria

- Not reported

Treatment

- ABVD or BEACOPP

Primary outcome measure(s)

- Progression-free survival

Hohaus 2015

Notes

Title: The risk of progression of Hodgkin lymphoma in patients with negative interim PET: A role for the number of tumor-infiltrating macrophages (CD68+ cell counts) and B symptoms.

(Conference abstract)

Aim

- To evaluate if integration of the response evaluation with iPET with parameters available at diagnosis could add prognostic information, allowing a better risk-stratification of individuals with HL

Study design

- Retrospective, 2007 to 2014

Country/treatment centre(s)

- Italy (Università Cattolica del Sacro Cuore, Rome)

Number of included participants

- 102

Inclusion criteria

- Diagnosed classic HL

Exclusion criteria

- Not reported

Treatment

Hohaus 2015 *(Continued)*

- ABVD

Primary outcome measure(s)

- Progression-free survival

Hutchings 2010

Notes

Title: correlation of FDG-PET results after one cycle and after two cycles of chemotherapy in Hodgkin lymphoma.

(Conference abstract)

Aim

- To study the correlation of PET results after one cycle and after two cycles of chemotherapy, and to investigate if the high predictive value of PET after two cycles is obtainable already after one cycle of chemotherapy

Study design

- Prospective trial

Country/treatment centre(s)

- Denmark (Copenhagen), USA (New York)

Number of included participants

- 36

Inclusion criteria

- Diagnosed HL

Exclusion criteria

- Not reported

Treatment

- ABVD or BEACOPPesc

Primary outcome measure(s)

- Negative predictive value

Knight-Greenfield 2013

Notes

Title: Interim FDG PET-CT to predict progression-free survival (PFS) better than clinical and baseline metabolic measurements in Hodgkin lymphoma (cHL).

(Conference abstract)

Aim

- To determine the best predictor of PFS among various variables of tumour metabolic measurements at baseline and at interim PET-CT compared to conventional methods in individuals with classic HL

Study design

Knight-Greenfield 2013 *(Continued)*

- Retrospective

Country/treatment centre(s)

- Not reported

Number of included participants

- 58

Inclusion criteria

- Diagnosed classic HL, ABVD therapy, minimal follow-up of 2 years

Exclusion criteria

- Not reported

Treatment

- ABVD

Primary outcome measure(s)

- Progression-free survival

Leontjeva 2016

Notes

Title: Significance of early interim PET results in advanced Hodgkin lymphoma treated intensive program EACOPP-14.

(Conference abstract)

Aim

- To evaluate the use of interim PET to guide treatment in advanced stage individuals with classic HL

Study design

- Not reported, December 2009 to December 2013

Country/treatment centre(s)

- Russia

Number of included participants

- 36

Inclusion criteria

- Newly diagnosed classic HL (stages IIB to IV, or IIA with bulk), adults

Exclusion criteria

- Not reported

Treatment

- EACOPP-14 with or without RT

Primary outcome measure(s)

- Progression-free survival

Luminari 2010

Notes	Title: The use of FDG positron emission tomography (FDG-PET) in patients with Hodgkin lymphoma (HL) in the "real world": A population based study from northern Italy. (Conference abstract) Aim • To assess how FDG-PET is currently used in individuals with HL Study design • Unclear, 2006 to 2008 Country/treatment centre(s) • Italy (Cancer Registries in Modena, Ferrara, Parma and Reggio Emilia) Number of included participants • 136 Inclusion criteria • Diagnosed HL, adults (18 to 75 years), HIV negative Exclusion criteria • None Treatment • Not reported Primary outcome measure(s) • Overall survival • Relapse-free survival • Failure-free survival

Luminari 2011

Notes	Title: Use of 2-[^{18}F]fluoro-2-deoxy-D-glucose positron emission tomography in patients with Hodgkin lymphoma in daily practice: a population-based study from Northern Italy Authors need to be contacted to clarify whether the treatment has been adapted based on the interim PET results. Aim • To investigate how PET is currently used in daily practice and whether results obtained in clinical trials and retrospective series can be generalised to all individuals with HL. Study design • Retrospective Country/treatment centre(s) • Italy • Participants were identified from archives of four population-based Italian cancer registries

Luminari 2011 *(Continued)*

Number of included participants

- Total: 136

Inclusion criteria

- Registration in one of the four population-based Italian cancer registries (Modena, Reggio Emilia, Parma, Ferrara)
- Histologicallyconfirmed diagnosis of HL between 1 January 2006 and 31 December 2008, age between 18 and 75 years, and human immunodeficiency virus (HIV) negativity

Exclusion criteria

- Missing data

Treatment

- N = 116 (85%) participants received ABVD chemotherapy, 11 participants (8%) received intensified regimens such as BEACOPP or COPP/EBV/CAD, six participants (4%) received chemotherapy without anthracycline such as VBM or MOPP, and three participants (3%) received other therapies such as radiotherapy alone

Primary outcome measure(s)

- Failure-free survival
- Overall survival

Medvedovskaya 2016

Notes

Title: The impact of outcome of interim PET-CT on advanced Hodgkin lymphoma treated with EA-COPP-14.

(Conference abstract)

Aim

- To assess the role of interim PET-CT and compare it with PET-CT results after the end of treatment in individuals with advanced stage classic HL

Study design

- Not reported

Country/treatment centre(s)

- Russia

Number of included participants

- 114

Inclusion criteria

- Newly diagnosed classic HL

Exclusion criteria

- None

Treatment

- 6 cycles of EACOPP-14 with or without RT

Medvedovskaya 2016 *(Continued)*

Primary outcome measure(s)

- Complete metabolic response

Molnar 2011

Notes	Title: The value of interim 18F-FDG PET-CT in Hodgkin lymphoma. (Conference abstract) Aim • To summarise our experience with 18F-FDG PET-CT in HL Study design • Retrospective, November 2006 to January 2010 Country/treatment centre(s) • Hungary (National Institute of Oncology, Budapest) Number of included participants • 60 Inclusion criteria • Not reported Exclusion criteria • Not reported Treatment • ABVD or BEACOPPesc, with or without RT Primary outcome measure(s) • Prognostic value

Molnar 2011a

Notes	Title: Interim FDG PET-CT examinations in advanced stage Hodgkin lymphoma. (Conference abstract) Aim • To summarise our experience with 18F-FDG PET-CT in interim staging Study design • Retrospective, November 2007 to January 2010 Country/treatment centre(s) • Hungary (National Institute of Oncology, Budapest) Number of included participants

Molnar 2011a *(Continued)*

- 19

Inclusion criteria

- Not reported

Exclusion criteria

- Not reported

Treatment

- ABVD or BEACOPPesc, with or without RT

Primary outcome measure(s)

- Prognostic value

Moreira 2013

Notes

Title: Prognostic value of interim vs. end-of-treatment PET scan in Hodgkin's lymphoma.

(Confernece abstract)

Aim

- To evaluate the prognostic value of interim PET scan (PET2) and end-of-treatment PET (PET6) in the outcome of individuals with HL

Study design

- Retrospective, January 2004 to December 2011

Country/treatment centre(s)

- Portugal (Porto, single-centre)

Number of included participants

- 261

Inclusion criteria

- Diagnosed HL

Exclusion criteria

- PET-guided treatment adaptation

Treatment

- ABVD, BEACOPPesc or CVP/CEB

Primary outcome measure(s)

- Complete remission
- Overall survival
- Progression-free survival

Perrone 2009

Notes	Title: Role of positron emission tomography (PET) after 2 and 4 courses of chemotherapy in patients with Hodgkin's lymphoma: A single center experience. (Conference abstract) Aim • To investigate the value of PET performed after 2 (PET2) and 4 (PET4) cycles of therapy for the management of patients with HL Study design • Not reported, September 2006 to September 2008 Country/treatment centre(s) • Italy (University of Bari) Number of included participants • 26 Inclusion criteria • Newly diagnosed HL Exclusion criteria • None Treatment • ABVD Primary outcome measure(s) • Complete remission • Partial remission • Progression-free survival

Pophali 2014

Notes	Title: Bulky disease does not adversely affect overall survival in early stage Hodgkin lymphoma: Role of interim PET and possible omission of radiotherapy in select patients. (Conference abstract) Aim • To assess the impact of disease bulk, interim PET and treatment modality on outcomes Study design • Retrospective, 1995 to 2011 Country/treatment centre(s) • USA (Cleveland Clinic, Cleveland OH) Number of included participants • 121

Pophali 2014 *(Continued)*

Inclusion criteria

- Previously untreated HL, early stages (I and II)

Exclusion criteria

- Missing clinical data

Treatment

- ABVD or other chemotherapy (not specified)

Primary outcome measure(s)

- Overall survival
- Progression-free survival

Rusconi 2010

Notes

Title: Baseline and dynamic prognostic factors in newly diagnosed classical Hodgkin's lymphoma.

(Conference abstract)

Aim

- To identify characteristics, both at baseline and during therapy, predictive for survival outcomes in HL

Study design

- Retrospective

Country/treatment centre(s)

- Italy (Niguarda Hospital, Milan)

Number of included participants

- 105

Inclusion criteria

- Diagnosed HL

Exclusion criteria

- None

Treatment

- 3 to 8 cycles of ABVD with or without IFRT

Primary outcome measure(s)

- Overall survival
- Event-free survival
- Relapse-free survival

Spallino 2017

Notes	Title: The Deauville criteria and QPET as prognostic factors in interim PET in adult Hodgkin lymphoma: A single centre experience. (Conference abstract) Aim • To explore the prognostic role of iPET in individuals with HL by correlating Deauville criteria and qPET to DFS and OS Study design • Retrospective Country/treatment centre(s) • Italy Number of included participants • 131 Inclusion criteria • Newly diagnosed HL, disease stages I to IV Exclusion criteria • None Treatment • Not reported Primary outcome measure(s) • Overall survival • Disease-free survival

Yaghmour 2012

Notes	Title: PET-negative at 2, 3 or 4 cycles of ABVD in Hodgkin's lymphoma is still good. (Conference abstract) Aim • To assess the prognostic value of anytime negative PET scan in the course of first line treatment in individuals with HL receiving ABVD Study design • Retrospective Country/treatment centre(s) • USA (Henry Ford Health System, Detroit MI) Number of included participants • 32 Inclusion criteria

Yaghmour 2012 *(Continued)*

- Newly diagnosed HL

Exclusion criteria

- Not reported

Treatment

- ABVD

Primary outcome measure(s)

- Overall survival

Zanoni 2011

Notes

Title: The predictive value of interim PET and immunohistochemical markers in Hodgkin lymphoma (HL).

(Conference abstract)

Aim

- To compare iPET with a series of histological and immunohistochemical parameters obtained on tissue-micro-arrays as possible predictive factors

Study design

- Retrospective

Country/treatment centre(s)

- Italy (Bologna)

Number of included participants

- 209

Inclusion criteria

- Biopsy-proven HL, complete clinical and iPET data

Exclusion criteria

- None

Treatment

- Not reported

Primary outcome measure(s)

- Overall survival
- Progression-free survival

ABVD: adriamycin/doxorubicin, bleomycin, vinblastine and dacarbazine; **BEACOPP:** bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine and prednisone; **EBVD:** epirubicin, bleomycin, vinblastine and dacarbazine; **FDG:** [18F]-fluorodeoxy-D-glucose**; HL:** Hodgkin lymphoma; **IF-RT:** involved-field radiation therapy; **iPET:** interim positron emission tomography**; MOPP:** mustargen, Oncovin, procarbazine and prednisone; **NPV:** negative predictive value**; PET:** positron emission tomography; **PET-CT:** positron emission tomography computed tomography; **RT:** radiotherapy.

Characteristics of ongoing studies *[ordered by study ID]*

NCT00736320

Study name	HD16 for Early Stages - Treatment optimisation trial in the first-line treatment of early stage Hodgkin lymphoma; treatment stratification by means of FDG-PET
Starting date	November 2009
Contact information	Prof. Dr. Andreas Engert, University of Cologne, Germany
Notes	Study design • Randomised clinical trial (phase III) including 1150 participants with HL Country/treatment centre • 1st Department of Medicine, Cologne University Hospital, Cologne, Germany Number of included participants • Total: 1150 Inclusion criteria • Hodgkin lymphoma • Adults (18 to 75 years) • CS I and II without risk factors (large mediastinal mass (> 1/3 of maximum transverse thorax diameter), extranodal involvement, elevated ESR, three or more involved nodal areas) • Written informed consent Exclusion criteria • Leukocytes < 3000/µl • Platelets < 100000/µl • Hodgkin lymphoma as composite lymphoma • Activity index (WHO) > 2 Arms and interventions • Active comparator (A): two cycles ABVD followed by 20 Gy IF-RT, irrespective of FDG-PET results after chemotherapy • Expertimental (B): two cycles ABVD followed by 20 Gy IF-RT if FDG-PET is positive after chemotherapy; 2 cycles ABVD and treatment stop if FDG-PET is negative after chemotherapy Primary outcome measure(s) • Progression-free survival • Time frame: five years Secondary outcome measure(s) • Overall survival • Acute toxicity • Late toxicity • Complete response rate • Time frame: five years Estimated study completion date • May 2020

ESR: erythrocyte sedimentation rate; **FDG:** [18F]-fluorodeoxy-D-glucose; **HL:** Hodgkin lymphoma; **IF-RT** :involved-field radiation therapy; **MDCT:** multi detector computed tomography; **WHO:** World Health Organization

DATA AND ANALYSES

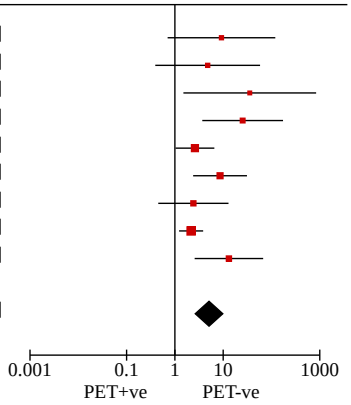
Comparison 1. Univariable comparison of PET+ve vs. PET-ve

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.1 Overall survival	9	1802	Hazard Ratio (IV, Random, 95% CI)	5.09 [2.64, 9.81]
1.2 Progression-free survival	14	2079	Hazard Ratio (IV, Random, 95% CI)	4.90 [3.47, 6.90]

Analysis 1.1. Comparison 1: Univariable comparison of PET+ve vs. PET-ve, Outcome 1: Overall survival

Study or Subgroup	log[Hazard Ratio]	SE	PET+ve Total	PET-ve Total	Weight	Hazard Ratio IV, Random, 95% CI
Barnes 2011	2.220623	1.309795	17	79	5.3%	9.21 [0.71 , 120.03]
Cerci 2010	1.563053	1.274372	30	74	5.6%	4.77 [0.39 , 58.02]
Hutchings 2005	3.570366	1.614442	22	63	3.7%	35.53 [1.50 , 841.02]
Hutchings 2014	3.231135	0.981951	37	89	8.3%	25.31 [3.69 , 173.42]
Kobe 2018	0.9555	0.4724	236	486	18.6%	2.60 [1.03 , 6.56]
Simon 2016	2.153725	0.654523	32	89	13.9%	8.62 [2.39 , 31.08]
Touati 2014	0.881751	0.854282	24	44	10.1%	2.42 [0.45 , 12.89]
Zaucha 2017	0.774	0.2928	24	152	24.1%	2.17 [1.22 , 3.85]
Zinzani 2012	2.575668	0.833487	53	251	10.4%	13.14 [2.57 , 67.31]
Total (95% CI)			**475**	**1327**	**100.0%**	**5.09 [2.64 , 9.81]**

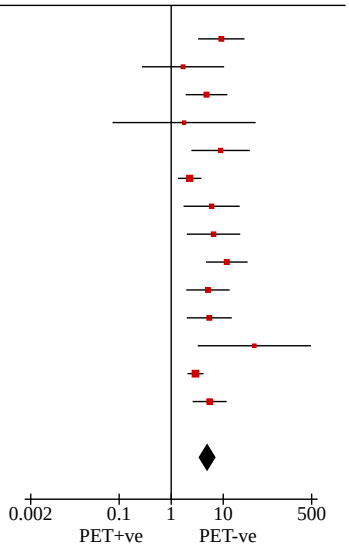
Heterogeneity: Tau² = 0.38; Chi² = 14.39, df = 8 (P = 0.07); I² = 44%
Test for overall effect: Z = 4.87 (P < 0.00001)
Test for subgroup differences: Not applicable



Analysis 1.2. Comparison 1: Univariable comparison of PET+ve vs. PET-ve, Outcome 2: Progression-free survival

Study or Subgroup	log[Hazard Ratio]	SE	PET+ve Total	PET-ve Total	Weight	Hazard Ratio IV, Random, 95% CI
Annunziata 2016	2.2192	0.5231	12	56	7.1%	9.20 [3.30 , 25.65]
Barnes 2011	0.5261	0.9261	17	79	3.0%	1.69 [0.28 , 10.39]
Cerci 2010	1.5624	0.4706	30	74	8.1%	4.77 [1.90 , 12.00]
Hutchings 2005	0.570366	1.6144	22	63	1.1%	1.77 [0.07 , 41.87]
Hutchings 2006	2.187	0.6587	16	61	5.2%	8.91 [2.45 , 32.40]
Kobe 2018	0.8198	0.2651	236	486	13.4%	2.27 [1.35 , 3.82]
Mesguich 2016	1.7891	0.6333	16	60	5.5%	5.98 [1.73 , 20.70]
Rossi 2014	1.873	0.6031	13	46	5.9%	6.51 [2.00 , 21.22]
Simon 2016	2.4596	0.4697	32	89	8.1%	11.70 [4.66 , 29.38]
Straus 2011	1.6268	0.49	24	64	7.7%	5.09 [1.95 , 13.29]
Touati 2014	1.685	0.5075	24	44	7.4%	5.39 [1.99 , 14.58]
Ying 2014	3.6783	1.278	10	25	1.7%	39.58 [3.23 , 484.51]
Zaucha 2017	1.0753	0.1812	24	152	16.0%	2.93 [2.05 , 4.18]
Zinzani 2012	1.6982	0.3845	53	251	10.0%	5.46 [2.57 , 11.61]
Total (95% CI)			**529**	**1550**	**100.0%**	**4.90 [3.47 , 6.90]**

Heterogeneity: Tau² = 0.16; Chi² = 23.52, df = 13 (P = 0.04); I² = 45%
Test for overall effect: Z = 9.07 (P < 0.00001)
Test for subgroup differences: Not applicable



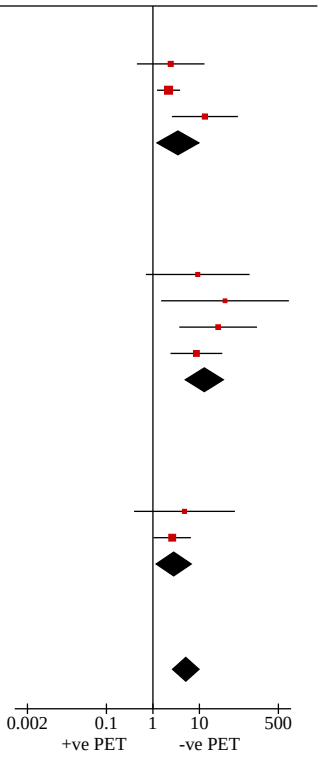
Comparison 2. Subgroups in univariable comparison of OS: PET+ve vs. PET-ve

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
2.1 OS by radiotherapy	9	1802	Hazard Ratio (IV, Random, 95% CI)	5.09 [2.64, 9.81]
2.1.1 Involved node and/or site	3	548	Hazard Ratio (IV, Random, 95% CI)	3.45 [1.22, 9.72]
2.1.2 involved field	4	428	Hazard Ratio (IV, Random, 95% CI)	12.75 [4.98, 32.68]
2.1.3 not specified	2	826	Hazard Ratio (IV, Random, 95% CI)	2.80 [1.17, 6.67]
2.2 OS by study design	8	1717	Hazard Ratio (IV, Random, 95% CI)	4.63 [2.43, 8.80]
2.2.1 Prospective	3	406	Hazard Ratio (IV, Random, 95% CI)	5.35 [1.07, 26.68]
2.2.2 Retrospective	4	589	Hazard Ratio (IV, Random, 95% CI)	7.12 [3.14, 16.14]
2.2.3 RCT	1	722	Hazard Ratio (IV, Random, 95% CI)	2.60 [1.03, 6.56]
2.3 OS by chemotherapy	9	1802	Hazard Ratio (IV, Random, 95% CI)	5.09 [2.64, 9.81]
2.3.1 ABVD	5	801	Hazard Ratio (IV, Random, 95% CI)	5.19 [2.11, 12.72]
2.3.2 ABVD and/or other	3	279	Hazard Ratio (IV, Random, 95% CI)	10.30 [1.71, 62.13]
2.3.3 BEACOPP	1	722	Hazard Ratio (IV, Random, 95% CI)	2.60 [1.03, 6.56]
2.4 OS for PET/CT vs PET	8	1706	Hazard Ratio (IV, Random, 95% CI)	5.01 [2.50, 10.02]
2.4.1 PET/CT	5	595	Hazard Ratio (IV, Random, 95% CI)	4.70 [1.86, 11.86]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
2.4.2 PET only	3	1111	Hazard Ratio (IV, Random, 95% CI)	6.99 [1.58, 30.90]
2.5 OS by disease stage	9	1802	Odds Ratio (IV, Random, 95% CI)	5.09 [2.64, 9.81]
2.5.1 Stages I and II with A and B symptoms	1	96	Odds Ratio (IV, Random, 95% CI)	9.21 [0.71, 120.03]
2.5.2 All stages	7	984	Odds Ratio (IV, Random, 95% CI)	6.28 [2.62, 15.05]
2.5.3 Advanced	1	722	Odds Ratio (IV, Random, 95% CI)	2.60 [1.03, 6.56]
2.6 Timing of interim PET	9	1802	Hazard Ratio (IV, Random, 95% CI)	5.09 [2.64, 9.81]
2.6.1 PET2	6	1495	Hazard Ratio (IV, Random, 95% CI)	3.53 [1.97, 6.32]
2.6.2 Other (including mixed)	3	307	Hazard Ratio (IV, Random, 95% CI)	20.13 [5.04, 80.38]
2.7 OS by HR type of estimation	9	1802	Hazard Ratio (IV, Random, 95% CI)	5.09 [2.64, 9.81]
2.7.1 precise	7	1638	Hazard Ratio (IV, Random, 95% CI)	5.70 [2.60, 12.48]
2.7.2 Imprecise	2	164	Hazard Ratio (IV, Random, 95% CI)	3.60 [0.89, 14.64]

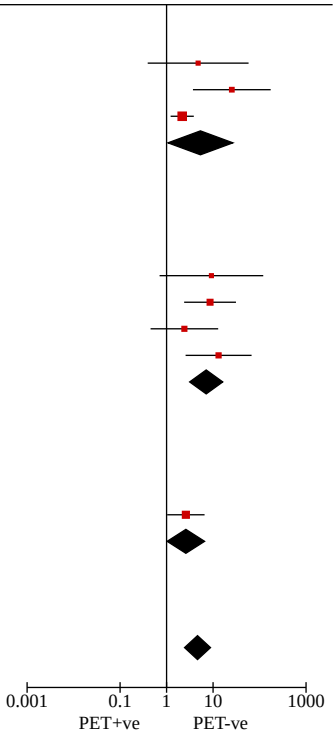
Analysis 2.1. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 1: OS by radiotherapy

Study or Subgroup	log[Hazard Ratio]	SE	Experimental Total	Control Total	Weight	Hazard Ratio IV, Random, 95% CI
2.1.1 Involved node and/or site						
Touati 2014	0.881751	0.8542821	24	44	10.1%	2.42 [0.45 , 12.89]
Zaucha 2017	0.774	0.2928	24	152	24.1%	2.17 [1.22 , 3.85]
Zinzani 2012	2.575668	0.8334869	53	251	10.4%	13.14 [2.57 , 67.31]
Subtotal (95% CI)			**101**	**447**	**44.6%**	**3.45 [1.22 , 9.72]**
Heterogeneity: Tau² = 0.45; Chi² = 4.17, df = 2 (P = 0.12); I² = 52%						
Test for overall effect: Z = 2.34 (P = 0.02)						
2.1.2 involved field						
Barnes 2011	2.220623	1.3097947	17	79	5.3%	9.21 [0.71 , 120.03]
Hutchings 2005	3.570366	1.6144415	22	63	3.7%	35.53 [1.50 , 841.02]
Hutchings 2014	3.231135	0.9819507	37	89	8.3%	25.31 [3.69 , 173.42]
Simon 2016	2.153725	0.6545228	32	89	13.9%	8.62 [2.39 , 31.08]
Subtotal (95% CI)			**108**	**320**	**31.3%**	**12.75 [4.98 , 32.68]**
Heterogeneity: Tau² = 0.00; Chi² = 1.31, df = 3 (P = 0.73); I² = 0%						
Test for overall effect: Z = 5.30 (P < 0.00001)						
2.1.3 not specified						
Cerci 2010	1.563053	1.274372	30	74	5.6%	4.77 [0.39 , 58.02]
Kobe 2018	0.9555	0.4724	236	486	18.6%	2.60 [1.03 , 6.56]
Subtotal (95% CI)			**266**	**560**	**24.2%**	**2.80 [1.17 , 6.67]**
Heterogeneity: Tau² = 0.00; Chi² = 0.20, df = 1 (P = 0.65); I² = 0%						
Test for overall effect: Z = 2.32 (P = 0.02)						
Total (95% CI)			**475**	**1327**	**100.0%**	**5.09 [2.64 , 9.81]**
Heterogeneity: Tau² = 0.38; Chi² = 14.39, df = 8 (P = 0.07); I² = 44%						
Test for overall effect: Z = 4.87 (P < 0.00001)						
Test for subgroup differences: Chi² = 6.01, df = 2 (P = 0.05), I² = 66.7%						

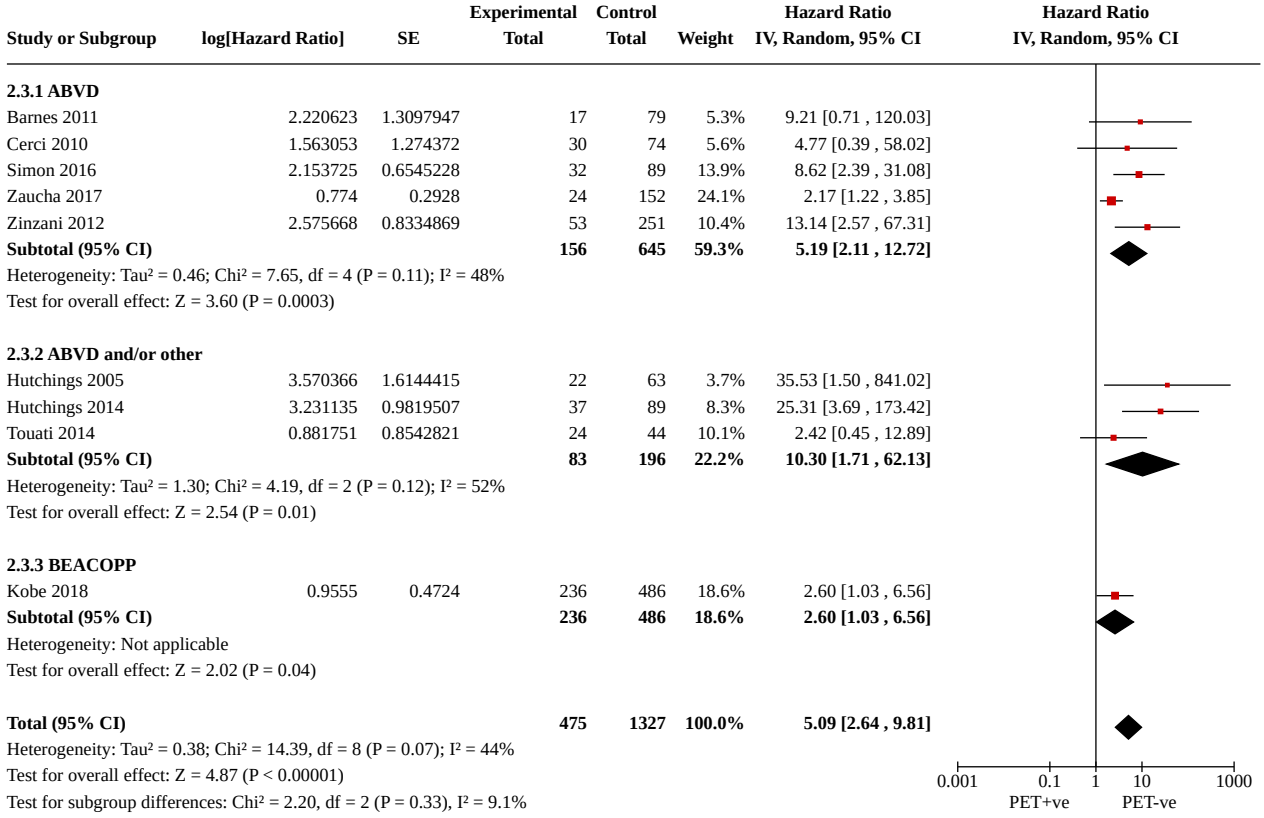


Analysis 2.2. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 2: OS by study design

Study or Subgroup	log[Hazard Ratio]	SE	Experimental Total	Control Total	Weight	Hazard Ratio IV, Random, 95% CI
2.2.1 Prospective						
Cerci 2010	1.563053	1.274372	30	74	5.5%	4.77 [0.39 , 58.02]
Hutchings 2014	3.231135	0.981951	37	89	8.3%	25.31 [3.69 , 173.42]
Zaucha 2017	0.774	0.2928	24	152	26.2%	2.17 [1.22 , 3.85]
Subtotal (95% CI)			**91**	**315**	**40.1%**	**5.35 [1.07 , 26.68]**
Heterogeneity: Tau² = 1.32; Chi² = 5.95, df = 2 (P = 0.05); I² = 66%						
Test for overall effect: Z = 2.04 (P = 0.04)						
2.2.2 Retrospective						
Barnes 2011	2.220623	1.309795	17	79	5.3%	9.21 [0.71 , 120.03]
Simon 2016	2.153725	0.654523	32	89	14.3%	8.62 [2.39 , 31.08]
Touati 2014	0.881751	0.854282	24	44	10.2%	2.42 [0.45 , 12.89]
Zinzani 2012	2.575668	0.833487	53	251	10.5%	13.14 [2.57 , 67.31]
Subtotal (95% CI)			**126**	**463**	**40.3%**	**7.12 [3.14 , 16.14]**
Heterogeneity: Tau² = 0.00; Chi² = 2.27, df = 3 (P = 0.52); I² = 0%						
Test for overall effect: Z = 4.70 (P < 0.00001)						
2.2.3 RCT						
Kobe 2018	0.9555	0.4724	236	486	19.6%	2.60 [1.03 , 6.56]
Subtotal (95% CI)			**236**	**486**	**19.6%**	**2.60 [1.03 , 6.56]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.02 (P = 0.04)						
Total (95% CI)			**453**	**1264**	**100.0%**	**4.63 [2.43 , 8.80]**
Heterogeneity: Tau² = 0.32; Chi² = 12.32, df = 7 (P = 0.09); I² = 43%						
Test for overall effect: Z = 4.67 (P < 0.00001)						
Test for subgroup differences: Chi² = 2.58, df = 2 (P = 0.28), I² = 22.5%						

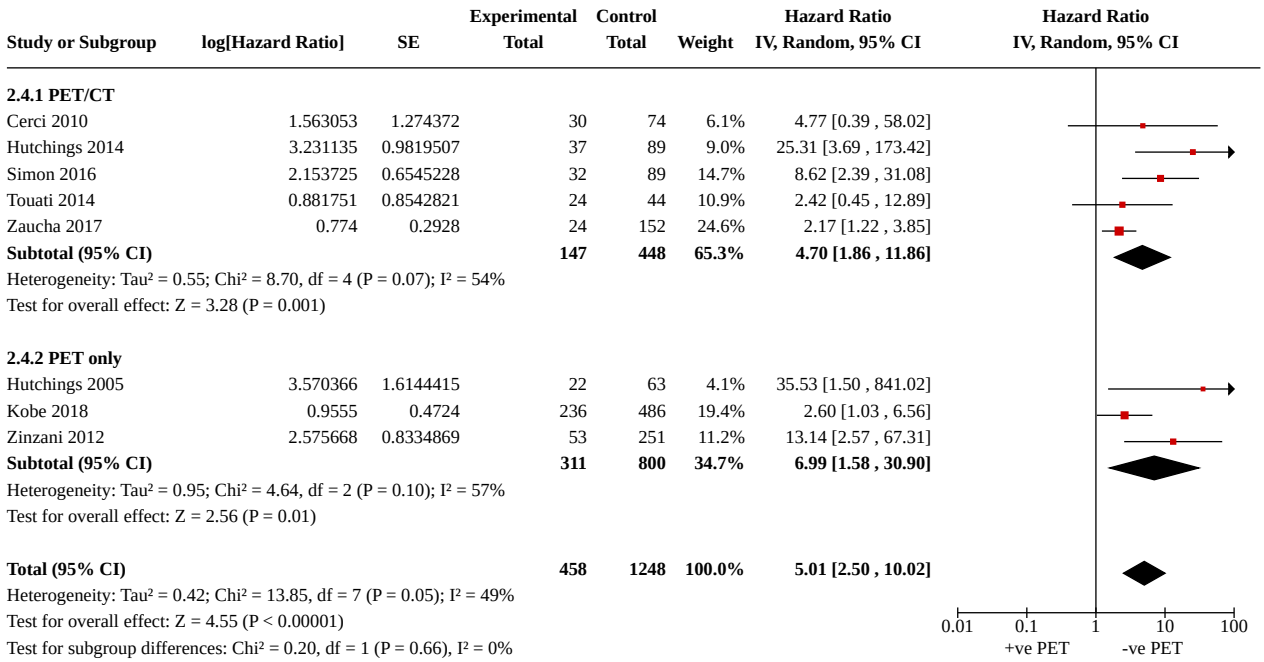


Analysis 2.3. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 3: OS by chemotherapy



Study or Subgroup	log[Hazard Ratio]	SE	Experimental Total	Control Total	Weight	Hazard Ratio IV, Random, 95% CI
2.3.1 ABVD						
Barnes 2011	2.220623	1.3097947	17	79	5.3%	9.21 [0.71 , 120.03]
Cerci 2010	1.563053	1.274372	30	74	5.6%	4.77 [0.39 , 58.02]
Simon 2016	2.153725	0.6545228	32	89	13.9%	8.62 [2.39 , 31.08]
Zaucha 2017	0.774	0.2928	24	152	24.1%	2.17 [1.22 , 3.85]
Zinzani 2012	2.575668	0.8334869	53	251	10.4%	13.14 [2.57 , 67.31]
Subtotal (95% CI)			**156**	**645**	**59.3%**	**5.19 [2.11 , 12.72]**
Heterogeneity: Tau² = 0.46; Chi² = 7.65, df = 4 (P = 0.11); I² = 48%						
Test for overall effect: Z = 3.60 (P = 0.0003)						
2.3.2 ABVD and/or other						
Hutchings 2005	3.570366	1.6144415	22	63	3.7%	35.53 [1.50 , 841.02]
Hutchings 2014	3.231135	0.9819507	37	89	8.3%	25.31 [3.69 , 173.42]
Touati 2014	0.881751	0.8542821	24	44	10.1%	2.42 [0.45 , 12.89]
Subtotal (95% CI)			**83**	**196**	**22.2%**	**10.30 [1.71 , 62.13]**
Heterogeneity: Tau² = 1.30; Chi² = 4.19, df = 2 (P = 0.12); I² = 52%						
Test for overall effect: Z = 2.54 (P = 0.01)						
2.3.3 BEACOPP						
Kobe 2018	0.9555	0.4724	236	486	18.6%	2.60 [1.03 , 6.56]
Subtotal (95% CI)			**236**	**486**	**18.6%**	**2.60 [1.03 , 6.56]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.02 (P = 0.04)						
Total (95% CI)			**475**	**1327**	**100.0%**	**5.09 [2.64 , 9.81]**
Heterogeneity: Tau² = 0.38; Chi² = 14.39, df = 8 (P = 0.07); I² = 44%						
Test for overall effect: Z = 4.87 (P < 0.00001)						
Test for subgroup differences: Chi² = 2.20, df = 2 (P = 0.33), I² = 9.1%						

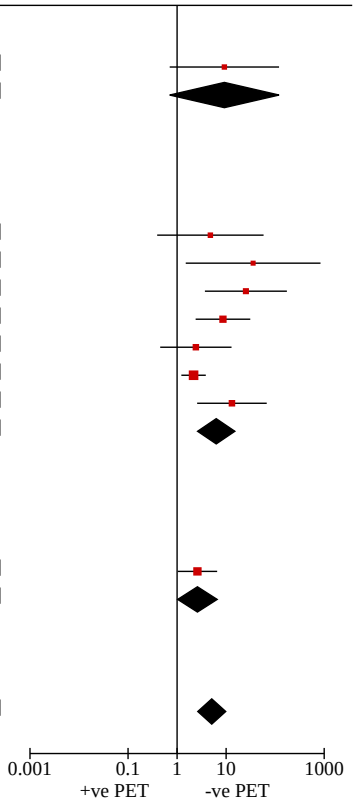
Analysis 2.4. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 4: OS for PET/CT vs PET



Study or Subgroup	log[Hazard Ratio]	SE	Experimental Total	Control Total	Weight	Hazard Ratio IV, Random, 95% CI
2.4.1 PET/CT						
Cerci 2010	1.563053	1.274372	30	74	6.1%	4.77 [0.39 , 58.02]
Hutchings 2014	3.231135	0.9819507	37	89	9.0%	25.31 [3.69 , 173.42]
Simon 2016	2.153725	0.6545228	32	89	14.7%	8.62 [2.39 , 31.08]
Touati 2014	0.881751	0.8542821	24	44	10.9%	2.42 [0.45 , 12.89]
Zaucha 2017	0.774	0.2928	24	152	24.6%	2.17 [1.22 , 3.85]
Subtotal (95% CI)			**147**	**448**	**65.3%**	**4.70 [1.86 , 11.86]**
Heterogeneity: Tau² = 0.55; Chi² = 8.70, df = 4 (P = 0.07); I² = 54%						
Test for overall effect: Z = 3.28 (P = 0.001)						
2.4.2 PET only						
Hutchings 2005	3.570366	1.6144415	22	63	4.1%	35.53 [1.50 , 841.02]
Kobe 2018	0.9555	0.4724	236	486	19.4%	2.60 [1.03 , 6.56]
Zinzani 2012	2.575668	0.8334869	53	251	11.2%	13.14 [2.57 , 67.31]
Subtotal (95% CI)			**311**	**800**	**34.7%**	**6.99 [1.58 , 30.90]**
Heterogeneity: Tau² = 0.95; Chi² = 4.64, df = 2 (P = 0.10); I² = 57%						
Test for overall effect: Z = 2.56 (P = 0.01)						
Total (95% CI)			**458**	**1248**	**100.0%**	**5.01 [2.50 , 10.02]**
Heterogeneity: Tau² = 0.42; Chi² = 13.85, df = 7 (P = 0.05); I² = 49%						
Test for overall effect: Z = 4.55 (P < 0.00001)						
Test for subgroup differences: Chi² = 0.20, df = 1 (P = 0.66), I² = 0%						

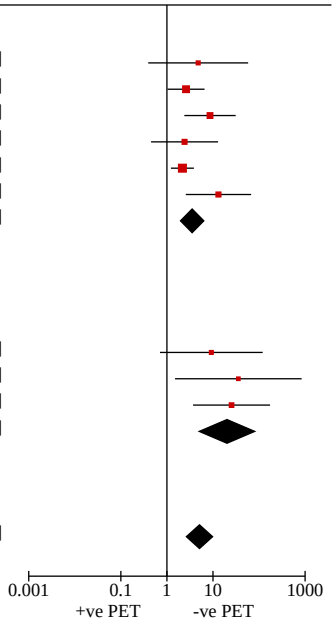
Analysis 2.5. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 5: OS by disease stage

Study or Subgroup	log[OR]	SE	Experimental Total	Control Total	Weight	Odds Ratio IV, Random, 95% CI
2.5.1 Stages I and II with A and B symptoms						
Barnes 2011	2.220623	1.3097947	17	79	5.3%	9.21 [0.71 , 120.03]
Subtotal (95% CI)			**17**	**79**	**5.3%**	**9.21 [0.71 , 120.03]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.70 (P = 0.09)						
2.5.2 All stages						
Cerci 2010	1.563053	1.274372	30	74	5.6%	4.77 [0.39 , 58.02]
Hutchings 2005	3.570366	1.6144415	22	63	3.7%	35.53 [1.50 , 841.02]
Hutchings 2014	3.231135	0.9819507	37	89	8.3%	25.31 [3.69 , 173.42]
Simon 2016	2.153725	0.6545228	32	89	13.9%	8.62 [2.39 , 31.08]
Touati 2014	0.881751	0.8542821	24	44	10.1%	2.42 [0.45 , 12.89]
Zaucha 2017	0.774	0.2928	24	152	24.1%	2.17 [1.22 , 3.85]
Zinzani 2012	2.575668	0.8334869	53	251	10.4%	13.14 [2.57 , 67.31]
Subtotal (95% CI)			**222**	**762**	**76.1%**	**6.28 [2.62 , 15.05]**
Heterogeneity: $Tau^2 = 0.67$; $Chi^2 = 13.40$, df = 6 (P = 0.04); $I^2 = 55\%$						
Test for overall effect: Z = 4.12 (P < 0.0001)						
2.5.3 Advanced						
Kobe 2018	0.9555	0.4724	236	486	18.6%	2.60 [1.03 , 6.56]
Subtotal (95% CI)			**236**	**486**	**18.6%**	**2.60 [1.03 , 6.56]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.02 (P = 0.04)						
Total (95% CI)			**475**	**1327**	**100.0%**	**5.09 [2.64 , 9.81]**
Heterogeneity: $Tau^2 = 0.38$; $Chi^2 = 14.39$, df = 8 (P = 0.07); $I^2 = 44\%$						
Test for overall effect: Z = 4.87 (P < 0.00001)						
Test for subgroup differences: $Chi^2 = 2.20$, df = 2 (P = 0.33), $I^2 = 8.9\%$						



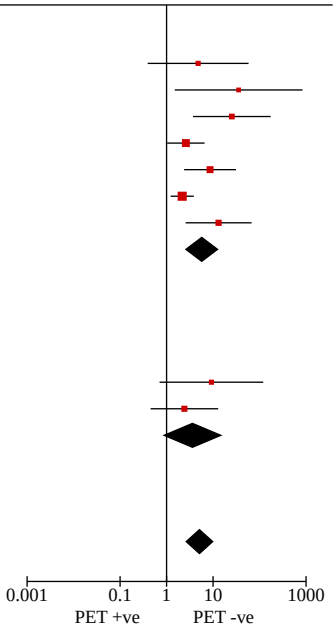
Analysis 2.6. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 6: Timing of interim PET

Study or Subgroup	log[Hazard Ratio]	SE	Experimental Total	Control Total	Weight	Hazard Ratio IV, Random, 95% CI
2.6.1 PET2						
Cerci 2010	1.563053	1.274372	30	74	5.6%	4.77 [0.39 , 58.02]
Kobe 2018	0.9555	0.4724	236	486	18.6%	2.60 [1.03 , 6.56]
Simon 2016	2.153725	0.6545228	32	89	13.9%	8.62 [2.39 , 31.08]
Touati 2014	0.881751	0.8542821	24	44	10.1%	2.42 [0.45 , 12.89]
Zaucha 2017	0.774	0.2928	24	152	24.1%	2.17 [1.22 , 3.85]
Zinzani 2012	2.575668	0.8334869	53	251	10.4%	13.14 [2.57 , 67.31]
Subtotal (95% CI)			**399**	**1096**	**82.6%**	**3.53 [1.97 , 6.32]**
Heterogeneity: Tau² = 0.16; Chi² = 7.26, df = 5 (P = 0.20); I² = 31%						
Test for overall effect: Z = 4.24 (P < 0.0001)						
2.6.2 Other (including mixed)						
Barnes 2011	2.220623	1.3097947	17	79	5.3%	9.21 [0.71 , 120.03]
Hutchings 2005	3.570366	1.6144415	22	63	3.7%	35.53 [1.50 , 841.02]
Hutchings 2014	3.231135	0.9819507	37	89	8.3%	25.31 [3.69 , 173.42]
Subtotal (95% CI)			**76**	**231**	**17.4%**	**20.13 [5.04 , 80.38]**
Heterogeneity: Tau² = 0.00; Chi² = 0.53, df = 2 (P = 0.77); I² = 0%						
Test for overall effect: Z = 4.25 (P < 0.0001)						
Total (95% CI)			**475**	**1327**	**100.0%**	**5.09 [2.64 , 9.81]**
Heterogeneity: Tau² = 0.38; Chi² = 14.39, df = 8 (P = 0.07); I² = 44%						
Test for overall effect: Z = 4.87 (P < 0.00001)						
Test for subgroup differences: Chi² = 5.16, df = 1 (P = 0.02), I² = 80.6%						



Analysis 2.7. Comparison 2: Subgroups in univariable comparison of OS: PET+ve vs. PET-ve, Outcome 7: OS by HR type of estimation

Study or Subgroup	log[Hazard Ratio]	SE	Experimental Total	Control Total	Weight	Hazard Ratio IV, Random, 95% CI
2.7.1 precise						
Cerci 2010	1.563053	1.274372	30	74	5.6%	4.77 [0.39 , 58.02]
Hutchings 2005	3.570366	1.614442	22	63	3.7%	35.53 [1.50 , 841.02]
Hutchings 2014	3.231135	0.981951	37	89	8.3%	25.31 [3.69 , 173.42]
Kobe 2018	0.9555	0.4724	236	486	18.6%	2.60 [1.03 , 6.56]
Simon 2016	2.153725	0.654523	32	89	13.9%	8.62 [2.39 , 31.08]
Zaucha 2017	0.774	0.2928	24	152	24.1%	2.17 [1.22 , 3.85]
Zinzani 2012	2.575668	0.833487	53	251	10.4%	13.14 [2.57 , 67.31]
Subtotal (95% CI)			**434**	**1204**	**84.6%**	**5.70 [2.60 , 12.48]**
Heterogeneity: Tau² = 0.53; Chi² = 13.66, df = 6 (P = 0.03); I² = 56%						
Test for overall effect: Z = 4.35 (P < 0.0001)						
2.7.2 Imprecise						
Barnes 2011	2.220623	1.309795	17	79	5.3%	9.21 [0.71 , 120.03]
Touati 2014	0.881751	0.854282	24	44	10.1%	2.42 [0.45 , 12.89]
Subtotal (95% CI)			**41**	**123**	**15.4%**	**3.60 [0.89 , 14.64]**
Heterogeneity: Tau² = 0.00; Chi² = 0.73, df = 1 (P = 0.39); I² = 0%						
Test for overall effect: Z = 1.79 (P = 0.07)						
Total (95% CI)			**475**	**1327**	**100.0%**	**5.09 [2.64 , 9.81]**
Heterogeneity: Tau² = 0.38; Chi² = 14.39, df = 8 (P = 0.07); I² = 44%						
Test for overall effect: Z = 4.87 (P < 0.00001)						
Test for subgroup differences: Chi² = 0.31, df = 1 (P = 0.58), I² = 0%						



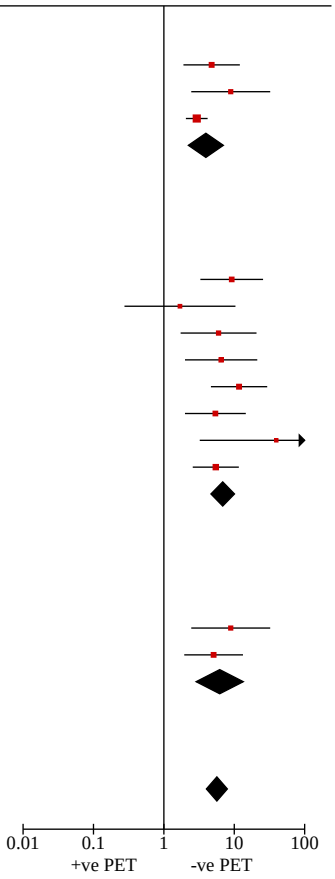
Comparison 3. Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
3.1 PFS by study design	13	1349	Hazard Ratio (IV, Random, 95% CI)	5.66 [4.02, 7.97]
3.1.1 prospective	3	357	Hazard Ratio (IV, Random, 95% CI)	3.95 [2.23, 7.00]
3.1.2 retrospective	8	827	Hazard Ratio (IV, Random, 95% CI)	6.85 [4.66, 10.08]
3.1.3 RCT	2	165	Hazard Ratio (IV, Random, 95% CI)	6.21 [2.87, 13.42]
3.2 PFS by chemotherapy	14	2079	Hazard Ratio (IV, Random, 95% CI)	4.90 [3.47, 6.90]
3.2.1 ABVD	7	945	Hazard Ratio (IV, Random, 95% CI)	5.13 [3.18, 8.27]
3.2.2 ABVD and/or other	4	265	Hazard Ratio (IV, Random, 95% CI)	7.07 [3.40, 14.70]
3.2.3 other NON-ABVD chemo	3	869	Hazard Ratio (IV, Random, 95% CI)	3.64 [1.83, 7.24]
3.3 PFS for PET/CT vs PET	13	1983	Hazard Ratio (IV, Random, 95% CI)	5.08 [3.57, 7.21]
3.3.1 PET/CT	8	707	Hazard Ratio (IV, Random, 95% CI)	6.03 [3.68, 9.90]
3.3.2 PET only	5	1276	Hazard Ratio (IV, Random, 95% CI)	4.06 [2.33, 7.08]
3.4 PFS by disease stage	14	2079	Hazard Ratio (IV, Random, 95% CI)	4.90 [3.47, 6.90]
3.4.1 Stages I and II with A and B symptoms	2	184	Hazard Ratio (IV, Random, 95% CI)	3.88 [1.54, 9.83]
3.4.2 All stages	11	1173	Hazard Ratio (IV, Random, 95% CI)	5.81 [3.93, 8.57]
3.4.3 Advanced	1	722	Hazard Ratio (IV, Random, 95% CI)	2.27 [1.35, 3.82]
3.5 PFS by radiotherapy	14	2079	Hazard Ratio (IV, Random, 95% CI)	4.90 [3.47, 6.90]
3.5.1 Involved node and/or site	5	651	Hazard Ratio (IV, Random, 95% CI)	5.35 [2.94, 9.75]
3.5.2 Involved field	6	514	Hazard Ratio (IV, Random, 95% CI)	7.06 [4.15, 12.00]
3.5.3 Not specified	2	826	Hazard Ratio (IV, Random, 95% CI)	2.97 [1.48, 5.98]
3.5.4 None	1	88	Hazard Ratio (IV, Random, 95% CI)	5.09 [1.95, 13.29]
3.6 Timing of interim PET	14	2079	Hazard Ratio (IV, Random, 95% CI)	4.90 [3.47, 6.90]
3.6.1 PET2	9	1677	Hazard Ratio (IV, Random, 95% CI)	4.68 [3.14, 6.98]
3.6.2 Other (including mixed)	5	402	Hazard Ratio (IV, Random, 95% CI)	6.32 [3.40, 11.75]
3.7 PFS by HR type of estimation	14	2079	Hazard Ratio (IV, Random, 95% CI)	4.90 [3.47, 6.90]

Outcome or subgroup title	No. of studies	No. of partici-pants	Statistical method	Effect size
3.7.1 precise	9	1450	Hazard Ratio (IV, Random, 95% CI)	4.69 [2.84, 7.73]
3.7.2 Imprecise	5	629	Hazard Ratio (IV, Random, 95% CI)	5.66 [3.65, 8.77]

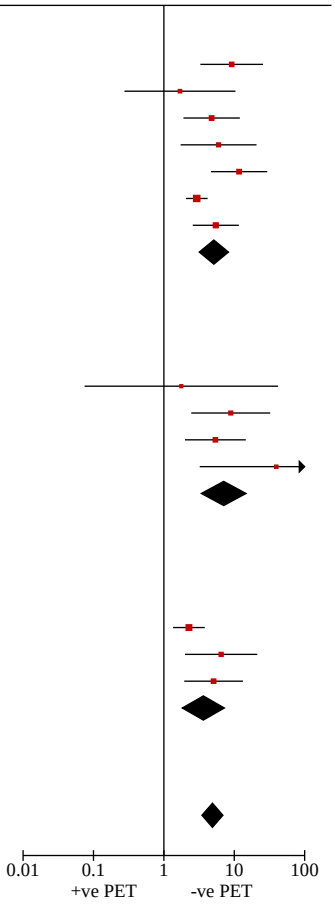
Analysis 3.1. Comparison 3: Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve, Outcome 1: PFS by study design

Study or Subgroup	log[Hazard Ratio]	SE	Experimental Total	Control Total	Weight	Hazard Ratio IV, Random, 95% CI
3.1.1 prospective						
Cerci 2010	1.5624	0.4706	30	74	8.8%	4.77 [1.90 , 12.00]
Hutchings 2006	2.187	0.6587	16	61	5.5%	8.91 [2.45 , 32.40]
Zaucha 2017	1.0753	0.1812	24	152	19.3%	2.93 [2.05 , 4.18]
Subtotal (95% CI)			**70**	**287**	**33.6%**	**3.95 [2.23 , 7.00]**
Heterogeneity: Tau² = 0.11; Chi² = 3.31, df = 2 (P = 0.19); I² = 40%						
Test for overall effect: Z = 4.70 (P < 0.00001)						
3.1.2 retrospective						
Annunziata 2016	2.2192	0.5231	12	56	7.7%	9.20 [3.30 , 25.65]
Barnes 2011	0.5261	0.9261	17	79	3.1%	1.69 [0.28 , 10.39]
Mesguich 2016	1.7891	0.6333	16	60	5.8%	5.98 [1.73 , 20.70]
Rossi 2014	1.873	0.6031	13	46	6.2%	6.51 [2.00 , 21.22]
Simon 2016	2.4596	0.4697	32	89	8.8%	11.70 [4.66 , 29.38]
Touati 2014	1.685	0.5075	24	44	8.0%	5.39 [1.99 , 14.58]
Ying 2014	3.6783	1.278	10	25	1.7%	39.58 [3.23 , 484.51]
Zinzani 2012	1.6982	0.3845	53	251	11.2%	5.46 [2.57 , 11.61]
Subtotal (95% CI)			**177**	**650**	**52.6%**	**6.85 [4.66 , 10.08]**
Heterogeneity: Tau² = 0.00; Chi² = 6.40, df = 7 (P = 0.49); I² = 0%						
Test for overall effect: Z = 9.79 (P < 0.00001)						
3.1.3 RCT						
Kobe 2018	2.187	0.6587	16	61	5.5%	8.91 [2.45 , 32.40]
Straus 2011	1.6268	0.49	24	64	8.4%	5.09 [1.95 , 13.29]
Subtotal (95% CI)			**40**	**125**	**13.8%**	**6.21 [2.87 , 13.42]**
Heterogeneity: Tau² = 0.00; Chi² = 0.47, df = 1 (P = 0.50); I² = 0%						
Test for overall effect: Z = 4.65 (P < 0.00001)						
Total (95% CI)			**287**	**1062**	**100.0%**	**5.66 [4.02 , 7.97]**
Heterogeneity: Tau² = 0.13; Chi² = 18.76, df = 12 (P = 0.09); I² = 36%						
Test for overall effect: Z = 9.92 (P < 0.00001)						
Test for subgroup differences: Chi² = 2.48, df = 2 (P = 0.29), I² = 19.5%						



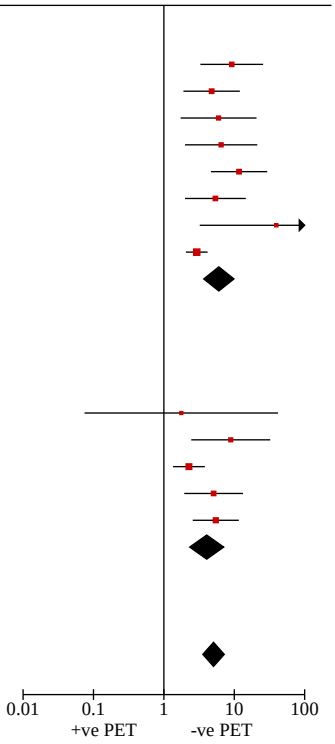
Analysis 3.2. Comparison 3: Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve, Outcome 2: PFS by chemotherapy

Study or Subgroup	log[Hazard Ratio]	SE	Experimental Total	Control Total	Weight	Hazard Ratio IV, Random, 95% CI
3.2.1 ABVD						
Annunziata 2016	2.2192	0.5231	12	56	7.1%	9.20 [3.30 , 25.65]
Barnes 2011	0.5261	0.9261	17	79	3.0%	1.69 [0.28 , 10.39]
Cerci 2010	1.5624	0.4706	30	74	8.1%	4.77 [1.90 , 12.00]
Mesguich 2016	1.7891	0.6333	16	60	5.5%	5.98 [1.73 , 20.70]
Simon 2016	2.4596	0.4697	32	89	8.1%	11.70 [4.66 , 29.38]
Zaucha 2017	1.0753	0.1812	24	152	16.0%	2.93 [2.05 , 4.18]
Zinzani 2012	1.6982	0.3845	53	251	10.0%	5.46 [2.57 , 11.61]
Subtotal (95% CI)			**184**	**761**	**57.7%**	**5.13 [3.18 , 8.27]**
Heterogeneity: Tau² = 0.20; Chi² = 12.73, df = 6 (P = 0.05); I² = 53%						
Test for overall effect: Z = 6.70 (P < 0.00001)						
3.2.2 ABVD and/or other						
Hutchings 2005	0.570366	1.6144	22	63	1.1%	1.77 [0.07 , 41.87]
Hutchings 2006	2.187	0.6587	16	61	5.2%	8.91 [2.45 , 32.40]
Touati 2014	1.685	0.5075	24	44	7.4%	5.39 [1.99 , 14.58]
Ying 2014	3.6783	1.278	10	25	1.7%	39.58 [3.23 , 484.51]
Subtotal (95% CI)			**72**	**193**	**15.4%**	**7.07 [3.40 , 14.70]**
Heterogeneity: Tau² = 0.00; Chi² = 2.96, df = 3 (P = 0.40); I² = 0%						
Test for overall effect: Z = 5.24 (P < 0.00001)						
3.2.3 other NON-ABVD chemo						
Kobe 2018	0.8198	0.2651	236	486	13.4%	2.27 [1.35 , 3.82]
Rossi 2014	1.873	0.6031	13	46	5.9%	6.51 [2.00 , 21.22]
Straus 2011	1.6268	0.49	24	64	7.7%	5.09 [1.95 , 13.29]
Subtotal (95% CI)			**273**	**596**	**26.9%**	**3.64 [1.83 , 7.24]**
Heterogeneity: Tau² = 0.18; Chi² = 3.91, df = 2 (P = 0.14); I² = 49%						
Test for overall effect: Z = 3.68 (P = 0.0002)						
Total (95% CI)			**529**	**1550**	**100.0%**	**4.90 [3.47 , 6.90]**
Heterogeneity: Tau² = 0.16; Chi² = 23.52, df = 13 (P = 0.04); I² = 45%						
Test for overall effect: Z = 9.07 (P < 0.00001)						
Test for subgroup differences: Chi² = 1.69, df = 2 (P = 0.43), I² = 0%						



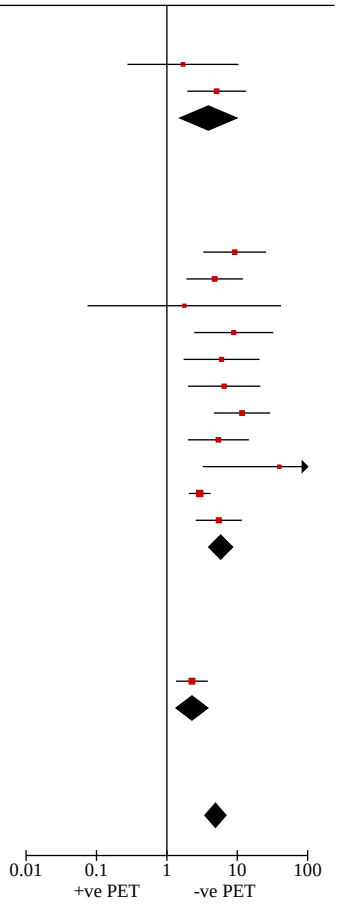
Analysis 3.3. Comparison 3: Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve, Outcome 3: PFS for PET/CT vs PET

Study or Subgroup	log[Hazard Ratio]	SE	Experimental Total	Control Total	Weight	Hazard Ratio IV, Random, 95% CI
3.3.1 PET/CT						
Annunziata 2016	2.2192	0.5231	12	56	7.3%	9.20 [3.30 , 25.65]
Cerci 2010	1.5624	0.4706	30	74	8.3%	4.77 [1.90 , 12.00]
Mesguich 2016	1.7891	0.6333	16	60	5.7%	5.98 [1.73 , 20.70]
Rossi 2014	1.873	0.6031	13	46	6.1%	6.51 [2.00 , 21.22]
Simon 2016	2.4596	0.4697	32	89	8.4%	11.70 [4.66 , 29.38]
Touati 2014	1.685	0.5075	24	44	7.6%	5.39 [1.99 , 14.58]
Ying 2014	3.6783	1.278	10	25	1.8%	39.58 [3.23 , 484.51]
Zaucha 2017	1.0753	0.1812	24	152	16.3%	2.93 [2.05 , 4.18]
Subtotal (95% CI)			**161**	**546**	**61.5%**	**6.03 [3.68 , 9.90]**
Heterogeneity: Tau² = 0.24; Chi² = 15.14, df = 7 (P = 0.03); I² = 54%						
Test for overall effect: Z = 7.12 (P < 0.00001)						
3.3.2 PET only						
Hutchings 2005	0.570366	1.6144	22	63	1.2%	1.77 [0.07 , 41.87]
Hutchings 2006	2.187	0.6587	16	61	5.4%	8.91 [2.45 , 32.40]
Kobe 2018	0.8198	0.2651	236	486	13.7%	2.27 [1.35 , 3.82]
Straus 2011	1.6268	0.49	24	64	7.9%	5.09 [1.95 , 13.29]
Zinzani 2012	1.6982	0.3845	53	251	10.3%	5.46 [2.57 , 11.61]
Subtotal (95% CI)			**351**	**925**	**38.5%**	**4.06 [2.33 , 7.08]**
Heterogeneity: Tau² = 0.15; Chi² = 6.78, df = 4 (P = 0.15); I² = 41%						
Test for overall effect: Z = 4.93 (P < 0.00001)						
Total (95% CI)			**512**	**1471**	**100.0%**	**5.08 [3.57 , 7.21]**
Heterogeneity: Tau² = 0.16; Chi² = 22.65, df = 12 (P = 0.03); I² = 47%						
Test for overall effect: Z = 9.06 (P < 0.00001)						
Test for subgroup differences: Chi² = 1.09, df = 1 (P = 0.30), I² = 8.0%						



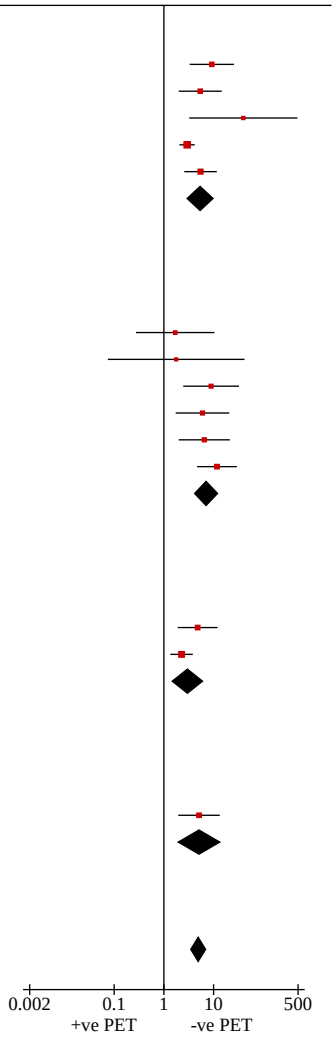
Analysis 3.4. Comparison 3: Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve, Outcome 4: PFS by disease stage

Study or Subgroup	log[Hazard Ratio]	SE	Experimental Total	Control Total	Weight	Hazard Ratio IV, Random, 95% CI
3.4.1 Stages I and II with A and B symptoms						
Barnes 2011	0.5261	0.9261	17	79	3.0%	1.69 [0.28 , 10.39]
Straus 2011	1.6268	0.49	24	64	7.7%	5.09 [1.95 , 13.29]
Subtotal (95% CI)			**41**	**143**	**10.7%**	**3.88 [1.54 , 9.83]**
Heterogeneity: Tau² = 0.06; Chi² = 1.10, df = 1 (P = 0.29); I² = 9%						
Test for overall effect: Z = 2.87 (P = 0.004)						
3.4.2 All stages						
Annunziata 2016	2.2192	0.5231	12	56	7.1%	9.20 [3.30 , 25.65]
Cerci 2010	1.5624	0.4706	30	74	8.1%	4.77 [1.90 , 12.00]
Hutchings 2005	0.570366	1.6144	22	63	1.1%	1.77 [0.07 , 41.87]
Hutchings 2006	2.187	0.6587	16	61	5.2%	8.91 [2.45 , 32.40]
Mesguich 2016	1.7891	0.6333	16	60	5.5%	5.98 [1.73 , 20.70]
Rossi 2014	1.873	0.6031	13	46	5.9%	6.51 [2.00 , 21.22]
Simon 2016	2.4596	0.4697	32	89	8.1%	11.70 [4.66 , 29.38]
Touati 2014	1.685	0.5075	24	44	7.4%	5.39 [1.99 , 14.58]
Ying 2014	3.6783	1.278	10	25	1.7%	39.58 [3.23 , 484.51]
Zaucha 2017	1.0753	0.1812	24	152	16.0%	2.93 [2.05 , 4.18]
Zinzani 2012	1.6982	0.3845	53	251	10.0%	5.46 [2.57 , 11.61]
Subtotal (95% CI)			**252**	**921**	**75.9%**	**5.81 [3.93 , 8.57]**
Heterogeneity: Tau² = 0.15; Chi² = 16.88, df = 10 (P = 0.08); I² = 41%						
Test for overall effect: Z = 8.85 (P < 0.00001)						
3.4.3 Advanced						
Kobe 2018	0.8198	0.2651	236	486	13.4%	2.27 [1.35 , 3.82]
Subtotal (95% CI)			**236**	**486**	**13.4%**	**2.27 [1.35 , 3.82]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 3.09 (P = 0.002)						
Total (95% CI)			**529**	**1550**	**100.0%**	**4.90 [3.47 , 6.90]**
Heterogeneity: Tau² = 0.16; Chi² = 23.52, df = 13 (P = 0.04); I² = 45%						
Test for overall effect: Z = 9.07 (P < 0.00001)						
Test for subgroup differences: Chi² = 8.05, df = 2 (P = 0.02), I² = 75.2%						



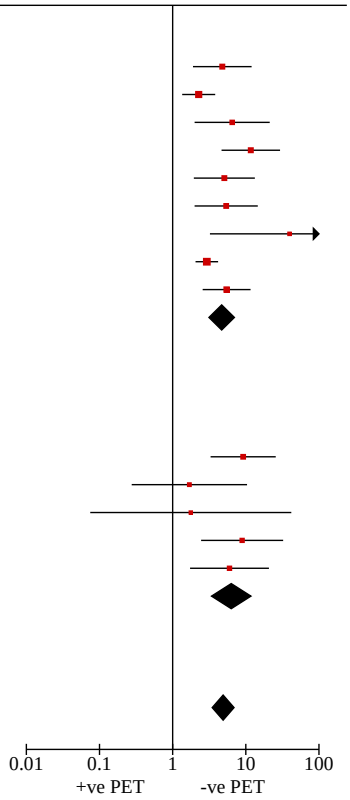
Analysis 3.5. Comparison 3: Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve, Outcome 5: PFS by radiotherapy

Study or Subgroup	log[Hazard Ratio]	SE	Experimental Total	Control Total	Weight	Hazard Ratio IV, Random, 95% CI
3.5.1 Involved node and/or site						
Annunziata 2016	2.2192	0.5231	12	56	7.1%	9.20 [3.30 , 25.65]
Touati 2014	1.685	0.5075	24	44	7.4%	5.39 [1.99 , 14.58]
Ying 2014	3.6783	1.278	10	25	1.7%	39.58 [3.23 , 484.51]
Zaucha 2017	1.0753	0.1812	24	152	16.0%	2.93 [2.05 , 4.18]
Zinzani 2012	1.6982	0.3845	53	251	10.0%	5.46 [2.57 , 11.61]
Subtotal (95% CI)			**123**	**528**	**42.2%**	**5.35 [2.94 , 9.75]**
Heterogeneity: Tau² = 0.24; Chi² = 9.63, df = 4 (P = 0.05); I² = 58%						
Test for overall effect: Z = 5.48 (P < 0.00001)						
3.5.2 Involved field						
Barnes 2011	0.5261	0.9261	17	79	3.0%	1.69 [0.28 , 10.39]
Hutchings 2005	0.570366	1.6144	22	63	1.1%	1.77 [0.07 , 41.87]
Hutchings 2006	2.187	0.6587	16	61	5.2%	8.91 [2.45 , 32.40]
Mesguich 2016	1.7891	0.6333	16	60	5.5%	5.98 [1.73 , 20.70]
Rossi 2014	1.873	0.6031	13	46	5.9%	6.51 [2.00 , 21.22]
Simon 2016	2.4596	0.4697	32	89	8.1%	11.70 [4.66 , 29.38]
Subtotal (95% CI)			**116**	**398**	**28.7%**	**7.06 [4.15 , 12.00]**
Heterogeneity: Tau² = 0.00; Chi² = 4.48, df = 5 (P = 0.48); I² = 0%						
Test for overall effect: Z = 7.21 (P < 0.00001)						
3.5.3 Not specified						
Cerci 2010	1.5624	0.4706	30	74	8.1%	4.77 [1.90 , 12.00]
Kobe 2018	0.8198	0.2651	236	486	13.4%	2.27 [1.35 , 3.82]
Subtotal (95% CI)			**266**	**560**	**21.4%**	**2.97 [1.48 , 5.98]**
Heterogeneity: Tau² = 0.13; Chi² = 1.89, df = 1 (P = 0.17); I² = 47%						
Test for overall effect: Z = 3.05 (P = 0.002)						
3.5.4 None						
Straus 2011	1.6268	0.49	24	64	7.7%	5.09 [1.95 , 13.29]
Subtotal (95% CI)			**24**	**64**	**7.7%**	**5.09 [1.95 , 13.29]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 3.32 (P = 0.0009)						
Total (95% CI)			**529**	**1550**	**100.0%**	**4.90 [3.47 , 6.90]**
Heterogeneity: Tau² = 0.16; Chi² = 23.52, df = 13 (P = 0.04); I² = 45%						
Test for overall effect: Z = 9.07 (P < 0.00001)						
Test for subgroup differences: Chi² = 3.73, df = 3 (P = 0.29), I² = 19.7%						

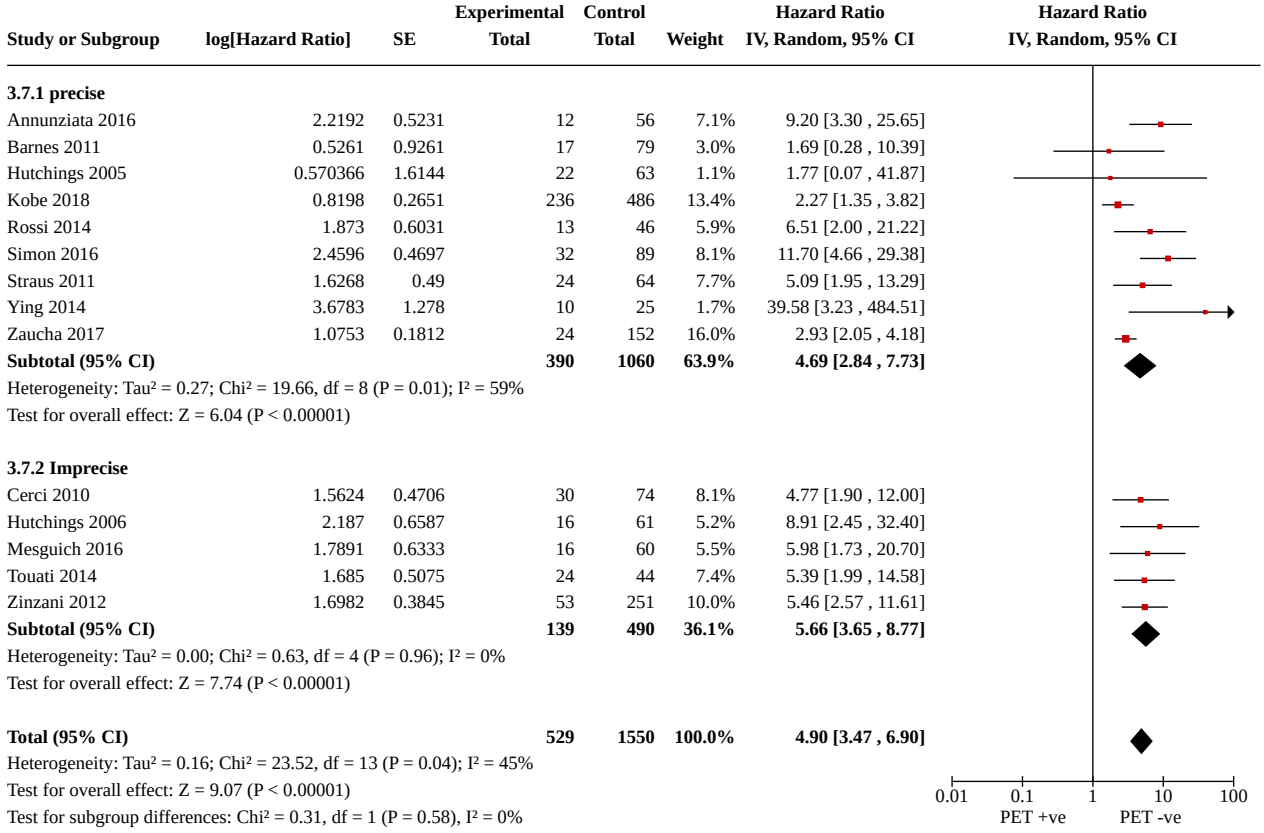


Analysis 3.6. Comparison 3: Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve, Outcome 6: Timing of interim PET

Study or Subgroup	log[Hazard Ratio]	SE	PET +ve Total	PET -ve Total	Weight	Hazard Ratio IV, Random, 95% CI
3.6.1 PET2						
Cerci 2010	1.5624	0.4706	30	74	8.1%	4.77 [1.90 , 12.00]
Kobe 2018	0.8198	0.2651	236	486	13.4%	2.27 [1.35 , 3.82]
Rossi 2014	1.873	0.6031	13	46	5.9%	6.51 [2.00 , 21.22]
Simon 2016	2.4596	0.4697	32	89	8.1%	11.70 [4.66 , 29.38]
Straus 2011	1.6268	0.49	24	64	7.7%	5.09 [1.95 , 13.29]
Touati 2014	1.685	0.5075	24	44	7.4%	5.39 [1.99 , 14.58]
Ying 2014	3.6783	1.278	10	25	1.7%	39.58 [3.23 , 484.51]
Zaucha 2017	1.0753	0.1812	24	152	16.0%	2.93 [2.05 , 4.18]
Zinzani 2012	1.6982	0.3845	53	251	10.0%	5.46 [2.57 , 11.61]
Subtotal (95% CI)			**446**	**1231**	**78.1%**	**4.68 [3.14 , 6.98]**
Heterogeneity: Tau² = 0.18; Chi² = 17.66, df = 8 (P = 0.02); I² = 55%						
Test for overall effect: Z = 7.56 (P < 0.00001)						
3.6.2 Other (including mixed)						
Annunziata 2016	2.2192	0.5231	12	56	7.1%	9.20 [3.30 , 25.65]
Barnes 2011	0.5261	0.9261	17	79	3.0%	1.69 [0.28 , 10.39]
Hutchings 2005	0.570366	1.6144	22	63	1.1%	1.77 [0.07 , 41.87]
Hutchings 2006	2.187	0.6587	16	61	5.2%	8.91 [2.45 , 32.40]
Mesguich 2016	1.7891	0.6333	16	60	5.5%	5.98 [1.73 , 20.70]
Subtotal (95% CI)			**83**	**319**	**21.9%**	**6.32 [3.40 , 11.75]**
Heterogeneity: Tau² = 0.00; Chi² = 3.44, df = 4 (P = 0.49); I² = 0%						
Test for overall effect: Z = 5.83 (P < 0.00001)						
Total (95% CI)			**529**	**1550**	**100.0%**	**4.90 [3.47 , 6.90]**
Heterogeneity: Tau² = 0.16; Chi² = 23.52, df = 13 (P = 0.04); I² = 45%						
Test for overall effect: Z = 9.07 (P < 0.00001)						
Test for subgroup differences: Chi² = 0.64, df = 1 (P = 0.42), I² = 0%						



Analysis 3.7. Comparison 3: Subgroups in univariable comparison of PFS: PET+ve vs. PET-ve, Outcome 7: PFS by HR type of estimation



Study or Subgroup	log[Hazard Ratio]	SE	Experimental Total	Control Total	Weight	Hazard Ratio IV, Random, 95% CI
3.7.1 precise						
Annunziata 2016	2.2192	0.5231	12	56	7.1%	9.20 [3.30 , 25.65]
Barnes 2011	0.5261	0.9261	17	79	3.0%	1.69 [0.28 , 10.39]
Hutchings 2005	0.570366	1.6144	22	63	1.1%	1.77 [0.07 , 41.87]
Kobe 2018	0.8198	0.2651	236	486	13.4%	2.27 [1.35 , 3.82]
Rossi 2014	1.873	0.6031	13	46	5.9%	6.51 [2.00 , 21.22]
Simon 2016	2.4596	0.4697	32	89	8.1%	11.70 [4.66 , 29.38]
Straus 2011	1.6268	0.49	24	64	7.7%	5.09 [1.95 , 13.29]
Ying 2014	3.6783	1.278	10	25	1.7%	39.58 [3.23 , 484.51]
Zaucha 2017	1.0753	0.1812	24	152	16.0%	2.93 [2.05 , 4.18]
Subtotal (95% CI)			**390**	**1060**	**63.9%**	**4.69 [2.84 , 7.73]**
Heterogeneity: Tau² = 0.27; Chi² = 19.66, df = 8 (P = 0.01); I² = 59%						
Test for overall effect: Z = 6.04 (P < 0.00001)						
3.7.2 Imprecise						
Cerci 2010	1.5624	0.4706	30	74	8.1%	4.77 [1.90 , 12.00]
Hutchings 2006	2.187	0.6587	16	61	5.2%	8.91 [2.45 , 32.40]
Mesguich 2016	1.7891	0.6333	16	60	5.5%	5.98 [1.73 , 20.70]
Touati 2014	1.685	0.5075	24	44	7.4%	5.39 [1.99 , 14.58]
Zinzani 2012	1.6982	0.3845	53	251	10.0%	5.46 [2.57 , 11.61]
Subtotal (95% CI)			**139**	**490**	**36.1%**	**5.66 [3.65 , 8.77]**
Heterogeneity: Tau² = 0.00; Chi² = 0.63, df = 4 (P = 0.96); I² = 0%						
Test for overall effect: Z = 7.74 (P < 0.00001)						
Total (95% CI)			**529**	**1550**	**100.0%**	**4.90 [3.47 , 6.90]**
Heterogeneity: Tau² = 0.16; Chi² = 23.52, df = 13 (P = 0.04); I² = 45%						
Test for overall effect: Z = 9.07 (P < 0.00001)						
Test for subgroup differences: Chi² = 0.31, df = 1 (P = 0.58), I² = 0%						

APPENDICES

Appendix 1. Cochrane Central Register of Controlled Trials search strategy

Searches until 07/02/2016

ID Search

#1 MeSH descriptor: [Lymphoma] this term only

#2 MeSH descriptor: [Hodgkin Disease] explode all trees

#3 Germinoblastom*

#4 Reticulolymphosarcom*

#5 Hodgkin*

#6 (malignan* near/2 (lymphogranulom* or granulom*))

#7 #1 or #2 or #3 or #4 or #5 or #6

#8 MeSH descriptor: [Positron-Emission Tomography] explode all trees

#9 (pet* or petscan*)

#10 tomograph*

#11 emission*

#12 MeSH descriptor: [Deoxyglucose] explode all trees

#13 MeSH descriptor: [Fluorodeoxyglucose F18] explode all trees

#14 (deoxyglucose* or desoxyglucose* or deoxy-glucose* or desoxy-glucose* or deoxy-d-glucose* or desoxy-d-glucose* or 2deoxyglucose* or 2deoxy-d-glucose* or fluorodeoxyglucose* or fluorodesoxyglucose* or fludeoxyglucose* or fluordeoxyglucose* or fluordesoxyglucose* or 18fluorodeoxyglucose* or 18fluorodesoxyglucose* or 18fluordeoxyglucose* or fdg* or 18fdg* or 18f-dg*)

#15 (fluor* or 2fluor* or fluoro* or fluorodeoxy* or fludeoxy* or fluorine* or 18f* or 18flu*)

#16 glucose*

#17 #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16

#18 #7 and #17 in trials

Searches from 08/02/2016 - 13/07/2017

ID Search

#1 MeSH descriptor: [Lymphoma] this term only

#2 MeSH descriptor: [Hodgkin Disease] explode all trees

#3 Germinoblastom*

#4 Reticulolymphosarcom*

#5 Hodgkin*

#6 (malignan* near/2 (lymphogranulom* or granulom*))

#7 #1 or #2 or #3 or #4 or #5 or #6

#8 MeSH descriptor: [Positron-Emission Tomography] explode all trees

#9 (pet* or petscan*)

#10 tomograph*

#11 emission*

#12 MeSH descriptor: [Deoxyglucose] explode all trees

#13 MeSH descriptor: [Fluorodeoxyglucose F18] explode all trees

#14 (deoxyglucose* or desoxyglucose* or deoxy-glucose* or desoxy-glucose* or deoxy-d-glucose* or desoxy-d-glucose* or 2deoxyglucose* or 2deoxy-d-glucose* or fluorodeoxyglucose* or fluorodesoxyglucose* or fludeoxyglucose* or fluordeoxyglucose* or fluordesoxyglucose* or 18fluorodeoxyglucose* or 18fluorodesoxyglucose* or 18fluordeoxyglucose* or fdg* or 18fdg* or 18f-dg*)

#15 (fluor* or 2fluor* or fluoro* or fluorodeoxy* or fludeoxy* or fluorine* or 18f* or 18flu*)

#16 glucose*

#17 #15 and #16

#18 #8 or #9 or #10 or #11 or #12 or #13 or #14 or #17

#19 #7 and #18 in Trials

#20 #19 Publication Year from 2016 to 2017

Searches from 12/07/2017 - 12/11/2018

Cochrane Central Register of Controlled Trials (Cochrane Library Issue 11, 2018)

ID Search

#1 MeSH descriptor: [Lymphoma] this term only

#2 MeSH descriptor: [Hodgkin Disease] this term only

#3 germinoblastom*

#4 reticulolymphosarcom*

#5 hodgkin*

#6 (malignan* near/2 (lymphogranulom* or granulom*))

#7 #1 or #2 or #3 or #4 or #5 or #6

#8 MeSH descriptor: [Positron-Emission Tomography] explode all trees

#9 (pet*)

#10 tomograph*

#11 emission*

#12 MeSH descriptor: [Deoxyglucose] explode all trees

#13 MeSH descriptor: [Fluorodeoxyglucose F18] this term only

#14 (deoxyglucose* or desoxyglucose* or deoxy-glucose* or desoxy-glucose* or deoxy-d-glucose* or desoxy-d-glucose* or 2deoxyglucose* or 2deoxy-d-glucose* or fluorodeoxyglucose* or fluorodesoxyglucose* or fludeoxyglucose* or fluordeoxyglucose* or fluordesoxyglucose* or 18fluorodeoxyglucose* or 18fluorodesoxyglucose* or 18fluordeoxyglucose* or fdg* or 18fdg* or 18f-dg*)

#15 (glucose* and (fluor* or 2fluor* or fludeoxy* or 18f*))

#16 #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15

#17 #7and #16 in Trials

key: *: truncation, near/#: adjacent within # number of words

Searches from 12/11/2018 - 02/04/2019

ID Search

#1 MeSH descriptor: [Lymphoma] this term only

#2 MeSH descriptor: [Hodgkin Disease] explode all trees

#3 Germinoblastom*

#4 Reticulolymphosarcom*

#5 Hodgkin*

#6 (malignan* near/2 (lymphogranulom* or granulom*))

#7 #1 or #2 or #3 or #4 or #5 or #6

#8 MeSH descriptor: [Positron-Emission Tomography] explode all trees

#9 (pet* or petscan*)

#10 tomograph*

#11 emission*

#12 MeSH descriptor: [Deoxyglucose] explode all trees

#13 MeSH descriptor: [Fluorodeoxyglucose F18] explode all trees

#14 (deoxyglucose* or desoxyglucose* or deoxy-glucose* or desoxy-glucose* or deoxy-d-glucose* or desoxy-d-glucose* or 2deoxyglucose* or 2deoxy-d-glucose* or fluorodeoxyglucose* or fluorodesoxyglucose* or fludeoxyglucose* or fluordeoxyglucose* or fluordesoxyglucose* or 18fluorodeoxyglucose* or 18fluorodesoxyglucose* or 18fluordeoxyglucose* or fdg* or 18fdg* or 18f-dg*)

#15 (fluor* or 2fluor* or fluoro* or fluorodeoxy* or fludeoxy* or fluorine* or 18f* or 18flu*)

#16 glucose*

#17 #15 and #16

#18 #8 or #9 or #10 or #11 or #12 or #13 or #14 or #17

#19 #7 and #18 in Trials

Appendix 2. MEDLINE Ovid search strategy

#	Searches until 02/02/2016
1	*LYMPHOMA/
2	exp HODGKIN DISEASE/
3	Germinoblastom$.tw,kf,ot.
4	Reticulolymphosarcom$.tw,kf,ot.
5	Hodgkin$.tw,kf,ot.
6	(malignan$ adj2 (lymphogranulom$ or granulom$)).tw,kf,ot.
7	or/1-6
8	POSITRON-EMISSION TOMOGRAPHY/
9	(pet$ or petscan$).tw,kf,ot.
10	tomograph$.tw,kf,ot.
11	emission$.tw,kf,ot.
12	exp DEOXYGLUCOSE/
13	FLUORODEOXYGLUCOSE F18/
14	(deoxyglucose$ or desoxyglucose$ or deoxy-glucose$ or desoxy-glucose$ or deoxy-d-glucose$ or desoxy-d-glucose$ or 2deoxyglucose$ or 2deoxy-d-glucose$ or fluorodeoxyglucose$ or fluorodesoxyglucose$ or fludeoxyglucose$ or fluordeoxyglucose$ or fluordesoxyglucose$ or 18fluorodeoxyglucose$ or 18fluorodesoxyglucose$ or 18fluordeoxyglucose$ or fdg$ or 18fdg$ or 18f-dg$).tw.
15	(fluor$ or 2fluor$ or fluoro$ or fluorodeoxy$ or fludeoxy$ or fluorine$ or 18f$ or 18flu$).tw.
16	glucose$.tw.
17	or/8-16
18	7 and 17
19	ANIMALS/ not HUMANS/
20	18 not 19

#	Searches from 03/02/2016 – 12/07/2017
1	*LYMPHOMA/
2	exp HODGKIN DISEASE/
3	Germinoblastom$.tw,kf,ot.
4	Reticulolymphosarcom$.tw,kf,ot.
5	Hodgkin$.tw,kf,ot.
6	(malignan$ adj2 (lymphogranulom$ or granulom$)).tw,kf,ot.
7	or/1-6
8	exp POSITRON-EMISSION TOMOGRAPHY/
9	(pet$ or petscan$).tw,kf,ot.
10	tomograph$.tw,kf,ot.
11	emission$.tw,kf,ot.
12	exp DEOXYGLUCOSE/
13	FLUORODEOXYGLUCOSE F18/
14	(deoxyglucose$ or desoxyglucose$ or deoxy-glucose$ or desoxy-glucose$ or deoxy-d-glucose$ or desoxy-d-glucose$ or 2deoxyglucose$ or 2deoxy-d-glucose$ or fluorodeoxyglucose$ or fluorodesoxyglucose$ or fludeoxyglucose$ or fluordeoxyglucose$ or fluordesoxyglucose$ or 18fluorodeoxyglucose$ or 18fluorodesoxyglucose$ or 18fluordeoxyglucose$ or fdg$ or 18fdg$ or 18f-dg$).tw.
15	(fluor$ or 2fluor$ or fluoro$ or fluorodeoxy$ or fludeoxy$ or fluorine$ or 18f$ or 18flu$).tw.
16	glucose$.tw.
17	15 and 16
18	or/8-14
19	17 or 18
20	7 and 19
21	ANIMALS/ not HUMANS/
22	20 not 21
23	limit 22 to ed=20160203-20170712

#	Searches from 12/07/2017 - 12/11/2018
1	*LYMPHOMA/
2	HODGKIN DISEASE/
3	germinoblastom$.tw,kf,ot.
4	reticulolymphosarcom$.tw,kf,ot.
5	hodgkin$.tw,kf,ot.
6	(malignan$ adj2 (lymphogranulom$ or granulom$)).tw,kf,ot.
7	or/1-6
8	exp POSITRON-EMISSION TOMOGRAPHY/
9	(pet$).tw,kf,ot.
10	tomograph$.tw,kf,ot.
11	emission$.tw,kf,ot.
12	exp DEOXYGLUCOSE/
13	FLUORODEOXYGLUCOSE F18/
14	(deoxyglucose$ or desoxyglucose$ or deoxy-glucose$ or desoxy-glucose$ or deoxy-d-glucose$ or desoxy-d-glucose$ or 2deoxyglucose$ or 2deoxy-d-glucose$ or fluorodeoxyglucose$ or fluorodesoxyglucose$ or fludeoxyglucose$ or fluordeoxyglucose$ or fluordesoxyglucose$ or 18fluorodeoxyglucose$ or 18fluorodesoxyglucose$ or 18fluordeoxyglucose$ or fdg$ or 18fdg$ or 18f-dg$).tw.
15	(glucose$ and (fluor$ or 2fluor$ or fludeoxy$ or 18f$)).tw.
16	or/8-15
17	7 and 16
18	exp ANIMALS/ not HUMANS/
19	17 not 18
20	limit 19 to ed=20160203-20170712
21	limit 19 to ed=20170712-20181112
22	ANIMALS/ not HUMANS/
23	21 not 22
24	limit 23 to ed=20160203-20170712
25	limit 23 to ed=20170712-20181112

#	Searches from 12/11/2018 - 02/04/2019
1	*LYMPHOMA/
2	HODGKIN DISEASE/
3	Germinoblastom$.tw,kf,ot.
4	Reticulolymphosarcom$.tw,kf,ot.
5	Hodgkin$.tw,kf,ot.
6	(malignan$ adj2 (lymphogranulom$ or granulom$)).tw,kf,ot.
7	or/1-6
8	exp POSITRON-EMISSION TOMOGRAPHY/
9	(pet$ or petscan$).tw,kf,ot.
10	tomograph$.tw,kf,ot.
11	emission$.tw,kf,ot.
12	exp DEOXYGLUCOSE/
13	Fluorodeoxyglucose F18/
14	(deoxyglucose$ or desoxyglucose$ or deoxy-glucose$ or desoxy-glucose$ or deoxy-d-glucose$ or desoxy-d-glucose$ or 2deoxyglucose$ or 2deoxy-d-glucose$ or fluorodeoxyglucose$ or fluorodesoxyglucose$ or fludeoxyglucose$ or fluordeoxyglucose$ or fluordesoxyglucose$ or 18fluorodeoxyglucose$ or 18fluorodesoxyglucose$ or 18fluordeoxyglucose$ or fdg$ or 18fdg$ or 18f-dg $).tw.
15	(glucose$ and (fluor$ or 2fluor$ or fludeoxy$ or 18f$)).tw.
16	or/8-15
17	7 and 16
18	exp ANIMALS/ not HUMANS/
19	17 not 18
20	limit 19 to ed=20160203-20170712
21	limit 19 to ed=20170712-20181112
22	limit 19 to ed=20181112-20190402

key: exp # /: explode # MeSH subject heading, tw: text word, kf: keyword heading word, ot: original title, ti: title, $: truncation, adj#: adjacent within # number of words

Appendix 3. Embase/Ovid search strategy

#	Searches
1	exp CLASSICAL HODGKIN LYMPHOMA/
2	*HODGKIN DISEASE/
3	germinoblastom*.tw,kw.
4	reticulolymphosarcom*.tw,kw.
5	hodgkin*.tw,kw.
6	(malignan* adj2 (lymphogranulom* or granulom*)).tw,kw.
7	or/1-6
8	exp POSITRON EMISSION TOMOGRAPHY/
9	pet*.tw,kw.
10	tomograph*.tw,kw.
11	emission*.tw,kw.
12	exp DEOXYGLUCOSE/
13	FLUORODEOXYGLUCOSE F 18/
14	(deoxyglucose* or desoxyglucose* or deoxy-glucose* or desoxy-glucose* or deoxy-d-glucose* or desoxy-d-glucose* or 2deoxyglucose* or 2deoxy-d-glucose* or fluorodeoxyglucose* or fluorodesoxyglucose* or fludeoxyglucose* or fluordeoxyglucose* or fluordesoxyglucose* or 18fluorodeoxyglucose* or 18fluorodesoxyglucose* or 18fluordeoxyglucose* or fdg* or 18fdg* or 18f-dg*).tw.
15	(glucose* and (fluor* or 2fluor* or fludeoxy* or 18f*)).tw.
16	or/8-15
17	7 and 16
18	exp ANIMAL/ not HUMAN/
19	17 not 18
20	limit 19 to yr="1990 -Current"
21	meta-analys:.mp. or search:.tw. or review.pt.
22	(child* or p?ediatric*).ti.
23	20 not (21 or 22)
24	limit 23 to embase

(Continued)

25	limit 23 to conference abstracts
26	24 or 25

key: exp # /: explode # EMTREE term, * # /: focus # EMTREE term, /: EMTREE term, tw: text word, kw: keyword, ti: title, mp: multiple purpose, pt: publication type, *: truncation, ?: wildcard

search line #21: Review Embase search filter - best balance of sensitivity and specificity https://hiru.mcmaster.ca/hiru/HIRU_Hedges_EMBASE_Strategies.aspx

WHAT'S NEW

Date	Event	Description
14 August 2020	Amended	Following correspondence between the editorial base and the funding institution of one of the authors, the internal sources of support and the acknowledging statement was updated.

HISTORY

Protocol first published: Issue 4, 2017
Review first published: Issue 9, 2019

Date	Event	Description
20 December 2019	New citation required but conclusions have not changed	Following correspondence between the authors and one of the peer reviewers post-publication, the authors have revised some of the risk of bias judgements. Some terminology around confounding has also been changed.
20 December 2019	Amended	Following correspondence between the authors and one of the peer reviewers post-publication, the authors have revised some of the risk of bias judgements. Some terminology around confounding has also been changed.

CONTRIBUTIONS OF AUTHORS

Angela Aldin: screening and selection of studies, development of data extraction form, data extraction, 'Risk of bias' assessment, GRADE assessment, data analysis interpretation, 'Summary of findings' tables, writing and drafting of the review, communication with and between authors.

Lisa Umlauff: 'Risk of bias' assessment, characteristics of included and excluded studies (texts and tables), abstract and Plain language summary, proofread and commented on the draft.

Karel Moons: methodological input on reviews of prognosis studies.

Lise J Estcourt: screening and selection of studies, data extraction, risk of bias assessment, clinical and methodological input.

Andreas Engert: medical and content input, particularly on the clinical comparability of studies and subgroup analysis.

Carsten Kobe: nuclear medical input on PET-CT.

Bastian von Tresckow: clinical input, particularly on the clinical comparability of studies and subgroup analysis.

Gary Collins: methodological input on reviews of prognostic studies.

Madhuri Haque: screening and selection of studies.

Farid Foroutan: input on risk of bias and GRADE assessment of prognostic factor studies.

Nina Kreuzberger: 'Risk of bias' assessment, proofread and commented on the review draft.

Marialena Trivella: screening and selection of studies, data extraction, risk of bias assessment, statistical analysis, proofread and commented on the review draft.

Nicole Skoetz: protocol development, screening and selection of studies, data extraction, risk of bias assessment, GRADE assessment, proofread and commented on the review draft.

DECLARATIONS OF INTEREST

Angela Aldin: award of the grant by Federal Ministry of Education and Research for the University Hospital of Cologne to perform this systematic review does not lead to a conflict of interest.

Lisa Umlauff: award of the grant by Federal Ministry of Education and Research for the University Hospital of Cologne to perform this systematic review does not lead to a conflict of interest.

Karel Moons: none known.

Lise J Estcourt: award of the grant by Federal Ministry of Education and Research to the University of Oxford to perform this systematic review does not lead to a conflict of interest.

Andreas Engert: award of the grant by Federal Ministry of Education and Research for the University Hospital of Cologne to perform this systematic review does not lead to a conflict of interest. Principal investigator of the HD18 trial, does not lead to a conflict of interest. Received funds from Takeda Pharma GmbH, BMS and MSD for consultancy and educational presentations, but these were not related to the intervention in this review. No competing interests.

Carsten Kobe: award of the grant by Federal Ministry of Education and Research for the University Hospital of Cologne to perform this systematic review does not lead to a conflict of interest.

Bastian von Tresckow: award of the grant by Federal Ministry of Education and Research for the University Hospital of Cologne to perform this systematic review does not lead to a conflict of interest. Received funds from Novartis Pharma GmbH, Takeda Pharma GmbH and MSD for consultancy and educational presentations, but these were not related to the intervention in this review. No competing interests.

Gary Collins: supported by the NIHR Biomedical Research Centre, Oxford, and Cancer Research UK (programme grant: C49297/A27294). No conflict of interest.

Madhuri Haque: award of the grant by Federal Ministry of Education and Research for the University Hospital of Cologne to perform this systematic review does not lead to a conflict of interest.

Farid Foroutan: none known.

Nina Kreuzberger: award of the grant by Federal Ministry of Education and Research for the University Hospital of Cologne to perform this systematic review does not lead to a conflict of interest.

Marialena Trivella: part of the grant from the Federal Ministry of Education and Research to the University Hospital of Cologne, was paid to the University of Oxford for author time spent working on this review. However, the funder played no part in the design and execution of the project and it does not constitute a conflict of interest.

Nicole Skoetz: award of the grant by Federal Ministry of Education and Research for the University Hospital of Cologne to perform this systematic review does not lead to a conflict of interest.

SOURCES OF SUPPORT

Internal sources

- University Hospital of Cologne, Germany

  Cochrane Haematological Malignancies, Department 1 of Internal Medicine
- NHS Blood and Transplant, UK

External sources

- Federal Ministry of Education and Research, Germany

  Funding Number 01KG1709

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

We included studies that evaluated both adult and adolescent participants (the youngest being 13 years old), as opposed to including adult participants (≥ 18 years of age) only as stated in the protocol of this review. Hodgkin lymphoma is a disease with a typical onset in adolescence to mid-adulthood, with little physiological differences between adolescents and adults. In the studies included in this review, participants under the age of 18 were treated in the same clinic and received the same treatment as participants over ≥ 18 years of age. We believe that the results regarding interim PET are equally relevant to adolescents as they are to adult participants and, therefore, did not see reasons against the inclusion of studies including both younger and older adults. Nevertheless, we did not include studies that evaluated solely paediatric participants. In studies where only paediatric participants are included, it is more likely they will be treated in paediatric clinics and receive a different treatment regimen than adult participants.

We used an amended version of the Quality in Prognostic Factor Studies (QUIPS) tool to assess the risk of bias of the included studies. In consultation with Hayden and colleagues (Hayden 2013), we adapted the QUIPS tool by adding 'unclear (no information)' as a fourth judgement in the tool. In addition, we renamed the fifth domain of the tool, originally named 'study confounding', into 'other prognostic factors (covariates)', to highlight the importance of adjusting for other prognostic factors and distinguish it from confounding. Lastly, we assessed all six domains (study participation, study attrition, prognostic factor measurement, outcome measurement, other prognostic factors (covariates), statistical analysis and reporting) per outcome (OS and PFS) in each study. The first three domains ended up always receiving the same judgement as they are indeed to be considered at study level. With regard to the outcomes, however, we identified differences in analysis and reporting within studies.

With regard to data extraction, we developed our own data extraction form specific to prognostic factor studies (particularly those that are included in this review), which includes more items than stipulated in the protocol of this review.

Lastly, we searched Embase as an additional database, as well as one trial registry (ClinicalTrials.gov).

INDEX TERMS

Medical Subject Headings (MeSH)

Antineoplastic Combined Chemotherapy Protocols [*therapeutic use]; Chemoradiotherapy; Decision Making; Disease Progression; Disease-Free Survival; Hodgkin Disease [*drug therapy]; Positron Emission Tomography Computed Tomography [*methods]; Prognosis

MeSH check words

Humans; Young Adult