



Cochrane
Library

Cochrane Database of Systematic Reviews

Positron emission tomography (PET) and magnetic resonance imaging (MRI) for assessing tumour resectability in advanced epithelial ovarian/fallopian tube/primary peritoneal cancer (Review)

Roze JF, Hoogendam JP, van de Wetering FT, Spijker R, Verleye L, Vlayen J, Veldhuis WB, Scholten RJPM, Zweemer RP

Roze JF, Hoogendam JP, van de Wetering FT, Spijker R, Verleye L, Vlayen J, Veldhuis WB, Scholten RJPM, Zweemer RP. Positron emission tomography (PET) and magnetic resonance imaging (MRI) for assessing tumour resectability in advanced epithelial ovarian/fallopian tube/primary peritoneal cancer. *Cochrane Database of Systematic Reviews* 2018, Issue 10. Art. No.: CD012567. DOI: [10.1002/14651858.CD012567.pub2](https://doi.org/10.1002/14651858.CD012567.pub2).

www.cochranelibrary.com

TABLE OF CONTENTS

ABSTRACT	1
PLAIN LANGUAGE SUMMARY	2
SUMMARY OF FINDINGS	3
BACKGROUND	5
OBJECTIVES	6
METHODS	6
Figure 1.	8
Figure 2.	9
Figure 3.	9
RESULTS	10
Figure 4.	11
Figure 5.	13
Figure 6.	14
DISCUSSION	15
AUTHORS' CONCLUSIONS	16
ACKNOWLEDGEMENTS	16
REFERENCES	17
CHARACTERISTICS OF STUDIES	19
DATA	30
Test 1. PET/CT for assessing incomplete debulking with residual disease of any size.	31
Test 2. MRI for assessing incomplete debulking with residual disease > 1 cm.	31
Test 3. MRI for assessing incomplete debulking with residual disease > 2 cm.	31
Test 4. MRI for assessing incomplete debulking with residual disease of any size.	31
ADDITIONAL TABLES	31
APPENDICES	32
CONTRIBUTIONS OF AUTHORS	39
DECLARATIONS OF INTEREST	39
SOURCES OF SUPPORT	39
DIFFERENCES BETWEEN PROTOCOL AND REVIEW	39
INDEX TERMS	39

[Diagnostic Test Accuracy Review]

Positron emission tomography (PET) and magnetic resonance imaging (MRI) for assessing tumour resectability in advanced epithelial ovarian/fallopian tube/primary peritoneal cancer

Joline F Roze¹, Jacob P Hoogendam¹, Fleur T van de Wetering², René Spijker², Leen Verleye³, Joan Vlayen³, Wouter B Veldhuis⁴, Rob JPM Scholten², Ronald P Zweemer¹

¹Department of Gynaecological Oncology, UMC Utrecht Cancer Center, Utrecht, Netherlands. ²Cochrane Netherlands, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands. ³Belgian Health Care Knowledge Centre, Brussels, Belgium. ⁴Department of Radiology, University Medical Center Utrecht, Utrecht, Netherlands

Contact: Joline F Roze, Department of Gynaecological Oncology, UMC Utrecht Cancer Center, Utrecht, 3508 GA, Netherlands. J.F.Roze@umcutrecht.nl.

Editorial group: Cochrane Gynaecological, Neuro-oncology and Orphan Cancer Group.

Publication status and date: New, published in Issue 10, 2018.

Citation: Roze JF, Hoogendam JP, van de Wetering FT, Spijker R, Verleye L, Vlayen J, Veldhuis WB, Scholten RJPM, Zweemer RP. Positron emission tomography (PET) and magnetic resonance imaging (MRI) for assessing tumour resectability in advanced epithelial ovarian/fallopian tube/primary peritoneal cancer. *Cochrane Database of Systematic Reviews* 2018, Issue 10. Art. No.: CD012567. DOI: [10.1002/14651858.CD012567.pub2](https://doi.org/10.1002/14651858.CD012567.pub2).

Copyright © 2018 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ABSTRACT

Background

Ovarian cancer is the leading cause of death from gynaecological cancer in developed countries. Surgery and chemotherapy are considered its mainstay of treatment and the completeness of surgery is a major prognostic factor for survival in these women. Currently, computed tomography (CT) is used to preoperatively assess tumour resectability. If considered feasible, women will be scheduled for primary debulking surgery (i.e. surgical efforts to remove the bulk of tumour with the aim of leaving no visible (macroscopic) tumour). If primary debulking is not considered feasible (i.e. the tumour load is too extensive), women will receive neoadjuvant chemotherapy to reduce tumour load and subsequently undergo (interval) surgery. However, CT is imperfect in assessing tumour resectability, so additional imaging modalities can be considered to optimise treatment selection.

Objectives

To assess the diagnostic accuracy of fluorodeoxyglucose-18 (FDG) PET/CT, conventional and diffusion-weighted (DW) MRI as replacement or add-on to abdominal CT, for assessing tumour resectability at primary debulking surgery in women with stage III to IV epithelial ovarian/fallopian tube/primary peritoneal cancer.

Search methods

We searched MEDLINE and Embase (OVID) for potential eligible studies (1946 to 23 February 2017). Additionally, ClinicalTrials.gov, WHO-ICTRP and the reference list of all relevant studies were searched.

Selection criteria

Diagnostic accuracy studies addressing the accuracy of preoperative FDG-PET/CT, conventional or DW-MRI on assessing tumour resectability in women with advanced stage (III to IV) epithelial ovarian/fallopian tube/primary peritoneal cancer who are scheduled to undergo primary debulking surgery.

Data collection and analysis

Two authors independently screened titles and abstracts for relevance and inclusion, extracted data and performed methodological quality assessment using QUADAS-2. The limited number of studies did not permit meta-analyses.

Main results

Five studies (544 participants) were included in the analysis. All studies performed the index test as replacement of abdominal CT. Two studies (366 participants) addressed the accuracy of FDG-PET/CT for assessing incomplete debulking with residual disease of any size (> 0 cm) with sensitivities of 1.0 (95% CI 0.54 to 1.0) and 0.66 (95% CI 0.60 to 0.73) and specificities of 1.0 (95% CI 0.80 to 1.0) and 0.88 (95% CI 0.80 to 0.93), respectively (low- and moderate-certainty evidence). Three studies (178 participants) investigated MRI for different target conditions, of which two investigated DW-MRI and one conventional MRI. The first study showed that DW-MRI determines incomplete debulking with residual disease of any size with a sensitivity of 0.94 (95% CI 0.83 to 0.99) and a specificity of 0.98 (95% CI 0.88 to 1.00) (low- and moderate-certainty evidence). For abdominal CT, the sensitivity for assessing incomplete debulking was 0.66 (95% CI 0.52 to 0.78) and the specificity 0.77 (95% CI 0.63 to 0.87) (low- and low-certainty evidence). The second study reported a sensitivity of DW-MRI of 0.75 (95% CI 0.35 to 0.97) and a specificity of 0.96 (95% CI 0.80 to 1.00) (very low-certainty evidence) for assessing incomplete debulking with residual disease > 1 cm. In the last study, the sensitivity for assessing incomplete debulking with residual disease of > 2 cm on conventional MRI was 0.91 (95% CI 0.59 to 1.00) and the specificity 0.97 (95% CI 0.87 to 1.00) (very low-certainty evidence). Overall, the certainty of evidence was very low to moderate (according to GRADE), mainly due to small sample sizes and imprecision.

Authors' conclusions

Studies suggested a high specificity and moderate sensitivity for FDG-PET/CT and MRI to assess macroscopic incomplete debulking. However, the certainty of the evidence was insufficient to advise routine addition of FDG-PET/CT or MRI to clinical practice..

In a research setting, adding an alternative imaging method could be considered for women identified as suitable for primary debulking by abdominal CT, in an attempt to filter out false-negatives (i.e. debulking, feasible based on abdominal CT, unfeasible at actual surgery).

PLAIN LANGUAGE SUMMARY

How accurate are the imaging techniques PET and MRI to determine the feasibility of primary debulking surgery for ovarian cancer?

Why is it important to determine the feasibility of ovarian tumour resection?

Ovarian cancer is a disease with a high mortality that affects 239,000 women each year across the world. By the time it is symptomatic and detected, cancer cells have spread throughout the abdomen in most women. Treatment consists of surgery to remove as much visible tumour as possible (also called debulking surgery) and chemotherapy. Randomised controlled trials have shown that in women where all visible cancer cannot be removed with surgery, giving chemotherapy first to shrink the tumour is an alternative treatment strategy. This can improve the number of women having successful removal of all visible tumour, known as macroscopic debulking. Therefore, it is important to determine beforehand if all visible tumour deposits can be removed by surgery, followed by chemotherapy, or if chemotherapy is needed first to reduce tumour size before surgery is performed.

Imaging with abdominal computed tomography (abdominal CT) is currently used to determine whether primary debulking surgery is feasible. However, it cannot determine the outcome correctly in all women. Other imaging techniques that can be used are positron emission tomography (PET) and magnetic resonance imaging (MRI). PET visualises glucose uptake by cells and allows detection of distant metastases and is frequently performed parallel with abdominal CT (FDG-PET/CT). MRI provides good soft tissue contrast to detect small lesions. These additional imaging techniques may improve treatment selection.

What is the aim of this review?

To investigate the accuracy of PET and MRI in women with advanced stage ovarian cancer to determine the feasibility of primary debulking surgery.

What are the main results of the review?

We identified two studies (with 366 participants) addressing the accuracy of FDG-PET/CT and three studies (with 178 participants) investigating the accuracy of MRI.

In a hypothetical group of 1000 women, of whom 620 would have residual tumour after surgery (prevalence 62%), 211 women would incorrectly be considered suitable for surgery according to FDG-PET/CT and 37 women according to MRI. However, the quality and quantity of these studies were insufficient for these imaging techniques to be used routinely in clinical practice. Therefore, the authors concluded that more research is needed before such a recommendation can be made.

SUMMARY OF FINDINGS

Summary of findings 1. Diagnostic accuracy of FDG-PET/CT and MRI for assessing tumour resectability in advanced epithelial ovarian/fallopian tube/primary peritoneal cancer

What is the diagnostic accuracy of FDG-PET/CT or MRI for assessing tumour resectability in advanced epithelial ovarian/fallopian tube/primary peritoneal cancer?

Patients Women suspected of ovarian cancer scheduled for surgery

Prior testing Conventional diagnostic work-up (e.g. physical examination, ultrasound)

Setting University hospitals or specialised cancer institutes

Index test FDG-PET/CT or MRI. In all studies, the index test was evaluated as a replacement of abdominal CT. No studies were identified that followed an add-on design.

Target condition Residual disease assessed after debulking surgery

Test	Target condition	No. of women (studies)	Prevalence in study	Sensitivity (95% CI)	Specificity (95% CI)	No. of false negatives* per 1000 tested	No. of false positives** per 1000 tested	Test accuracy certainty (quality) of evidence (sensitivity/specificity) ^a
FDG-PET/CT	Residual disease > 0 cm	23/343 (2)	26%/65%	1.0 (0.54 to 1.0) and 0.66 (0.60 to 0.73)	1.0 (0.80 to 1.0) and 0.88 (0.80 to 0.93)	211 (167 to 248) ^b	46 (27 to 76) ^b	Low ^c /moderate ^d
DW-MRI	Residual disease > 0 cm	94 (1)	53%	0.94 (0.83 to 0.99)	0.98 (0.88 to 1.00)	37 (6 to 105) ^b	8 (0 to 46) ^b	Low ^c /moderate ^d
DW-MRI	Residual disease > 1 cm	34 (1)	23.5%	0.75 (0.35 to 0.97)	0.96 (0.80 to 1.00)	59 (7 to 153)	31 (0 to 153)	Very low/very low ^{e, f}
Conventional MRI	Residual disease > 2 cm	50 (1)	22%	0.91 (0.59 to 1.00)	0.97 (0.87 to 1.00)	20 (0 to 90)	23 (0 to 101)	Very low/very low ^{e, g}
CT^h	Residual disease > 0 cm	94 (1)	53%	0.66 (95% CI 0.52 to 0.78)	0.77 (95% CI 0.63 to 0.87)	211 (136 to 298) ^b	87 (49 to 141) ^b	Low/low ^c

CI: confidence interval

CT: computed tomography

DW-MRI: diffusion-weighted Magnetic Resonance Imaging

FDG: fluorodeoxyglucose-18

PET: positron emission tomography

* False negatives (FNs): judged as feasible for surgery based on imaging, with an incomplete debulking at surgery.

** False positives (FPs): judged as not feasible for surgery based on imaging, with a complete debulking at surgery.

- a. According to GRADE for sensitivity (false negatives (FNs)) and specificity (false positives (FPs)), respectively
- b. Numbers are calculated based on the results of the largest study (Shim 2015) at the mean prevalence of incomplete debulking (62%) of the two largest studies that addressed debulking with residual disease of any size (Michielsen 2017; Shim 2015). The prevalence of incomplete debulking was calculated as $(TP + FN)/\text{total study subjects}$ ($273/437 = 62\%$).
- c. Downgraded two levels for very wide confidence interval for number of FN's (sensitivity)
- d. Downgraded one level for wide confidence interval for number of FP's (specificity)
- e. Downgraded two levels as very small sample size; very wide confidence intervals for number of FN's (sensitivity) and number of FP's (specificity).
- f. Downgraded one level due to applicability concerns for the Index test since the radiologists were blinded for (presurgical) clinical data.
- g. Downgraded one level due to high risk of bias for patient selection and flow and timing.
- h. To compare the findings of the included studies (performing PET/CT or MRI to assess tumour resectability) with CT (the current gold standard), we provided the diagnostic accuracy of CT from the study with the best quality of evidence and with the target condition that is currently used in practice (Michielsen 2017).

BACKGROUND

Epithelial ovarian, fallopian tube, and primary peritoneal cancers are malignancies of the internal female genital tract. Clinically, these tumours are often regarded as a single entity, due to their similarity and overlap in pathophysiology, symptomatology, diagnostic approach, staging, treatment, and prognosis (Prat 2014). Globally, ovarian cancer affects 239,000 women each year (Ferlay 2012). It is most commonly identified at an advanced stage due to the absence of symptoms in early stage-disease. When symptoms do occur, they are often nonspecific and include abdominal pain or discomfort, bloating, and fatigue (Olson 2001). The extent of ovarian cancer is categorised using the International Federation of Gynaecology and Obstetrics (FIGO) staging criteria (Prat 2014). In advanced stage-disease, the tumour is not confined to the ovaries (stage I) or true pelvis (stage II), but has spread outside the pelvis through the peritoneal (abdominal) cavity or towards regional lymph nodes (stage III), or to extra-abdominal lymph nodes and/or with haematogenous spread resulting in distant metastasis (e.g. lungs or liver parenchyma, stage IV) (Mutch 2014; Prat 2014). This late presentation makes ovarian cancer the leading cause of death from gynaecological cancer in developed countries worldwide, with an absolute global mortality of 152,000 women each year (Ferlay 2012).

In women with advanced stage epithelial ovarian, fallopian tube, and primary peritoneal cancer, a combination of chemotherapy and debulking surgery is considered the mainstay of treatment. Debulking surgery (i.e. surgical efforts to remove the bulk of tumour) usually encompasses removal of the uterus (hysterectomy) and adnexa, resection of the omentum (an apron of fatty tissue attached to the greater curvature of the stomach, containing veins, arteries, lymphatics), and the attempted resection of all visible tumour deposits (NCI 2015). The actual feasibility of the latter, in reality, is limited by the location of lesions (e.g. around blood vessels) and the potential morbidity that each resection induces. At the end of each surgical procedure, a conclusion can be drawn on the completeness of debulking (cytoreductive) surgery, categorised into: no visible tumour deposits left (i.e. macroscopic ('complete') debulking); debulking with residual disease ≤ 1 cm (in the past often called 'optimal debulking'); or debulking with residual disease > 1 cm (i.e. incomplete debulking). This distinction is important since, along with tumour response to chemotherapy, the completeness of debulking surgery is the most important prognostic factor for survival in women with advanced stage epithelial ovarian cancer (Bristow 2002; Elattar 2011; NCI 2015; Vergote 2010). Unfortunately, despite chemotherapy and macroscopic debulking surgery, the majority of women still develop recurrent disease (Du Bois 2009). As 'macroscopic complete debulking' is determined by the naked eye of the surgeon, this does not imply that the resections are 'complete' in the sense of cancer-free surgical margins determined by histopathological examination of the specimen. Therefore, recurrences can be partly due to remaining microscopic disease (i.e. occult disease) after treatment.

Preoperative diagnostic imaging is used to estimate tumour extension and thus the feasibility of surgical debulking. If macroscopic debulking (removal of all visible tumour) seems feasible, based on imaging, primary debulking surgery is attempted. If imaging indicates that the chance of macroscopic debulking is small, women receive neoadjuvant chemotherapy

(in order to reduce tumour load) and subsequently debulking surgery (i.e. interval debulking). Currently, diagnostic imaging is predominantly based on abdominal computed tomography (CT). Unfortunately, this preoperative assessment is imperfect since small tumour deposits can be missed and distinguishing malignant from benign tissue can be challenging. This can lead to cases where primary surgery is attempted in which not all visible tumour can be removed. This causes unnecessary morbidity and negatively influences prognosis (Vergote 2010). In contrast, macroscopic debulking is the strongest independent predictor of patient outcome and should be attempted whenever deemed possible (Vergote 2010). Recent randomised controlled trials have demonstrated equivalence in survival between primary surgery and the alternative approach with neoadjuvant chemotherapy and interval debulking surgery, with reduced morbidity in the latter (Kehoe 2015; Morrison 2012; Vergote 2010).

Bristow 2002 demonstrated the extensive heterogeneity between centres in their percentage of macroscopically debulking and incomplete debulking with residual disease limited to 1 cm in diameter, or 2 cm in the earlier studies (Baker 1994), which ranged from 0% to 100% with a weighted mean of only 41.9%. Even with careful patient selection using laparoscopy, the percentage of women with residual tumour after primary debulking surgery still ranges up to 31% to 43% (Rutten 2017, Rutten 2014).

In conclusion, it is important to conscientiously select women for either primary debulking surgery with adjuvant chemotherapy or neoadjuvant chemotherapy followed by interval debulking. The aim should be to macroscopically debulk those women upfront who can be surgically resected and reduce surgical morbidity in those who cannot, who would benefit from chemotherapy first.

Target condition being diagnosed

The target condition is the outcome of primary debulking surgery for advanced stage epithelial ovarian, fallopian tube, and/or primary peritoneal cancer. The outcome is defined by the diameter of the largest tumour deposit remaining after surgery and is determined by the surgeon performing the procedure. The term 'primary' specifies those women in whom no treatment, surgical or chemotherapy, has been given prior to this surgery. Three target condition categories were considered.

- Macroscopic debulking, which was defined as no macroscopically visible tumour deposits at the end of surgery. Debulking of all deposits is the objective, though not always clinically feasible (NICE 2011). This can be due to their location (e.g. situated on the mesentery or liver hilum) or when the number of (small) metastases is innumerable (i.e. miliary pattern of spread). In general, deposit resection needs to be abandoned when continuing would induce unacceptable morbidity (e.g. compromising the blood supply to the entire small bowel in case of mesenteric resections). Consequently, this leads to an incomplete debulking with residual deposits of ovarian cancer.
- Incomplete debulking with visible residual disease, divided into two subcategories, depending on whether there were macroscopically visible tumour deposits:
 - ≤ 1 cm in diameter remaining at the end of surgery; or
 - > 1 cm in diameter remaining at the end of surgery.

Index test(s)

In this systematic review, we considered the following three noninvasive and commonly available index tests.

- Whole body fluorodeoxyglucose-18 (FDG) positron emission tomography (PET), with or without a parallel conventional abdominal CT for anatomical reference (PET-CT).
- Conventional T1w/T2w (i.e. anatomical) magnetic resonance imaging (MRI), with or without intravenously administered gadolinium contrast.
- Diffusion-weighted MRI (DW-MRI), in addition to conventional MRI, an imaging method that uses the diffusion of water molecules to generate contrast.

Clinical pathway

With (subtle) symptoms, or based on accidental discovery of an abdominal mass, women suspected of ovarian cancer preferably present to a gynaecological oncologist. Here, a standard diagnostic work-up is performed starting with obtaining information about medical history, symptoms, family history, known allergies, use of medication, and social background. This is followed by a general physical and pelvic examination (Roett 2009). In most centres, ultrasound (transvaginal and/or abdominal) is routinely added to assess the size and composition of the adnexal mass as well as the presence of free fluid in the rectouterine excavation (i.e. pouch of Douglas) (NICE 2011).

Blood tests are performed to assess both general health as well as specific tumour marker levels and a CT scan of the pelvis, abdomen and, optionally, the chest is performed (NICE 2011). The presence, location, and extent of the adnexal mass, ascites, peritoneal tumour deposits, omental caking (abnormally thickened greater omentum which indicates infiltration of tumour tissue), lymph node enlargement, pleural effusion and haematogenous metastases are specifically assessed. In some centres, chest CT is substituted by two-directional plain film chest radiography.

A multidisciplinary tumour board of experts discuss all findings and determine the diagnosis, stage and treatment plan, and, in particular, the feasibility of ('complete') tumour debulking. When considered feasible, primary debulking surgery followed by adjuvant chemotherapy is preferred. The tumour stage is macroscopically estimated at surgery and definitively after histopathological examination. When the feasibility of debulking surgery is questionable, women are commonly treated with three or six cycles of neoadjuvant chemotherapy (usually a combination of carboplatin and paclitaxel) and subsequently, in the case of no disease progression, with interval debulking surgery.

Alternative test(s)

Laparoscopy, performed either as ambulatory surgery or directly before the laparotomy, was considered as an alternative test. A Cochrane systematic review on laparoscopy for the assessment of tumour resectability in ovarian cancer remained inconclusive (Rutten 2014). However, a recent randomised controlled trial found that the number of incomplete debulking surgeries with residual disease > 1 cm in diameter can be reduced from 39% to 10% by performing diagnostic laparoscopy prior to debulking surgery (Rutten 2017).

Rationale

Abdominal CT is imperfect in assessing the (non-)resectability of advanced stage ovarian cancer in primary debulking surgery (Borley 2015; Suidan 2014, Vergote 2008). Alternative imaging options, such as PET(-CT), conventional and diffusion-weighted MRI, are currently widely available in the developed world and may possibly yield a superior diagnostic test accuracy (DTA) to assess preoperatively if macroscopic debulking can be achieved. First, PET(-CT) provides information on tumour extension, based on the enhanced glucose metabolism of cancer cells, and is particularly useful for identification of distant metastases. Second, MRI has good soft tissue image contrast and gives a detailed view of structures and its position towards surrounding tissue. These imaging tests can be added to the preoperative work-up (if the healthcare system permits with respect to costs), either as an alternative to abdominal CT (i.e. replacement test) or in combination with abdominal CT (i.e. as an add-on test). Adding an alternative imaging method can be considered in women with a tumour load determined resectable by abdominal CT, in an attempt to filter out false-negatives (i.e. resectable based on abdominal CT, not resectable according to the alternative method). In these women with non-resectable tumours, additional imaging studies such as MRI or PET(-CT) may possibly reduce the percentage of women with residual disease after primary debulking surgery. If PET(-CT) and/or MRI show superior accuracy, more adequate selection of women for either primary debulking or neoadjuvant chemotherapy can be performed.

Unfortunately, there is currently no systematic review which addresses the DTA of these imaging modalities (see; [Index test\(s\)](#)) in this context.

OBJECTIVES

To assess the diagnostic accuracy of fluorodeoxyglucose-18 (FDG) PET/CT, conventional and diffusion-weighted (DW) MRI as replacement or add-on to abdominal CT, for assessing tumour resectability at primary debulking surgery in women with stage III to IV epithelial ovarian/fallopian tube/primary peritoneal cancer.

Secondary objectives

To investigate the year of study initiation, the annual surgical caseload, and whether surgery is performed by a gynaecological oncologist as possible sources of heterogeneity. For further details, please see [Investigations of heterogeneity](#).

METHODS

Criteria for considering studies for this review

Types of studies

We included randomised comparisons of diagnostic tests, cross-sectional, retrospective and prospective cohort studies, that address the DTA of preoperative PET(-CT), conventional or (additional) diffusion-weighted MRI on assessing tumour resectability in women who are scheduled to undergo primary debulking surgery. Studies which added the index test(s) on to abdominal CT or when the index test replaced abdominal CT, were included. To evaluate the add-on effect, the alternative imaging test had to be performed within four weeks before or after the abdominal CT. Studies following a case-control design, which carry

an inherent high risk of bias in a DTA research objective, were excluded.

Participants

Studies had to include adult (18 years of age or more) women diagnosed with advanced stage (stage III to IV) epithelial ovarian/fallopian tube/primary peritoneal cancer, considered eligible for primary debulking surgery (i.e. no adjuvant chemotherapy treatment or prior surgery to assess tumour extension was performed). Also, studies with participants in stage I to IV disease were included if data from women with stage III to IV disease could be extracted.

Index tests

The index tests of interest were preoperatively performed fluorodeoxyglucose-18 PET(-CT), conventional and diffusion-weighted MRI (see; [Index test\(s\)](#)). All these imaging modalities were used as a replacement or as an add-on to abdominal CT in women with advanced epithelial ovarian/fallopian tube/primary peritoneal cancer.

A positive index test was defined as an assessment of tumour spread in which resection at primary debulking surgery was judged to be unfeasible (i.e. index test indicates ‘tumour is not resectable’) by the radiologist or multidisciplinary tumour board. Conversely, a negative index test was defined as a tumour for which resection by primary debulking surgery was considered feasible.

Target conditions

The target condition was defined as the resectability of all deposits from epithelial ovarian/fallopian tube/primary peritoneal cancer at primary debulking surgery. This target condition had three categories (see; [Target condition being diagnosed](#)) which makes two commonly studied and clinically relevant dichotomisations possible (see: [Statistical analysis and data synthesis](#)).

Reference standards

The reference standard was the process of debulking surgery. This is most commonly performed via a laparotomy, although in recent years laparoscopy has also been performed in cases of limited disease volume. During such a procedure, the abdomen is systematically explored to assess the tumour spread and its resectability. The outcome category (size of residual tumour after surgery) was determined by the surgeon at the end of this surgery.

Search methods for identification of studies

Our search for relevant literature involved both electronic databases (see [Electronic searches](#)) and additional sources (see [Searching other resources](#)).

Electronic searches

We searched MEDLINE (Ovid) and Embase (Ovid) systematically for potentially eligible studies. We did not use search filters (collections of terms aimed at reducing the number needed to screen) as an overall limiter because those published have not proved sensitive enough ([Beynon 2013](#)) and we applied no language restriction. The Medline search strategy was developed in conjunction with Cochrane Gynaecological, Neuro-oncology and Orphan Cancers and this along with the Embase strategy were executed by co-

author René Spijker who has extensive experience in systematic reviews.

- MEDLINE Ovid (January 1946 to 23 February 2017) ([Appendix 1](#)).
- Embase (January 1946 to 23 February 2017) ([Appendix 2](#))

Searching other resources

We searched both [ClinicalTrials.gov](#) ([Appendix 3](#)) and [WHO-ICTRP](#) ([Appendix 4](#)) to identify prospectively registered trials. Furthermore, the reference lists of all relevant studies were searched for additional relevant studies using [Web of Science](#).

Data collection and analysis

The data collection and analysis adhered to the guidelines provided in the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* ([Deeks 2013](#)).

Selection of studies

All titles and abstracts retrieved by electronic searching were downloaded into a reference management database and duplicates were removed. The remaining references were independently examined by two review authors (JFR and JPH) using the pre-set inclusion and exclusion criteria, as stated above. Afterwards, discrepancies in judgement between both review authors were discussed until consensus was reached. When the possible inclusion or exclusion of an individual study remained unclear, full-text assessment was independently performed by the same two review authors for a final decision. Articles considered directly eligible based on title and abstract screening were also read in full text to definitively confirm adherence to the inclusion and exclusion criteria. Excluded studies were documented and the reasons for exclusion were stated according to the guidance provided in the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*.

Data extraction and management

Two review authors (JFR and JPH) independently performed data extraction from the selected studies. Data were checked and entered into RevMan 5 by one review author and checked by another review author.

For the included studies, general information (title, aim of study, setting, study design, inclusion period), data on characteristics of women (inclusion criteria, exclusion criteria, age, FIGO stage, number of enrolled and eligible women) and index test (type, criteria to consider primary debulking unfeasible), outcomes and deviations from the protocol were abstracted onto a data abstraction form specially designed for the review (see [Characteristics of included studies](#)). We contacted the authors of the original studies in case of missing data.

Assessment of methodological quality

The QUADAS-2 assessment tool for diagnostic accuracy studies in the context of systematic reviews was completed for all included studies ([Whiting 2011](#)). This assessment was performed independently by two review authors (JFR and JPH) and final results were based on consensus discussion. Operational definitions of QUADAS-2 items were derived from [Rutten 2014](#) and are described in [Appendix 5](#).

Statistical analysis and data synthesis

We performed separate analyses for different target conditions based on the size of residual disease after debulking surgery (see: [Target condition being diagnosed](#)):

- Incomplete debulking with residual disease of any size (> 0 cm in diameter) versus macroscopic debulking.

- Debulking with residual disease > 1 cm versus residual disease ≤ 1 cm in diameter.

From each study, we extracted the numbers of true and false negatives and positives to calculate sensitivity and specificity. [Figure 1](#) and [Figure 2](#) outline the definitions of the two by two table for these analyses. [Figure 3](#) shows a visual representation of the 2 x 2 tables.

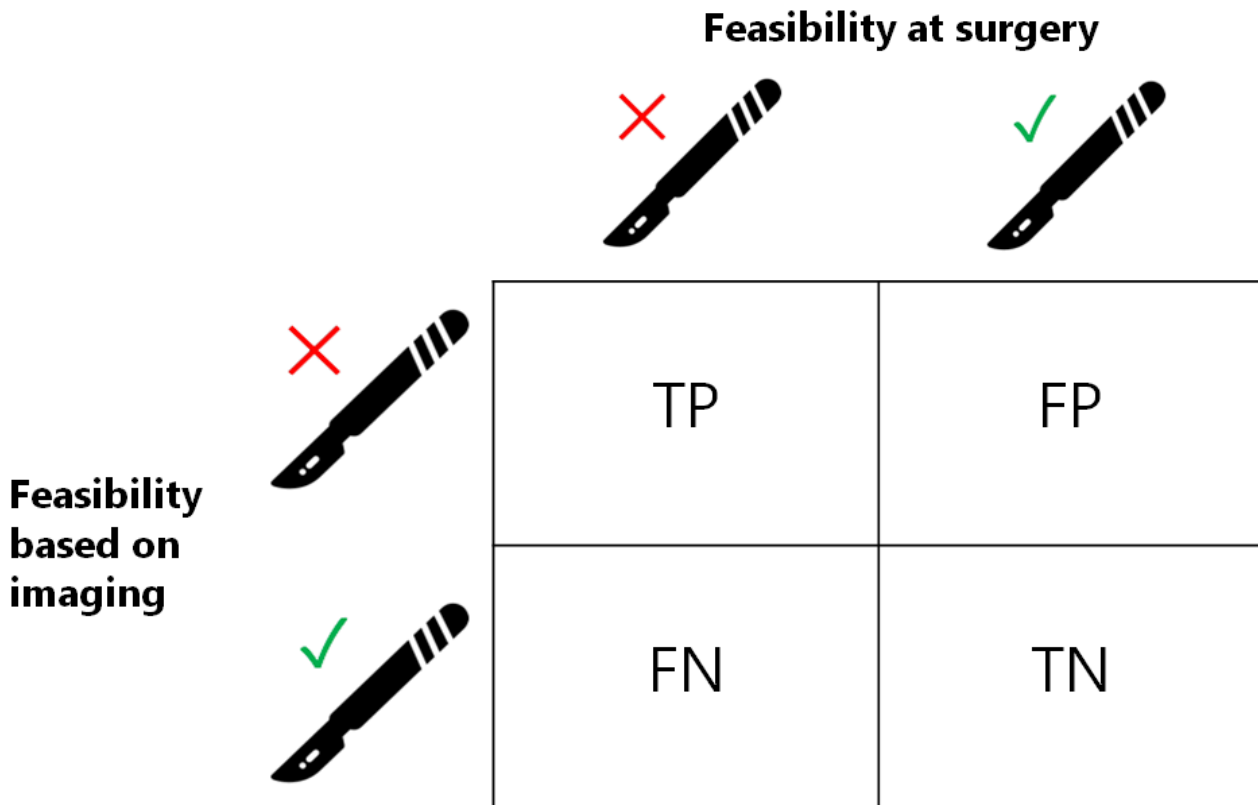
Figure 1. Definitions of the two by two table, wherein the index tests are tabulated against the reference standard outcome, on the analysis: macroscopic debulking versus incomplete debulking with residual disease of any size (i.e. consisting of deposits ≤ 1 cm and > 1 cm in diameter). TP = true positive, FP = false positive, FN = false negative, TN = true negative.

		Reference standard: primary cytoreductive surgery	
		Positive Incomplete resection with residual disease of any size	Negative Macroscopically complete resection
Index test: Judgment based on MRI or PET-CT	Positive Macroscopically complete resection is unfeasible (i.e. incomplete resection with residual disease of any size is anticipated)	TP: Macroscopically complete resection was <u>correctly</u> judged as <u>unfeasible</u> on MRI or PET(-CT) whereas surgery resulted in an incomplete resection with residual disease of any size	FP: Macroscopically complete resection was <u>incorrectly</u> judged as <u>unfeasible</u> on MRI or PET(-CT) whereas surgery resulted in a macroscopically complete resection
	Negative Macroscopically complete resection is feasible	FN: Macroscopically complete resection was <u>incorrectly</u> judged as <u>feasible</u> on MRI or PET(-CT) whereas surgery resulted in an incomplete resection with residual disease of any size	TN: Macroscopically complete resection was <u>correctly</u> judged as <u>feasible</u> on MRI or PET(-CT) whereas surgery resulted in a macroscopically complete resection

Figure 2. Definitions of the two by two table, wherein the index tests are tabulated against the reference standard outcome, on the analysis: macroscopic debulking or incomplete debulking with residual disease ≤ 1 cm in diameter versus incomplete resection with residual disease > 1 cm in diameter. TP = true positive, FP = false positive, FN = false negative, TN = true negative.

		Reference standard: primary cytoreductive surgery	
		Positive Incomplete resection with residual disease >1 cm in diameter	Negative Macroscopically complete resection or incomplete resection with residual disease ≤1 cm in diameter
Index test: Judgment based on MRI or PET-CT	Positive Macroscopically complete resection or incomplete resection with residual disease ≤1 cm in diameter is unfeasible (i.e. incomplete resection with residual disease >1 cm in diameter is anticipated)	TP: Macroscopically complete resection or incomplete resection with residual disease ≤1 cm was <u>correctly</u> judged as <u>un</u> feasible on MRI or PET(-CT) whereas surgery resulted in an incomplete resection with residual disease >1 cm in diameter	FP: Macroscopically complete resection or incomplete resection with residual disease ≤1 cm was <u>incorrectly</u> judged as <u>un</u> feasible on MRI or PET(-CT) whereas surgery resulted in a macroscopically complete resection or an incomplete resection with residual disease ≤1 cm
	Negative Macroscopically complete resection or incomplete resection with residual disease ≤1 cm in diameter is feasible	FN: Macroscopically complete resection or incomplete resection with residual disease ≤1 cm was <u>incorrectly</u> judged as feasible on MRI or PET(-CT) whereas surgery resulted in an incomplete resection with residual disease >1 cm in diameter	TN: Macroscopically complete resection or incomplete resection with residual disease ≤1 cm was <u>correctly</u> judged as feasible on MRI or PET(-CT) whereas surgery resulted in a macroscopically complete resection or an incomplete resection with residual disease ≤1 cm

Figure 3. Visual representation of 2 x 2 table. TP = true positive, FP = false positive, FN = false negative, TN = true negative.



We intended to perform analyses for the index tests as add-on tests in women who were considered resectable based on abdominal CT (CT 'negatives') to filter out women who were erroneously considered resectable by abdominal CT (false-negatives). We planned to perform meta-analyses according to the guidelines described in Chapter 10 of the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (Macaskill 2010). Unfortunately, we could not perform a meta-analysis for the three MRI studies since the target condition differed between these studies. For the PET-studies, we could not perform meta-analysis, because of the limited number of studies. [Review Manager 2014](#) was used to prepare forest plots of sensitivity and specificity of the included studies.

We assigned levels of evidence to the various outcome categories (true positive (TP), false positive (FP), false negative (FN) and true negative (TN), see [Figure 1](#) and [Figure 2](#)) according to GRADE and prepared 'Summary of findings' tables (Hsu 2011; Schünemann 2008). Labelling the tumour status erroneously as resectable ('false negatives') was considered worse than labelling the tumour status erroneously as non-resectable ('false positives'). For GRADE, therefore, the DTA outcome 'false negative' was deemed 'critical' (9) and the DTA outcome 'false positive' as less critical (8). The other outcomes (TP and TN) were considered 'important'. To create the GRADE profiles and 'Summary of findings' tables, we used [GRADEpro GDT](#).

Investigations of heterogeneity

We had planned to explore heterogeneity by adding covariates to the statistical model but the limited number of studies prevented this.

Sensitivity analyses

We had intended to perform sensitivity analyses by excluding studies at high risk of bias for each of the QUADAS-2 domains, but we were unable to do so due to too few studies.

Assessment of reporting bias

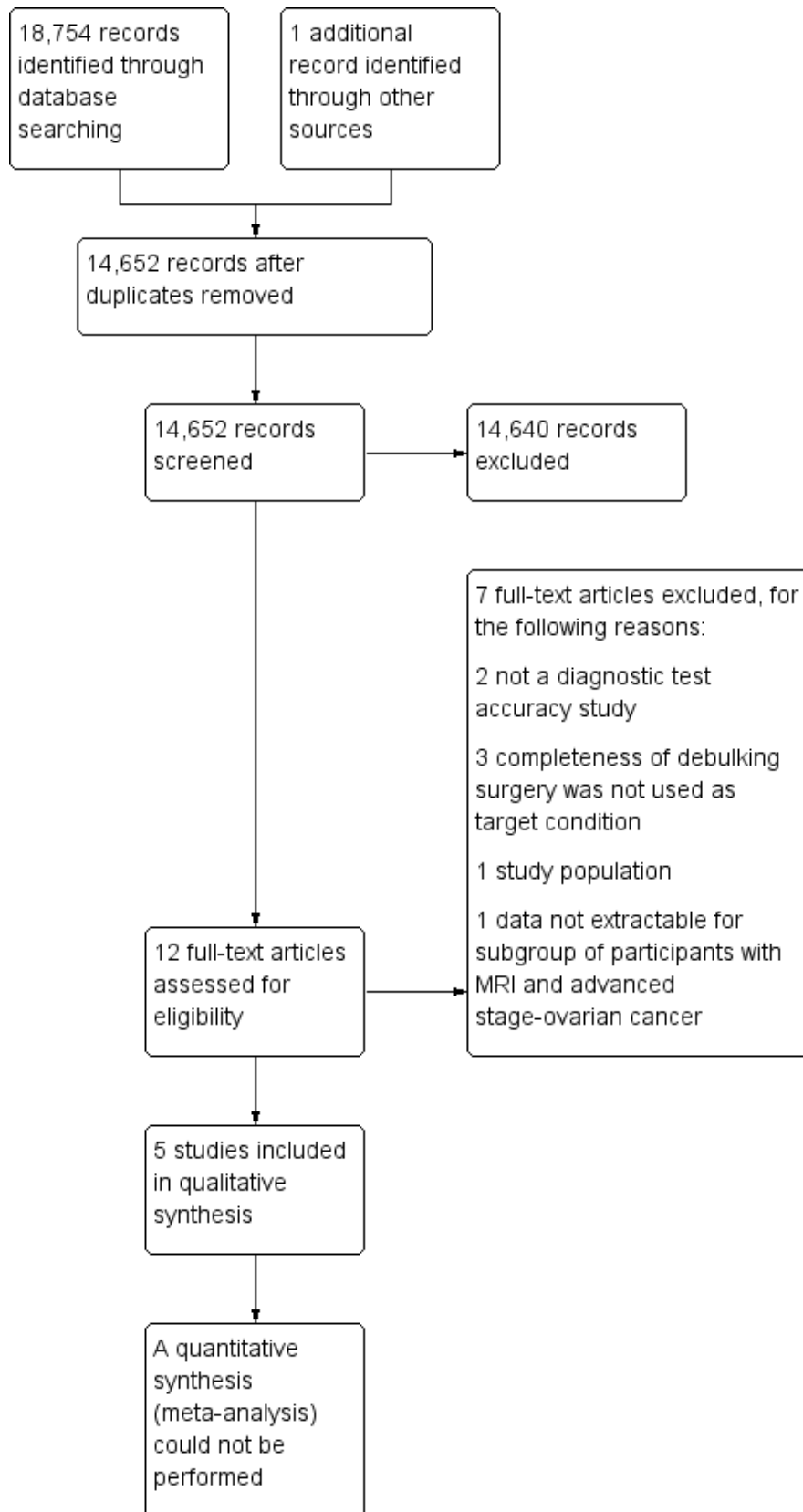
No assessment of reporting bias was performed. Currently, no uniformly accepted and validated method for assessing this type of bias, in the context of a review based on DTA studies, exists (Van Enst 2014).

RESULTS

Results of the search

Our search identified 7,101 citations in MEDLINE and 11,653 in Embase. After removing duplicates, 14,789 articles remained for title and abstract screening. A total of 11 articles were deemed potentially eligible and were reviewed in full text. Of these, seven did not meet the inclusion criteria and are listed in the [Characteristics of excluded studies](#) table, along with their reasons for exclusion. We included the four remaining articles in this review. Searching [ClinicalTrials.gov](#) and [WHO-ICTRP](#) revealed 119 and 64 additional trials, respectively, and, out of these, one additional study eligible for inclusion was identified. Reference checking with Web of Science revealed 160 citations, but no additional studies were found. Five studies were therefore finally included in this analysis. An overview of the search results is presented in [Figure 4](#).

Figure 4. Study flow diagram.



Results are presented separately for FDG-PET/CT and MRI in this review. We could not identify studies that addressed the accuracy of FDG-PET/CT and MRI as add-on tests to abdominal CT. Characteristics and quality assessments of the individual studies can be found in the [Characteristics of included studies](#) table.

FDG-PET/CT

Two studies investigated the accuracy of FDG-PET/CT for assessing tumour resectability ([Alessi 2016](#); [Shim 2015](#)).

The first study prospectively investigated the accuracy of FDG-PET/CT to assess the outcome of debulking surgery ([Alessi 2016](#)). The target condition was, after clarification was provided by the study authors, macroscopic debulking with no visible tumour remaining after surgery. In 29 consecutive women with an ovarian mass, total body FDG-PET/CT was performed within 20 days of debulking surgery. All women underwent explorative laparotomy. Where debulking was considered feasible, women received primary debulking surgery and the remaining women received neoadjuvant chemotherapy. Criteria to consider primary debulking unfeasible are summarised in [Table 1](#). Out of 29 women, 23 were diagnosed with ovarian cancer (of whom four had early stage-disease), and are included in our analysis.

The second study developed and validated a model to determine incomplete debulking with residual disease of any size in women with advanced stage ovarian cancer ([Shim 2015](#)). A total of 343 women were included and allocated to a development ($n = 240$) or validation ($n = 103$) cohort. All received primary debulking surgery. Women undergoing neoadjuvant chemotherapy, due to insufficient physical condition for surgery or presence of extra-abdominal disease, were excluded. The prediction model consisted of five FDG-PET/CT features (four anatomical structures, see [Table 1](#), and the tumour FDG uptake ratio) and one non-imaging related feature (an unvalidated surgical aggressiveness index). FDG-PET/CT was performed within four weeks of surgery.

MRI

Three studies addressed the accuracy of MRI for assessing tumour resectability ([Espada 2013](#); [Forstner 1995](#); [Michielsen 2017](#)). One study addressed conventional MRI ([Forstner 1995](#)), and two studies addressed DW-MRI ([Espada 2013](#); [Michielsen 2017](#)). Two of the studies also addressed the accuracy of abdominal CT ([Forstner 1995](#); [Michielsen 2017](#)).

The first study assessed the diagnostic accuracy of MRI in combination with diffusion-weighted imaging (DW-MRI) compared to explorative laparotomy for assessing incomplete debulking surgery in women with advanced stage ovarian cancer ([Espada 2013](#)). Surgery was performed by a gynaecological oncologist and incomplete debulking was defined as residual tumour > 1 cm in diameter. Within 15 days of surgery, 3-Tesla (DW-)MRI of the abdomen and pelvis was performed. Criteria to consider primary debulking surgery unfeasible are summarised in [Table 1](#). From the 36 recruited women, 34 were diagnosed with ovarian cancer and included in the analysis.

The second study prospectively evaluated ovarian cancer staging and tumour resectability with abdominal CT or conventional T1w/T2w MRI, or both ([Forstner 1995](#)). A total of 128 women were enrolled, of whom 82 received imaging by abdominal CT, MRI, or both. After inclusion, women with neoadjuvant chemotherapy, benign disease, other intra-abdominal malignancies or those who had undergone surgery more than one month after MRI were excluded from the statistical analysis ($n = 46$). In our analysis, data from the subgroup of 50 women with MRI were included, of whom 30 had FIGO stage III/IV ovarian cancer. The target condition was defined as debulking with residual disease < 2 cm in diameter. Criteria to consider whether primary debulking was unfeasible are summarised in [Table 1](#). MRI was performed within four weeks of surgery and all women received debulking surgery.

The third study compared the accuracy of abdominal CT and whole body DW-MRI to assess incomplete debulking with residual disease of any size in women with ovarian cancer ([Michielsen 2017](#)). This prospective study enrolled 126 women, of whom 94 were diagnosed with ovarian cancer and were eligible for analysis. All women received (primary or interval) debulking surgery, except for four women, who were physically unfit to undergo surgery. If surgery was considered unfeasible, a diagnostic laparoscopy was performed as a reference standard to confirm non-resectability. Criteria to consider whether primary debulking was unfeasible are summarised in [Table 1](#). Out of the 94 women with ovarian cancer, 73 had advanced stage (III or IV) disease. No details were provided on the time period between the index test and reference standard.

Methodological quality of included studies

The results of the QUADAS-2 assessments are presented in [Figure 5](#).

Figure 5. Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study

	Risk of Bias					Applicability Concerns			
	Patient Selection	Index Test: PET/CT	Index Test: MRI	Reference Standard	Flow and Timing	Patient Selection	Index Test: PET/CT	Index Test: MRI	Reference Standard
Alessi 2016	+	+		?	+	+	+		+
Espada 2013	?		+	+	+	?		-	+
Forstner 1995	-		?	+	-	?		+	?
Michielsen 2017	+		+	+	?	+		+	+
Shim 2015	+	+		+	+	+	+		+

- High
 ? Unclear
 + Low

FDG-PET/CT

The two FDG-PET/CT studies were considered at low risk of bias in all domains, except for the reference standard domain (Alessi 2016; Shim 2015). For both studies, it was unclear whether the surgery was performed by a gynaecologist or a specialised gynaecological oncologist and, for one study, it was not reported whether the outcome of debulking surgery was interpreted blind to the FDG-PET/CT results (Alessi 2016). There were no applicability concerns for both studies.

MRI

One study addressing DW-MRI was judged to be at low risk of bias in all domains, except for participant selection (Espada 2013). Enrolment procedures and exclusion criteria were not described, resulting in an unclear risk of bias for this domain. There were concerns about applicability for the Index test domain since the radiologists were blinded for (presurgical) clinical data contrary to standard practice. Furthermore, the applicability of participant selection remained unclear because no details were provided on the diagnostic assessment leading to their selection.

A second study addressing conventional T1w/T2w MRI was judged to have a high risk of bias in two domains (Forstner 1995). There were concerns about participant selection since their allocation to the imaging modality was decided on a variety of factors,

including the preference of the referring physician. Additionally, during its execution the study design and methodology were changed. The initial goal was to perform both abdominal CT and MRI in all women for an inpatient comparison. However, due to difficulties in participant recruitment, the study design was changed into a non randomised inter-participant comparison that required either abdominal CT or MRI imaging. From the initial 128 women recruited, 82 underwent both surgery and imaging and formed the study population. Women treated with neoadjuvant chemotherapy were excluded after enrolment. Consequently, the participant flow could have introduced bias. Applicability concerns for this study remained unclear for two domains. First, it was unclear whether all women were scheduled for debulking surgery after diagnostic assessment. Secondly, the study did not provide a clear definition of what was considered a positive result for the reference standard. However, in the discussion, the study authors specified that debulking was considered optimal when, after surgery, no tumour of > 2 cm remained, which was the standard for debulking surgery during the study period (1990 to 1994). Thirdly, with changing attitudes towards the goal of debulking surgery over the past two decades, from < 2 cm in the 1990s to no visible residual tumour nowadays, applicability concerns were present for this study.

In a third study investigating DW-MRI, the overall risk of bias was judged as low and there were no applicability concerns (Michielsen 2017). It remained unclear if the study flow and timing could have introduced bias because no information was provided on the time period between the index test and reference standard.

Findings

FDG-PET/CT

Two studies evaluated the accuracy of FDG-PET/CT to assess tumour resectability. The target condition was incomplete

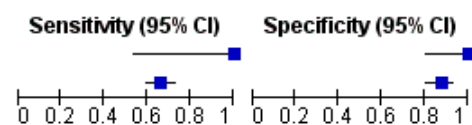
debulking with residual disease of any size (> 0 cm in diameter) versus macroscopic debulking. Definitions of true positives, false positives, false negatives, and true negatives are shown in (Figure 1).

In the first study, the prevalence of incomplete debulking was 6/23 (26%) (Alessi 2016). The sensitivity for assessing incomplete debulking (with residual disease of any size) of FDG-PET/CT was 1.00 (95% CI 0.54 to 1.00) and the specificity 1.00 (95% CI 0.80 to 1.00), as displayed in Figure 6.

Figure 6. Forest plot of tests: 1 PET/CT for assessing incomplete debulking with residual disease of any size, 4 MRI for assessing incomplete debulking with residual disease of any size, 2 MRI for assessing incomplete debulking with residual disease > 1 cm, 3 MRI for assessing incomplete debulking with residual disease > 2 cm.

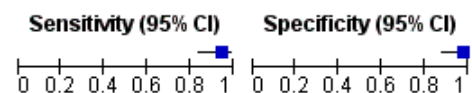
PET/CT for assessing incomplete debulking with residual disease of any size

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Alessi 2016	6	0	0	17	1.00 [0.54, 1.00]	1.00 [0.80, 1.00]
Shim 2015	148	15	75	105	0.66 [0.60, 0.73]	0.88 [0.80, 0.93]



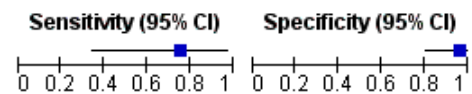
MRI for assessing incomplete debulking with residual disease of any size

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Michielsen 2017	47	1	3	43	0.94 [0.83, 0.99]	0.98 [0.88, 1.00]



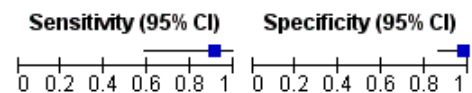
MRI for assessing incomplete debulking with residual disease > 1 cm

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Espada 2013	6	1	2	25	0.75 [0.35, 0.97]	0.96 [0.80, 1.00]



MRI for assessing incomplete debulking with residual disease > 2 cm

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Forstner 1995	10	1	1	38	0.91 [0.59, 1.00]	0.97 [0.87, 1.00]



The second FDG-PET/CT study used a prediction model including five FDG-PET/CT features and a surgical aggressiveness index to assess incomplete debulking with residual disease of any size in 343 women with ovarian cancer (Shim 2015). The study authors defined the high-risk group as having a predicted probability of incomplete debulking of greater than 80%. With this prediction model, the prevalence of incomplete debulking was 65% and 163 women would be classified as being unsuitable for debulking (positive index test), of whom 148 would have incomplete debulking with residual disease (positive reference standard). The sensitivity of FDG-PET/CT for incomplete debulking (with residual disease of any size) was 0.66 (95% CI 0.60 to 0.73) and the specificity 0.88 (95% CI 0.80 to 0.93).

If the very small first study (Alessi 2016) was ignored, the following results would apply to a hypothetical group of 1,000 women with an incomplete debulking prevalence of 62% (mean prevalence of Michielsen 2017 and Shim 2015): 211 women (95% CI 167 to 248) would be incorrectly classified as having no residual tumour (FNs) after surgery and 46 women (95% CI 27 to 76) would be incorrectly

classified as having residual disease (FPs) after surgery (Summary of findings 1).

MRI

All three MRI studies assessed the diagnostic accuracy of MRI for a different target condition. Figure 6 displays the paired forest plots of sensitivity and specificity for assessing incomplete debulking for the different target conditions.

The first study (Espada 2013) used a self-developed predictive score based on abdominal sites and tumour extension on DW-MRI to assess incomplete debulking with residual disease > 1 cm. Debulking was incomplete in 8 of 34 women (23.5%). A score ≥ 6 had the highest overall accuracy at 91%. The sensitivity for assessing incomplete debulking of DW-MRI was 0.75 (95% CI 0.35 to 0.97) and the specificity 0.96 (95% CI 0.80 to 1.00).

In the second study (Forstner 1995), 11 out of 50 women had incomplete debulking surgery with residual disease > 2 cm (22%). The sensitivity for assessing incomplete debulking on conventional MRI was 0.91 (95% CI 0.59 to 1.0) and the specificity 0.97 (95% CI 0.87

to 1.0). For abdominal CT, the sensitivity for assessing incomplete debulking was 0.50 (95% CI 0.12 to 0.88) and the specificity 1.0 (95% CI 0.91 to 1.0).

The third MRI study (Michielsen 2017) compared the diagnostic accuracy of DW-MRI and abdominal CT. From the 94 included women, 44 underwent primary debulking surgery. Macroscopic debulking was performed in 39 women (89%), two women had residual tumour < 1 cm, one woman had residual disease > 1 cm and two women were unfit for surgery. In the 50 remaining women, treated with neoadjuvant chemotherapy and interval debulking, non-resectability was confirmed with laparoscopy or biopsy from a distant metastasis. In this study, the prevalence of incomplete debulking with residual disease of any size was 53% and the sensitivity for assessing incomplete debulking (with residual disease of any size) of DW-MRI was 0.94 (95% CI 0.83 to 0.99) and the specificity 0.98 (95% CI 0.88 to 1.00). For abdominal CT, the sensitivity for assessing incomplete debulking was 0.66 (95% CI 0.52 to 0.78) and the specificity 0.77 (95% CI 0.63 to 0.87).

An overview of the results is provided in [Summary of findings 1](#).

DISCUSSION

Summary of main results

The aim of this systematic review was to determine the diagnostic accuracy of FDG-PET/CT and MRI for assessing incomplete debulking surgery in women with advanced stage epithelial ovarian cancer. We included five studies: two addressing FDG-PET/CT (Alessi 2016; Shim 2015); one conventional MRI (Forstner 1995) and two DW-MRI (Espada 2013; Michielsen 2017). Both FDG-PET/CT and MRI showed high specificity and moderate sensitivity (see [Summary of findings 1](#)). In a hypothetical group of 1000 women, of whom 620 would have incomplete debulking of any size (prevalence 62%), in 211 women (95% CI 167 to 248), surgery would incorrectly be considered feasible according to FDG-PET/CT and in 37 women (95% CI 6 to 105) according to MRI. However, the quality of evidence was very low to moderate according to GRADE, mainly due to the small sample sizes of the included studies.

In all studies, FDG-PET/CT or MRI were used as an initial test and the sensitivity and specificity were determined irrespective of abdominal CT results. Therefore, this review does not provide information on the accuracy of FDG-PET/CT or MRI as add-on tests to abdominal CT, only as its replacement. The two studies that addressed the accuracy of abdominal CT (Forstner 1995; Michielsen 2017) found low sensitivity and moderate specificity for assessing incomplete debulking of residual disease > 2 cm and of any size, respectively. A review comparing 10 studies that used abdominal CT-based models to assess residual disease showed a sensitivity ranging from 19.2% to 100% and specificity from 56.7% to 100% (Rutten 2015). This broad range can be explained by the different definitions used for the size of residual disease (e.g. the sensitivity and specificity can be different for assessing residual disease > 2 cm and > 0 cm).

In our included studies, the prevalence of incomplete debulking varied from 22% to 63%. This wide range may in part be due to changes in the goal of debulking surgery over the past decades, previously < 2 cm residual disease to the current standard of no macroscopically visible residual disease. Therefore, the prevalence

of incomplete debulking is likely to increase when the accepted size of remaining tumour after surgery decreases.

Two studies (Forstner 1995; Shim 2015) excluded women who had received neoadjuvant chemotherapy instead of primary debulking surgery. This exclusion has affected the numbers in the two by two tables, possibly leading to an underestimation of the sensitivity, since most of the women would have been considered unsuitable for surgery.

All studies used laparotomy as a reference standard and one study (Michielsen 2017) used also laparoscopy or biopsy from a suspect distant lesion as reference standards to confirm tumour non-resectability in women in which primary debulking was considered unfeasible. However, there might be ethical concerns with respect to operating on women where debulking surgery was considered not feasible.

While a number of studies have tried to identify specific radiologic predictors of incomplete debulking, no accepted universally validated scoring instrument exists. For clinical and study purposes, a standardised image-based instrument that can assess the feasibility of debulking surgery is desired. As a result, methodological heterogeneity exists between the included studies due to their different criteria on the feasibility of primary debulking (Table 1). The management of recurrent disease has also been widely investigated over the past years. A tool including performance status, the completeness of primary debulking, and the presence of ascites has been developed and used to assess the feasibility of secondary debulking in women with recurrent disease (Harter 2011). As the outcome of primary surgery is one of the criteria for this review, this tool cannot be used for assessing the feasibility of primary debulking.

Determining tumour resectability remains a complex and heterogeneous decision, since the feasibility of surgery depends not only on imaging results (which captures the dissemination pattern), but also on the experience and degree of specialisation of the surgeon, the institutional policies, patient's physical condition, and her personal preferences (e.g. willingness to risk a colostomy).

Strengths and weaknesses of the review

Our extensive search with comprehensive inclusion criteria yielded a large number of screened publications. However, only a small number of publications addressed the review question. Several studies performed analyses based on specific tumour sites but lacked an overall judgement on tumour (non)resectability (Hynninen 2013; Pfannenberger 2009; Risum 2008). Therefore, we could not perform meta-analyses or correct for possible sources of heterogeneity such as year of study initiation. We successfully contacted study authors for clarification on their study methods and results when details were missing. Unfortunately, the sample size of included studies limits the ability to draw robust conclusions and no studies addressing the accuracy of FDG-PET/CT or MRI as additional tests to abdominal CT were found. Also, as the accepted size of remaining tumours after surgery was different for the three MRI studies, it was impossible to estimate summary sensitivity and specificity. Another limitation of this review is that it is uncertain if in some studies the index test was used to exclude participants. This could have introduced bias in estimating the positive predictive value (PPV) women for which surgery was considered unfeasible by

the index test with residual disease after debulking) and, therefore, the focus should lie with the negative predictive value (NPV).

Applicability of findings to the review question

All studies addressed the diagnostic test accuracy for FDG-PET/CT or MRI as an initial test and showed sensitivity and specificity independent of abdominal CT results. Therefore, we were unable to provide information on the accuracy of the index tests as add-on tests to abdominal CT. The proposed study population of this review had ovarian cancer in an advanced stage. Nevertheless, from the included studies, four out of 366 women (1%) undergoing FDG-PET/CT and 26 out of 178 women (15%) undergoing MRI had early stage-disease at surgery. We decided to include these women in our analysis as this reflects clinical practice.

AUTHORS' CONCLUSIONS

Implications for practice

In women with advanced stage ovarian cancer, no firm conclusions can be drawn regarding the accuracy of FDG-PET/CT, conventional MRI, or (DW-)MRI to assess incomplete debulking surgery. FDG-PET/CT and MRI are commonly available in hospitals and they suggested there was a high specificity and moderate sensitivity to assess incomplete debulking. Potential advantages included the ability of FDG-PET/CT to detect extra-abdominal (distant) disease and the soft tissue contrast of MRI for (small) lesion detection.

Importantly, the level of evidence is insufficient to advise routine addition of FDG-PET/CT or MRI to clinical practice.

Implications for research

When a patient is suspected of ovarian cancer with extensive tumour load, it is difficult to judge tumour resectability based on abdominal CT alone. However, the size of tumour tissue remaining after surgery is one of the main prognostic factors in women

with ovarian cancer and necessitates careful patient selection for either primary debulking or neoadjuvant chemotherapy treatment. Therefore, additional tools are needed, ideally more accurate than abdominal CT and less invasive than laparoscopy.

Future research should focus on the additional value of FDG-PET/CT and MRI compared to abdominal CT in order to reduce the number of women with incomplete debulking. A cohort of women with advanced stage ovarian cancer for whom debulking surgery by abdominal CT is considered feasible could receive either FDG-PET/CT or MRI before primary debulking is performed. A radiologist that is blinded to the results of the other test could systematically score both imaging modalities, ideally by using a universally accepted and validated scoring system that has yet to be determined.

As previously described, it remains challenging to determine the feasibility of tumour resection since validated prediction models are lacking. Therefore, future research should focus on the construction and verification of predictive algorithms based on radiological findings and other predictors including biochemical parameters, tumour biopsies, and patient characteristics. Ideally, centre-specific features (e.g. the level of specialisation and annual caseload) should be incorporated as covariates.

ACKNOWLEDGEMENTS

We thank Jo Morrison for clinical and editorial advice, Jo Platt for designing the search strategy, and Clare Jess, Gail Quinn, and Tracey Harrison for their contribution to the editorial process.

This project was supported by the National Institute for Health Research, via Cochrane Infrastructure funding to the Cochrane Gynaecological, Neuro-oncology and Orphan Cancer Group. The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the Systematic Reviews Programme, NIHR, NHS or the Department of Health.

REFERENCES

References to studies included in this review
Alessi 2016 {published data only}

Alessi A, Martinelli F, Padovano B, Serafini G, Lorusso D, Lorenzoni A, et al. FDG-PET/CT to predict optimal primary cytoreductive surgery in patients with advanced ovarian cancer: preliminary results. *Tumori* 2016;**102**:103-7. [DOI: [10.5301/tj.5000396](https://doi.org/10.5301/tj.5000396)]

Espada 2013 {published data only}

Espada M, Garcia-Flores JR, Jimenez M, Alvarez-Moreno E, De Haro M, Gonzalez-Cortijo L, et al. Diffusion-weighted magnetic resonance imaging evaluation of intra-abdominal sites of implants to predict likelihood of suboptimal cytoreductive surgery in patients with ovarian carcinoma. *European Radiology* 2013;**23**:2636-42. [DOI: [10.1007/s00330-013-2837-7](https://doi.org/10.1007/s00330-013-2837-7)]

Forstner 1995 {published data only}

Forstner R, Hricak H, Occhipinti KA, Powell CB, Frankel SD, Stern JL. Ovarian cancer: staging with CT and MR imaging. *Radiology* 1995;**197**:619-26.

Michielsen 2017 {published data only}

Michielsen K, Dresen R, Vanslebrouck R, De Keyzer F, Amant F, Mussen E, et al. Diagnostic value of whole body diffusion-weighted MRI compared to computed tomography for pre-operative assessment of patients suspected for ovarian cancer. *European Journal of Cancer* 2017;**83**:88-98. [PUBMED: 28734146]

Shim 2015 {published data only}

Shim SH, Lee SJ, Kim SO, Kim SN, Kim DY, Lee JJ, et al. Nomogram for predicting incomplete cytoreduction in advanced ovarian cancer patients. *Gynecologic Oncology* 2015;**136**:30-6.

References to studies excluded from this review
Cotton 2006 {published data only}

Cotton F, Pellet O, Gilly FN, Granier A, Sournac L, Glehen O. MRI evaluation of bulky tumor masses in the mesentery and bladder involvement in peritoneal carcinomatosis. *European Journal of Surgical Oncology* 2006;**32**(10):1212-6. [PUBMED: 16762527]

Lopez-Lopez 2016 {published data only}

Lopez-Lopez V, Cascales-Campos PA, Gil J, Frutos L, Andrade RJ, Fuster-Quinonero M, et al. Use of (18)F-FDG PET/CT in the preoperative evaluation of patients diagnosed with peritoneal carcinomatosis of ovarian origin, candidates to cytoreduction and hipec. A pending issue. *European Journal of Radiology* 2016;**85**(10):1824-8. [PUBMED: 27666623]

Low 2012 {published data only}

Low RN, Barone RM. Combined diffusion-weighted and gadolinium-enhanced MRI can accurately predict the peritoneal cancer index preoperatively in patients being considered for cytoreductive surgical procedures. *Annals of Surgical Oncology* 2012;**19**(5):1394-401. [PUBMED: 22302265]

Pfannenbergs 2009 {published data only}

Pfannenbergs C, Konigsrainer I, Aschoff P, Oksuz MO, Zieker D, Beckert S, et al. (18)F-FDG-PET/CT to select patients with peritoneal carcinomatosis for cytoreductive surgery and hyperthermic intraperitoneal chemotherapy. *Annals of Surgical Oncology* 2009;**16**(5):1295-303. [PUBMED: 19252950]

Qayyum 2005 {published data only}

Qayyum A, Coakley FV, Westphalen AC, Hricak H, Okuno WT, Powell B. Role of CT and MR imaging in predicting optimal cytoreduction of newly diagnosed primary epithelial ovarian cancer. *Gynecologic Oncology* 2005;**96**(2):301-6. [PUBMED: 15661212]

Risum 2008 {published data only}

Risum S, Hogdall C, Loft A, Berthelsen AK, Hogdall E, Nedergaard L, et al. Prediction of suboptimal primary cytoreduction in primary ovarian cancer with combined positron emission tomography/computed tomography - a prospective study. *Gynecologic Oncology* 2008;**108**:265-70.

Risum 2011 {published data only}

Risum S, Loft A, Hogdall C, Berthelsen AK, Hogdall E, Lundvall L, et al. Standardized FDG uptake as a prognostic variable and as a predictor of incomplete cytoreduction in primary advanced ovarian cancer. *Acta Oncologica* 2011;**50**(3):415-9. [PUBMED: 20698810]

Additional references
Baker 1994

Baker TR, Piver MS, Hempling RE. Long term survival by cytoreductive surgery to less than 1 cm, induction weekly cisplatin and monthly cisplatin, doxorubicin, and cyclophosphamide therapy in advanced ovarian adenocarcinoma. *Cancer* 1994;**74**(2):656-63. [PUBMED: 8033045]

Beynon 2013

Beynon R, Leeflang MM, McDonald S, Eisinga A, Mitchell RL, Whiting P, et al. Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE. *Cochrane Database of Systematic Reviews* 2013, Issue 9. [DOI: [10.1002/14651858.MR000022.pub3](https://doi.org/10.1002/14651858.MR000022.pub3); PUBMED: 24022476]

Borley 2015

Borley J, Wilhelm-Benartzi C, Yazbek J, Williamson R, Bharwani N, Stewart V, et al. Radiological predictors of cytoreductive outcomes in patients with advanced ovarian cancer. *British Journal of Obstetrics and Gynaecology* 2015;**122**(6):843-9.

Bristow 2002

Bristow RE, Tomacruz RS, Armstrong DK, Trimble EL, Montz FJ. Survival effect of maximal cytoreductive surgery for advanced ovarian carcinoma during the platinum era: a meta-analysis. *Journal of Clinical Oncology* 2002;**20**(5):1248-59.

Deeks 2013

Deeks JJ, Bossuyt PM, Gatsonis C, (editors). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* Version 1.0.0. The Cochrane Collaboration, 2013. Available from srdta.cochrane.org/ 2013.

Du Bois 2009

Du Bois A, Reuss A, Pujade-Lauraine E, Harter P, Ray-Coquard I, Pfisterer J, Arbeitsgemeinschaft Gynaekologische Onkologie Studiengruppe Ovarialkarzinom (AGO-OVAR) and the Groupe d'Investigateurs Nationaux Pour les Etudes des Cancers de l'Ovaire (GINECO). Role of surgical outcome as prognostic factor in advanced epithelial ovarian cancer: a combined exploratory analysis of 3 prospectively randomized phase 3 multicenter trials. *Cancer* 2009;**115**(6):1234-44. [PUBMED: 19189349]

Earle 2006

Earle CC, Schrag D, Neville BA, Yabroff KR, Topor M, Fahey A, et al. Effect of surgeon specialty on processes of care and outcomes for ovarian cancer patients. *Journal of the National Cancer Institute* 2006;**98**(3):172-80.

Elattar 2011

Elattar A, Bryant A, Winter-Roach BA, Hatem M, Naik R. Optimal primary surgical treatment for advanced epithelial ovarian cancer. *Cochrane Database of Systematic Reviews* 2011, Issue 8. [DOI: [10.1002/14651858.CD007565.pub2](https://doi.org/10.1002/14651858.CD007565.pub2)]

Ferlay 2012

Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. globocan.iarc.fr (accessed 27 May 2015).

Harter 2011

Harter P, Sehouli J, Reuss A, Hasenburg A, Scambia G, Cibula D, et al. AGO Study Group. Prospective validation study of a predictive score for operability of recurrent ovarian cancer: the Multicenter Intergroup Study DESKTOP II. *International Journal of Gynecological Cancer* 2011;**21**(2):289-95. [PUBMED: 21270612]

Hsu 2011

Hsu J, Brozek JL, Terracciano L, Kreis J, Compalati E, Stein AT, et al. Application of GRADE: making evidence-based recommendations about diagnostic tests in clinical practice guidelines. *Implementation Science* 2011;**6**:62. [PUBMED: 21663655]

Hynninen 2013

Hynninen J, Kempainen J, Lavonius M, Virtanen J, Matomaki J, Oksa S, et al. A prospective comparison of integrated FDG-PET/contrast-enhanced CT and contrast-enhanced CT for pretreatment imaging of advanced epithelial ovarian cancer. *Gynecologic Oncology* 2013;**131**(2):389-94. [PUBMED: 23994535]

Kehoe 2015

Kehoe S, Hook J, Nankivell M, Jayson GC, Kitchener H, Lopes T, et al. Primary chemotherapy versus primary surgery for newly diagnosed advanced ovarian cancer (CHORUS): an open-label, randomised, controlled, non-inferiority trial. *Lancet* 2015;**386**(9990):249-57.

Macaskill 2010

Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C, (editors). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* Version 1.0.0. The Cochrane Collaboration, 2013. Available from srdta.cochrane.org.

Mercado 2010

Mercado C, Zingmond D, Karlan BY, Sekaris E, Gross J, Maggard-Gibbons M, et al. Quality of care in advanced ovarian cancer: the importance of provider specialty. *Gynecologic Oncology* 2010;**117**(1):18-22.

Morrison 2012

Morrison J, Haldar K, Kehoe S, Lawrie TA. Chemotherapy versus surgery for initial treatment in advanced ovarian epithelial cancer. *Cochrane Database of Systematic Reviews* 2012, Issue 8. [DOI: [10.1002/14651858.CD005343.pub3](https://doi.org/10.1002/14651858.CD005343.pub3)]

Mutch 2014

Mutch DG, Prat J. 2014 FIGO staging for ovarian, fallopian tube and peritoneal cancer. *Gynecologic Oncology* 2014;**133**(3):401-4.

NCI 2015

National Cancer Institute. PDQ® ovarian epithelial, fallopian tube, and primary peritoneal cancer treatment. www.cancer.gov/types/ovarian/hp/ovarian-epithelial-treatment-pdq (accessed 8 June 2015).

NICE 2011

National Institute for Health and Clinical Excellence (NICE). Ovarian cancer: the recognition and initial management of ovarian cancer. www.nice.org.uk/guidance/CG122/ (accessed prior to 16 September 2018).

Olson 2001

Olson SH, Mignone L, Nakraseive C, Caputo TA, Barakat RR, Harlap S. Symptoms of ovarian cancer. *Obstetrics and Gynecology* 2001;**98**(2):212-7.

Prat 2014

Prat J, FIGO Committee on Gynecologic Oncology. Staging classification for cancer of the ovary, fallopian tube, and peritoneum. *International Journal of Gynaecology and Obstetrics* 2014;**124**(1):1-5.

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.

Roett 2009

Roett MA, Evans P. Ovarian cancer: an overview. *American Family Physician* 2009 Sep;**80**(6):609-16.

Rutten 2014

Rutten MJ, Leeflang MM, Kenter GG, Mol BW, Buist M. Laparoscopy for diagnosing resectability of disease in patients with advanced ovarian cancer. *Cochrane Database of Systematic Reviews* 2014, Issue 2. [DOI: [10.1002/14651858.CD009786.pub2](https://doi.org/10.1002/14651858.CD009786.pub2)]

Rutten 2015

Rutten MJ, Van de Vrie R, Bruining A, Spijkerboer AM, Mol BW, Kenter GG, et al. Predicting surgical outcome in patients with International Federation of Gynecology and Obstetrics stage III or IV ovarian cancer using computed tomography: a systematic review of prediction models. *International Journal of Gynecological Cancer* 2015;**25**(3):407-15. [PUBMED: 25695545]

Rutten 2017

Rutten MJ, Van Meurs HS, Van de Vrie R, Gaarenstroom KN, Naaktgeboren CA, Van Gorp T, et al. Laparoscopy to predict the result of primary cytoreductive surgery in patients with advanced ovarian cancer: a randomized controlled trial. *Journal of Clinical Oncology* 2017;**35**(6):613-21. [PUBMED: 28029317]

Schrag 2006

Schrag D, Earle C, Xu F, Panageas KS, Yabroff KR, Bristow RE, et al. Associations between hospital and surgeon procedure volumes and patient outcomes after ovarian cancer resection. *Journal of the National Cancer Institute* 2006;**98**(3):163-71.

Schünemann 2008

Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, et al. GRADE: grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* 2008;**336**:1106-10.

Suidan 2014

Suidan RS, Ramirez PT, Sarasohn DM, Teitcher JB, Mironov S, Iyer RB, et al. A multicenter prospective trial evaluating the

ability of preoperative computed tomography scan and serum CA-125 to predict suboptimal cytoreduction at primary debulking surgery for advanced ovarian, fallopian tube, and peritoneal cancer. *Gynecologic Oncology* 2014;**134**(3):455-61.

Van Enst 2014

Van Enst WA, Ochodo E, Scholten RJ, Hooft L, Leeflang MM. Investigation of publication bias in meta-analyses of diagnostic test accuracy: a meta-epidemiological study. *BMC Medical Research Methodology* 2014;**14**:70. [DOI: [10.1186/1471-2288-14-70](https://doi.org/10.1186/1471-2288-14-70); PUBMED: 24884381]

Vergote 2008

Vergote I, Van Gorp T, Amant F, Leunen K, Neven P, Berteloot P. Timing of debulking surgery in advanced ovarian cancer. *International Journal of Gynecological Cancer* 2008;**18**(suppl 1):11-9.

Vergote 2010

Vergote I, Tropé CG, Amant F, Kristensen GB, Ehlen T, Johnson N, et al. Neoadjuvant chemotherapy or primary surgery in stage IIIC or IV ovarian cancer. *New England Journal of Medicine* 2010;**363**(10):943-53.

Whiting 2011

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011;**155**:529-36.

CHARACTERISTICS OF STUDIES

Characteristics of included studies [ordered by study ID]

Alessi 2016

Study characteristics	
Patient sampling	<p>Aim of the study: to investigate the role of PET(-CT) in characterisation of ovarian masses and identification of critical areas of tumour spread affecting results of debulking surgery</p> <p>Type of study: prospective study</p> <p>Enrolled/eligible: 29/23</p> <p>Inclusion period: 2013 to 2014</p>
Patient characteristics and setting	<p>Inclusion criteria: elevated serum CA125 and ultrasound detection of suspected ovarian malignancies</p> <p>Exclusion criteria: blood glucose levels > 140 mg/dL</p> <p>Mean age (range): 62 years (21 to 82)</p> <p>Setting: Gynaecologic Oncology Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy</p>
Index tests	Whole body FDG-PET/CT

Alessi 2016 (Continued)

	Criteria to consider primary debulking unfeasible: involvement of porta hepatis, diffuse deep infiltration of root mesentery, diffuse carcinomatosis requiring complete colectomy or more than 4 bowel resections or total gastrectomy, deep infiltration of pancreas and duodenum, multiple liver metastases
Target condition and reference standard(s)	<p>Target condition: debulking with no macroscopically visible tumour remaining after surgery</p> <p>Reference standard: all patients underwent explorative laparotomy and, where surgery was considered feasible, patients had primary debulking</p>
Flow and timing	<p>PET/CT was performed within 20 days of surgery. All patients received debulking surgery.</p> <p>23 out of 29 patients were diagnosed with ovarian cancer and were eligible for analysis.</p>
Comparative	
Notes	Four patients had stage IC disease, 14 stage IIIC and three stage IV so it seems that two patients were missing in the stage description (n = 23).

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Were the patients diagnosed by conventional diagnostic work-up for advanced stage cancer?	Yes		
Were the patients planned for primary debulking surgery after conventional diagnostic work-up?	Yes		
		Low	Low
DOMAIN 2: Index Test PET/CT			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Did the whole sample, or a random selection of the sample, receive verification using a reference standard of diagnosis?	Yes		
Did patients receive the same reference standard regardless of the index test result?	Yes		

Alessi 2016 (Continued)

Were the same clinical data available when test results were interpreted as would be available when the test is used in clinical practice?	Yes		
Did the study provide a clear definition of what was considered to be a 'positive' result for the index test?	Yes		
		Low	Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Is the surgeon's expertise adequate to perform the reference standard?	Unclear		
Did the study provide a clear definition of what was considered to be a 'positive' result for the reference standard?	Yes		
		Unclear	Low
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Were withdrawals from the study reported?	Yes		
		Low	

Espada 2013
Study characteristics

Patient sampling	Aim of the study: to analyse the diagnostic accuracy of diffusion-weighted MRI for predicting suboptimal cytoreductive surgery Type of study: prospective study Enrolled/eligible: 36/34 Inclusion period: 2006 to 2012
Patient characteristics and setting	Inclusion criteria: patients undergoing surgery for suspected ovarian carcinoma Exclusion criteria: none

Espada 2013 (Continued)

	<p>Mean age (SD): 53 years (11)</p> <p>Setting: Gynaecology Department, Hospital Universitario Quiron, Madrid, Spain</p>
Index tests	<p>Pelvic and abdominal diffusion-weighted MRI</p> <p>Criteria to consider primary debulking unfeasible: involvement of stomach, lesser sac, liver, small bowel mesentery, splenic hilum, para-aortic lymph nodes above level of renal vessels</p>
Target condition and reference standard(s)	<p>Target condition: optimal debulking with residual disease of maximal 1 cm in diameter</p> <p>Reference standard: primary debulking surgery</p>
Flow and timing	<p>MRI was performed within 15 days of surgery. All patients received debulking surgery.</p> <p>34 out of 36 patients had ovarian cancer and were eligible for analysis.</p>
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
Were the patients diagnosed by conventional diagnostic work-up for advanced stage cancer?	Unclear		
Were the patients planned for primary debulking surgery after conventional diagnostic work-up?	Yes		
		Unclear	Unclear
DOMAIN 2: Index Test MRI			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Did the whole sample, or a random selection of the sample, receive verification using a reference standard of diagnosis?	Yes		
Did patients receive the same reference standard regardless of the index test result?	Yes		

Espada 2013 (Continued)

Were the same clinical data available when test results were interpreted as would be available when the test is used in clinical practice? No

Did the study provide a clear definition of what was considered to be a 'positive' result for the index test? Yes

Low
High
DOMAIN 3: Reference Standard

Is the reference standards likely to correctly classify the target condition? Yes

Were the reference standard results interpreted without knowledge of the results of the index tests? Yes

Is the surgeon's expertise adequate to perform the reference standard? Yes

Did the study provide a clear definition of what was considered to be a 'positive' result for the reference standard? Yes

Low
Low
DOMAIN 4: Flow and Timing

Was there an appropriate interval between index test and reference standard? Yes

Did all patients receive the same reference standard? Yes

Were all patients included in the analysis? Yes

Were withdrawals from the study reported? Yes

Low
Forstner 1995
Study characteristics

Patient sampling

Aim of the study: to evaluate ovarian cancer staging and tumour resectability with abdominal CT or MRI

Type of study: prospective study

Enrolled/eligible: 128 were enrolled of whom 82 underwent abdominal CT or MRI. 50/82 patients underwent MRI and were included in our analysis.

Inclusion period: 1990 to 1994

Patient characteristics and setting

Inclusion criteria: patients suspected of ovarian cancer scheduled for surgical staging

Forstner 1995 (Continued)

Exclusion criteria: after inclusion, patients with neoadjuvant chemotherapy, benign disease, other intra-abdominal malignancies, or those who had undergone surgery more than one month after MRI were excluded from the statistical analysis (n = 46)

Mean age (range): 52 years (17 to 82)

Setting: Department of Gynecologic Oncology, University of California School of Medicine, San Francisco, America

Index tests	MRI and/or abdominal CT. Patients undergoing MRI (with or without abdominal CT) were included in our analysis. Criteria to consider primary debulking unfeasible: tumour larger than 2 cm at root of mesentery, porta hepatis, omentum of lesser sac, intersegmental fissure of the liver, gastrosplenic ligament, diaphragm, dome of liver, enlarged lymph nodes around coeliac axis, and pre-sacral extraperitoneal disease
Target condition and reference standard(s)	Target condition: debulking with residual disease < 2 cm Reference standard: primary debulking surgery
Flow and timing	MRI was performed within four weeks of surgery. All patients received debulking surgery.
Comparative	
Notes	Patient scheduling was based on a variety of factors, including scheduling availability, preference of referring physician, and contraindications to abdominal CT or MRI. Also, there was a change in study design. From the initial 128 recruited patients, 82 patients underwent surgery and imaging and formed the study population.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	No		
Were the patients diagnosed by conventional diagnostic work-up for advanced stage cancer?	Yes		
Were the patients planned for primary debulking surgery after conventional diagnostic work-up?	Unclear		
		High	Unclear
DOMAIN 2: Index Test MRI			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Forstner 1995 *(Continued)*

If a threshold was used, was it pre-specified?	Yes
Did the whole sample, or a random selection of the sample, receive verification using a reference standard of diagnosis?	Yes
Did patients receive the same reference standard regardless of the index test result?	Unclear
Were the same clinical data available when test results were interpreted as would be available when the test is used in clinical practice?	Yes
Did the study provide a clear definition of what was considered to be a 'positive' result for the index test?	Yes
Unclear Low	

DOMAIN 3: Reference Standard

Is the reference standards likely to correctly classify the target condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Is the surgeon's expertise adequate to perform the reference standard?	Yes
Did the study provide a clear definition of what was considered to be a 'positive' result for the reference standard?	Unclear
Low Unclear	

DOMAIN 4: Flow and Timing

Was there an appropriate interval between index test and reference standard?	Yes
Did all patients receive the same reference standard?	No
Were all patients included in the analysis?	No
Were withdrawals from the study reported?	Yes
High	

Michielsen 2017
Study characteristics

Patient sampling	Aim of the study: to evaluate whole body DW-MRI for diagnosis, staging, and operability assessment of patients suspected for ovarian cancer compared to abdominal CT Type of study: prospective study
------------------	--

Michielsen 2017 (Continued)

Enrolled/eligible: 167/94

Inclusion period: 2010 to 2013

Patient characteristics and setting	Inclusion criteria: - suspicion of ovarian cancer by clinical assessment, serum CA-125, carcinoembryonic antigen (CEA) and gynaecological ultrasound, and - staging by abdominal CT Exclusion criteria: contraindication for MRI Median age (range): 61 years (14 to 88) Setting: Department of Obstetrics and Gynaecology, University Hospitals, Leuven, Belgium
Index tests	Whole body diffusion-weighted MRI Criteria to consider primary debulking unfeasible: extra-abdominal distant metastasis, hepatic metastases, tumour infiltration of duodenum, stomach, pancreas, large vessels of coeliac trunk, hepatoduodenal ligament, metastases behind the portal vein, bowel involvement necessitating multiple bowel resections, deep tumoural involvement of superior mesenteric artery and root, retroperitoneal lymph node metastases above level of renal veins
Target condition and reference standard(s)	Target condition: debulking with no macroscopically visible tumour remaining after surgery Reference standard: explorative laparotomy, diagnostic laparoscopy or image-guided biopsy of surgical-critical distant lesions
Flow and timing	No information was provided about the time period between the index test and reference standard. All patients received (primary or interval) debulking surgery except for 4 patients who were medically unfit to undergo surgery. In patients where surgery was considered unfeasible, diagnostic laparoscopy was used as a reference standard to confirm irresectability.
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Were the patients diagnosed by conventional diagnostic work-up for advanced stage cancer?	Yes		

Michielsen 2017 (Continued)

Were the patients planned for primary debulking surgery after conventional diagnostic work-up? Yes

Low
Low
DOMAIN 2: Index Test MRI

Were the index test results interpreted without knowledge of the results of the reference standard? Yes

If a threshold was used, was it pre-specified? Yes

Did the whole sample, or a random selection of the sample, receive verification using a reference standard of diagnosis? Yes

Did patients receive the same reference standard regardless of the index test result? No

Were the same clinical data available when test results were interpreted as would be available when the test is used in clinical practice? Yes

Did the study provide a clear definition of what was considered to be a 'positive' result for the index test? Yes

Low
Low
DOMAIN 3: Reference Standard

Is the reference standards likely to correctly classify the target condition? Yes

Were the reference standard results interpreted without knowledge of the results of the index tests? Yes

Is the surgeon's expertise adequate to perform the reference standard? Yes

Did the study provide a clear definition of what was considered to be a 'positive' result for the reference standard? Yes

Low
Low
DOMAIN 4: Flow and Timing

Was there an appropriate interval between index test and reference standard? Unclear

Did all patients receive the same reference standard? No

Were all patients included in the analysis? Yes

Were withdrawals from the study reported? Yes

Unclear

Shim 2015

Study characteristics

Patient sampling	<p>Aim of the study: to develop a PET/CT-based nomogram for predicting incomplete cytoreduction in advanced-ovarian cancer patients.</p> <p>Type of study: retrospective study. A nomogram predicting incomplete debulking was constructed in a model development cohort (n = 240) and used in the validation cohort (n = 103).</p> <p>Enrolled/eligible: 343/343</p> <p>Inclusion period: 2006 to 2012</p>
Patient characteristics and setting	<p>Inclusion criteria: patients between 18 and 80 years with pathologically confirmed ovarian cancer FIGO stage III to IV undergoing cytoreductive surgery</p> <p>Exclusion criteria: patients receiving neoadjuvant chemotherapy, patients with history of other malignancies, and patients treated in another institute</p> <p>Median age (range): 55 years (27 to 80)</p> <p>Setting: Department of Obstetrics and Gynecology, Asan Medical Center, Seoul, Republic of Korea</p>
Index tests	<p>A nomogram including five FDG-PET/CT features: involvement of diaphragm, small bowel mesentery, presence of ascites, peritoneal carcinomatosis, and tumoral uptake ratio and one non-imaging related feature (an unvalidated surgical aggressiveness index)</p>
Target condition and reference standard(s)	<p>Target condition: macroscopic complete debulking</p> <p>Reference standard: primary debulking surgery</p>
Flow and timing	<p>PET/CT was performed within 4 weeks of surgery. Patients undergoing neoadjuvant chemotherapy (due to poor physical condition for surgery or presence of extra-abdominal disease) were excluded.</p>

Comparative

Notes

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Did the study avoid inappropriate exclusions?	Yes		

Shim 2015 (Continued)

Were the patients diagnosed by conventional diagnostic work-up for advanced stage cancer?	Yes		
Were the patients planned for primary debulking surgery after conventional diagnostic work-up?	Yes		
		Low	Low
DOMAIN 2: Index Test PET/CT			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Did the whole sample, or a random selection of the sample, receive verification using a reference standard of diagnosis?	Yes		
Did patients receive the same reference standard regardless of the index test result?	Yes		
Were the same clinical data available when test results were interpreted as would be available when the test is used in clinical practice?	Unclear		
Did the study provide a clear definition of what was considered to be a 'positive' result for the index test?	Yes		
		Low	Low
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Is the surgeon's expertise adequate to perform the reference standard?	Unclear		
Did the study provide a clear definition of what was considered to be a 'positive' result for the reference standard?	Yes		
		Low	Low
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Were withdrawals from the study reported?	Yes		

Shim 2015 *(Continued)*
Low

CEA: carcinoembryonic antigen
 CT: computed tomography
 FDG: fluorodeoxyglucose-18
 FIGO: International Federation of Gynaecology and Obstetrics
 PET: positron emission tomography

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

Study	Reason for exclusion
Cotton 2006	Other target condition. In patients with peritoneal carcinomatosis, MRI was used to evaluate tumour masses in the mesentery and bladder involvement.
Lopez-Lopez 2016	Study population: only one patient received primary debulking surgery after preoperative evaluation by PET-CT. Also, Peritoneal Carcinomatosis Index was used as target condition instead of the completeness of debulking surgery.
Low 2012	Other target condition. MRI was used to predict Peritoneal Cancer Index in patients being considered for cytoreductive surgery, of whom 5 were diagnosed with ovarian cancer.
Pfannenbergl 2009	Other target condition. Peritoneal Cancer Index was estimated using PET-CT to select patients for cytoreductive surgery or hyperthermic intraperitoneal chemotherapy, of whom 7 were diagnosed with ovarian cancer.
Qayyum 2005	Required data could not be extracted from the published article and was not provided by study authors.
Risum 2008	Not a diagnostic test accuracy study
Risum 2011	Not a diagnostic test accuracy study

CT: computed tomography
 MRI: magnetic resonance imaging
 PET: positron emission tomography

DATA

Presented below are all the data for all of the tests entered into the review.

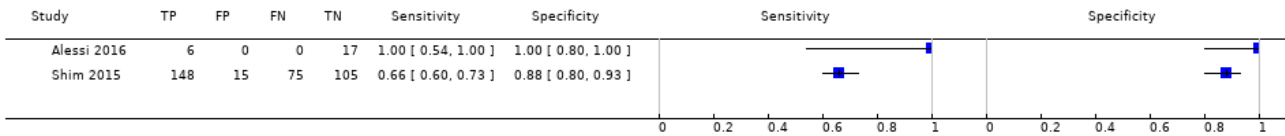
Table Tests. Data tables by test

Test	No. of studies	No. of participants
1 PET/CT for assessing incomplete debulking with residual disease of any size	2	366
2 MRI for assessing incomplete debulking with residual disease > 1 cm	1	34
3 MRI for assessing incomplete debulking with residual disease > 2 cm	1	50

Test	No. of studies	No. of participants
4 MRI for assessing incomplete debulking with residual disease of any size	1	94

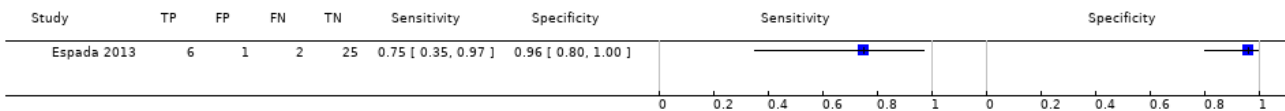
Test 1. PET/CT for assessing incomplete debulking with residual disease of any size.

Review: Positron emission tomography (PET) and magnetic resonance imaging (MRI) for assessing tumour resectability in advanced epithelial ovarian/fallopian tube/primary peritoneal cancer
Test: 1 PET/CT for assessing incomplete debulking with residual disease of any size



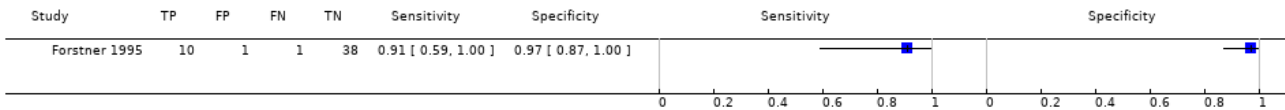
Test 2. MRI for assessing incomplete debulking with residual disease > 1 cm.

Review: Positron emission tomography (PET) and magnetic resonance imaging (MRI) for assessing tumour resectability in advanced epithelial ovarian/fallopian tube/primary peritoneal cancer
Test: 2 MRI for assessing incomplete debulking with residual disease > 1 cm



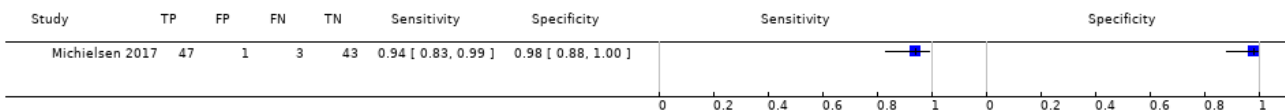
Test 3. MRI for assessing incomplete debulking with residual disease > 2 cm.

Review: Positron emission tomography (PET) and magnetic resonance imaging (MRI) for assessing tumour resectability in advanced epithelial ovarian/fallopian tube/primary peritoneal cancer
Test: 3 MRI for assessing incomplete debulking with residual disease > 2 cm



Test 4. MRI for assessing incomplete debulking with residual disease of any size.

Review: Positron emission tomography (PET) and magnetic resonance imaging (MRI) for assessing tumour resectability in advanced epithelial ovarian/fallopian tube/primary peritoneal cancer
Test: 4 MRI for assessing incomplete debulking with residual disease of any size



ADDITIONAL TABLES

Table 1. Criteria to consider primary debulking unfeasible

Criteria to consider primary debulking unfeasible according to study methods	Alessi	Shim	Espada	Forstner	Michielsen
Site of tumour involvement					

Table 1. Criteria to consider primary debulking unfeasible (Continued)

Liver/porta hepatis	Yes	No	No	Yes	Yes
Mesentery	Yes	Yes	Yes	Yes	No
Colon	Yes, when necessitating > 4 bowel resections	No	No	No	Yes, when necessitating multiple bowel resections
Stomach	Yes	No	Yes	No	Yes
Pancreas	Yes	No	No	No	Yes
Duodenum	Yes	No	No	No	Yes
Diaphragm	No	Yes	No	Yes	No
Ascites	No	Yes	No	No	No
Peritoneal carcinomatosis	Yes	Yes	No	No	No
Lesser sac/bursa omentalis	No	No	Yes	Yes	No
Spleen/splenic hilum	No	No	Yes	No	No
Lymph nodes above level of renal vessels/at coeliac axis	No	No	Yes	Yes	Yes
Gastrosplenic ligament	No	No	No	Yes	No
Presacral extraperitoneal disease	No	No	No	Yes	No
Extra-abdominal distant metastasis	No	No	No	No	Yes
Vessels of coeliac trunk	No	No	No	No	Yes
Hepatoduodenal ligament	No	No	No	No	Yes
Superior mesenteric artery	No	No	No	No	Yes

Yes: site of tumour involvement is selected as one of the criteria to consider primary debulking unfeasible

No: site of tumour involvement is not selected as a criterion to consider primary debulking unfeasible

APPENDICES

Appendix 1. MEDLINE search strategy

Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to 2017 February 23rd

Searches

1 exp Ovarian Neoplasms/

2 Fallopian Tube Neoplasms/

3 Peritoneal Neoplasms/

4 ((ovar* or fallopian* or peritone*) adj5 (cancer* or neoplasm* or carcin* or cystadenocarcinoma* or malign* or tumo?r*)).ti,ab,kw,kf.

5 1 or 2 or 3 or 4

6 exp MAGNETIC RESONANCE IMAGING/
 7 (MRI or MRi or NMRI or NMRI).ti,ab,kw,kf.
 8 ((magn* or MR or MTC or MT or NMR or spin or chemical shift or diffus*) adj3 (imag* or scan* or resonance* or tomogra\$)).ti,ab,kw,kf.
 9 Diffusion-weighted.ti,ab,kw,kf.
 10 exp POSITRON-EMISSION TOMOGRAPHY/
 11 (pet adj3 scan*).ti,ab,kw,kf.
 12 (positr* adj4 tomogr*).ti,ab,kw,kf.
 13 (pet-ct or petct or fdg-pet).ti,ab,kw,kf.
 14 (CT adj3 (cine or scan* or x-ray* or xray*)).ab,ti,kw,kf.
 15 (ct or mdct).ti.
 16 ((electron beam* or comput* or axial) adj3 tomography).ab,ti,kw,kf.
 17 tomodensitometry.ab,ti,kw,kf.
 18 exp TOMOGRAPHY, X-RAY COMPUTED/
 19 radiography.fs.
 20 radionuclide imaging.fs.
 21 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20
 22 5 and 21

Appendix 2. Embase search strategy

Embase Classic+Embase 1946 to 2017 February 23rd

1 ((exp ovary tumor/ or uterine tube tumor/ or exp peritoneum tumor/) or (((ovar* or fallopian* or peritone*) adj5 (cancer* or neoplasm* or carcin* or cystadenocarcinoma* or malign* or tumo?*)).ti,ab,kw.))
 2 (exp nuclear magnetic resonance imaging/ or (MRI or MRi or NMRI or NMRI or Diffusion-weighted or ((magn* or MR or MTC or MT or NMR or spin or chemical shift or diffus*) adj3 (imag* or scan* or resonance* or tomogra\$))).ti,ab,kw.)
 3 ((positron emission tomography/ or exp computer assisted tomography/ or computer assisted emission tomography/) or ((pet adj3 scan*) or (positr* adj4 tomogr*) or (pet-ct or petct or fdg-pet) or (CT adj3 (cine or scan* or x-ray* or xray*))) or ((electron beam* or comput* or axial) adj3 tomography) or tomodensitometry).ti,ab,kw or (ct or mdct).ti.)
 4 1 AND (2 OR 3)

Appendix 3. Clinicaltrials.gov search strategy

ClinicalTrials.gov search strategy

(MRI OR MRi OR NMRI OR NMRI OR Diffusion-weighted OR magnetic imaging OR chemical-shift OR pet-ct or petct or fdg-pet OR PET-scan OR CT-scan) | ((ovarian OR ovary OR fallopian OR peritoneal) AND (cancer OR neoplasm OR carcinoma OR cystadenocarcinoma OR malignant OR malignancy OR tumor OR tumour))

Appendix 4. ICTRP search strategy

ICTRP search strategy

ovarian cancer AND MRI OR "fallopian cancer" AND MRI OR "ovarian tumor" AND MRI OR "ovarian tumour" AND MRI OR "peritoneal cancer" AND MRI OR "peritoneal tumor" AND MRI OR "peritoneal tumour" AND MRI OR "ovarian neoplasm" AND MRI OR "ovarian carcinoma" AND MRI
 OR
 ovarian cancer AND "magnetic imaging" OR "fallopian cancer" AND "magnetic imaging" OR "ovarian tumor" AND "magnetic imaging" OR "ovarian tumour" AND "magnetic imaging" OR "peritoneal cancer" AND "magnetic imaging" OR "peritoneal tumor" AND "magnetic imaging" OR "peritoneal tumour" AND "magnetic imaging" OR "ovarian neoplasm" AND "magnetic imaging" OR "ovarian carcinoma" AND "magnetic imaging"
 OR
 ovarian cancer AND "Diffusion-weighted" OR "fallopian cancer" AND "Diffusion-weighted" OR "ovarian tumor" AND "Diffusion-weighted" OR "ovarian tumour" AND "Diffusion-weighted" OR "peritoneal cancer" AND "Diffusion-weighted" OR "peritoneal tumor" AND "Diffusion-weighted" OR "peritoneal tumour" AND "Diffusion-weighted" OR "ovarian neoplasm" AND "Diffusion-weighted" OR "ovarian carcinoma" AND "Diffusion-weighted"
 OR
 ovarian cancer AND "chemical-shift" OR "fallopian cancer" AND "chemical-shift" OR "ovarian tumor" AND "chemical-shift" OR "ovarian tumour" AND "chemical-shift" OR "peritoneal cancer" AND "chemical-shift" OR "peritoneal tumor" AND "chemical-shift" OR "peritoneal tumour" AND "chemical-shift" OR "ovarian neoplasm" AND "chemical-shift" OR "ovarian carcinoma" AND "chemical-shift"
 OR
 ovarian cancer AND "CT-scan" OR "fallopian cancer" AND "CT-scan" OR "ovarian tumor" AND "CT-scan" OR "ovarian tumour" AND "CT-scan" OR "peritoneal cancer" AND "CT-scan" OR "peritoneal tumor" AND "CT-scan" OR "peritoneal tumour" AND "CT-scan" OR "ovarian neoplasm" AND "CT-scan" OR "ovarian carcinoma" AND "CT-scan"

OR
ovarian cancer AND "pet-scan" OR "fallopian cancer" AND "pet-scan" OR "ovarian tumor" AND "pet-scan" OR "ovarian tumour" AND "pet-scan" OR "peritoneal cancer" AND "pet-scan" OR "peritoneal tumor" AND "pet-scan" OR "peritoneal tumour" AND "pet-scan" OR "ovarian neoplasm" AND "pet-scan" OR "ovarian carcinoma" AND "pet-scan"

OR
ovarian cancer AND "fdg-pet" OR "fallopian cancer" AND "fdg-pet" OR "ovarian tumor" AND "fdg-pet" OR "ovarian tumour" AND "fdg-pet" OR "peritoneal cancer" AND "fdg-pet" OR "peritoneal tumor" AND "fdg-pet" OR "peritoneal tumour" AND "fdg-pet" OR "ovarian neoplasm" AND "fdg-pet" OR "ovarian carcinoma" AND "fdg-pet"

OR
ovarian cancer AND "pet-ct" OR "fallopian cancer" AND "pet-ct" OR "ovarian tumor" AND "pet-ct" OR "ovarian tumour" AND "pet-ct" OR "peritoneal cancer" AND "pet-ct" OR "peritoneal tumor" AND "pet-ct" OR "peritoneal tumour" AND "pet-ct" OR "ovarian neoplasm" AND "pet-ct" OR "ovarian carcinoma" AND "pet-ct"

Appendix 5. Operational definitions of QUADAS-2 items

Risk of bias		Applicability	
Quality indicator	Notes	Quality indicator	Notes
Domain 1	Could the selection of patients have introduced bias? (High/low/unclear)	Are there concerns that the included patients and settings do not match the review question? (High/low/unclear)	
Patient Selection	<p>1. Was a consecutive or random sample of patients enrolled?</p> <p>'Yes' if a consecutive or random sample of patients was enrolled.</p> <p>'No' if a selected group of patients was enrolled.</p> <p>'Unclear' if there was insufficient information on enrolment.</p>	<p>1. Were the patients diagnosed by conventional diagnostic work-up for advanced stage ovarian cancer?</p> <p>1. Were the patients diagnosed by conventional diagnostic work-up for advanced stage ovarian cancer?</p>	<p>'Yes' if patients were diagnosed by conventional diagnostic work-up with advanced stage-ovarian cancer.</p> <p>'No' if patients included in the trial were diagnosed with low stage-disease (FIGO I or II) only. No high stage-disease patients in the trial.</p> <p>'Unclear' if there was insufficient information on recruitment method, criteria for diagnosis of ovarian cancer.</p>
	<p>2. Did the study avoid inappropriate exclusions?</p> <p>'Yes' if there were no inappropriate exclusions.</p> <p>'No' if there were inappropriate exclusions.</p> <p>'Unclear' if there was insufficient information on exclusions.</p>	<p>2. Were the patients scheduled for primary debulking surgery after conventional diagnostic work-up?</p>	<p>'Yes' if the patients were scheduled for primary debulking surgery after conventional diagnostic work-up.</p> <p>'No' if none of the patients were scheduled for primary debulking surgery.</p> <p>'Unclear' if there was insufficient information.</p>
Domain 2	Could the interpretation of the Index test have introduced bias? (High/low/unclear)	Were there concerns that the index test, its conduct, or the interpretation differed from the review question? (High/low/unclear)	
Index Test	<p>1. Were the index test results interpreted without the knowledge of the results of the reference standard?</p> <p>This will always be rated as 'yes', because the index test is performed before the reference standard.</p>	<p>1. Were the same clinical data available when test results were interpreted as would</p>	<p>'Yes' if all usual clinical data (except laparotomy results) were available when the index test was interpreted, including details of physical examination, serum tumour markers, ultrasound, and CT/MRI imaging.</p>

(Continued)			
		be available when the test is used in clinical practice?	Also answer 'yes' if one of the items was missing. 'No' if clinical information (as mentioned by 'yes') was not available to the gynaecologist. 'Unclear' if insufficient information was reported.
	2. Was the threshold used pre-specified?	'Yes' if a clear description of the threshold was given which was specified before start of the study. 'No' if no clear description was given beforehand. 'Unclear' if there was insufficient information within the paper to determine whether or not a prespecified threshold was used.	2. Did the study provide a clear definition of what was considered to be a 'positive' result for the index test? 'Yes' if a clear description was given about when the index test was positive or negative (e.g. what the cut-off for too extensive abdominal disease was). 'No' if there was no clear description given about what was classified as too extensive disease or not. 'Unclear' if there was insufficient information within the paper to determine whether or not a defined threshold was used for a positive test result.
	3. Did the whole sample, or a random selection of the sample, receive verification using a reference standard of diagnosis?	'Yes' if all patients underwent the reference standard (laparotomy). 'No' if not all patients underwent reference standard. 'Unclear' if insufficient information was provided.	
	4. Did patients receive the same reference standard regardless of the index test result?	'Yes' if patients who underwent the reference standard had laparotomy. 'No' if patients did not undergo laparotomy. 'Unclear' if insufficient information was provided.	
Domain 3	Could the interpretation of the reference standard have introduced bias? (High/low/unclear)		Were there concerns that the target condition as defined by the reference standard did not match the question? (High/low/unclear)
Reference Standard	1. Was the reference standard likely to correctly classify the target condition?	'Yes' if the reference standard was laparotomy. 'No' if the reference standard used was not the one defined in the protocol.	1. Did the study provide a clear definition of what 'Yes' if a clear description was given about when the reference standard was positive or negative (e.g. if description was given

(Continued)	'Unclear' if the information was insufficient.	was considered to be a 'positive' result for the reference standard?	about the size of the tumour deposits left after surgery). 'No' if there was no clear description of tumour deposit size after surgery. 'Unclear' if there was insufficient information within the paper that described tumour size after surgery.
2. Were the reference standard results interpreted without the knowledge of the results of the index test?	'Yes' if the report stated that the reference test was performed by individuals who did not perform the index test. 'No' if the reference test was done by the same person performing the index test. 'Unclear' if not reported.		
3. Was the surgeon's expertise adequate to perform the reference standard?	'Yes' if the reference test was performed by a gynaecological oncologist. 'No' if the reference test was not performed by a gynaecological oncologist. 'Unclear' if not reported.		
Domain 4	Could the patient flow have introduced bias? (High/low/unclear)		
Flow and Timing			
1. Was the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	'Yes' if the time period between the index test and reference standard did not extend 6 weeks. 'No' if the time period was more than 6 weeks for an unacceptably high proportion of patients. 'Unclear' if the information on the timing of tests was not provided.		
2. Did all patients receive the same reference standard?	'Yes' if all patients underwent the reference standard (laparotomy, diagnostic laparoscopy or image-guided biopsy of distant metastases).		

(Continued)

	'No' if not all patients underwent the reference standard.
	'Unclear' if insufficient information was provided.
3. Were all patients included in the analysis?	'Yes' if for all patients entered in the study were included in the analysis.
	'No' if not all the patients in the study were included in the analysis.
	'Unclear' if it was not clear whether all patients were accounted for.
4. Were withdrawals from the study reported?	'Yes' if, for all patients entered in the study, it was reported what happened
	during the study, also those who withdrew or answered 'Yes' if no withdrawals were reported, and results were reported for all patients who entered in the study.
	'No' if not all the patients in the study completed the study and these patients were not accounted for.
	'Unclear' if it was not clear whether all patients were accounted for.

CONTRIBUTIONS OF AUTHORS

The review was written by JFR, JPH, RS, RJPMS and RP and subsequently revised by all authors. In addition, RS conducted the search strategy for this review and performed the search. JFR and JPH assessed titles and abstracts for inclusion and the quality of included studies. JFR and RJPMS extracted data, performed the statistical analyses and assigned level of evidence of the included studies.

DECLARATIONS OF INTEREST

JFR: no conflicts of interest relevant to the presented research.

JPH: no conflicts of interest relevant to the presented research.

FTW: no conflicts of interest relevant to the presented research.

RS: no conflicts of interest relevant to the presented research.

LV: no conflicts of interest relevant to the presented research.

JV: no conflicts of interest relevant to the presented research.

WBV: no conflicts of interest relevant to the presented research.

RJPMS: his institution has received grant funding from the Belgian Health Care Knowledge Centre for work related to this review. His institution has also received funding from the WHO and World Federation of Haemophilia for travel and consultancy unrelated to this review.

RPZ: no conflicts of interest relevant to the presented research.

SOURCES OF SUPPORT

Internal sources

- No sources of support supplied

External sources

- KCE Belgian Healthcare Knowledge Centre, Belgium.

Practical support during all stages of the review process.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

Compared to the review protocol, three minor adjustments have been made in the review process.

First of all, when we performed the electronic searches, we expected no additional records in the Science Citation Index Expanded (Web of Knowledge), Social Sciences Citation Index (Web of Knowledge), and Arts & Humanities Citation Index (Web of Knowledge). Therefore, we did not perform a separate search of these sources.

Secondly, due to possible confusion about terminology, we decided to define the outcome (target condition) as the size of residual disease and not on completeness of the debulking surgery (macroscopic 'complete', optimal, or incomplete).

Thirdly, for the statistical analysis, we planned to explore heterogeneity by adding covariates to the statistical model, but the limited number of studies prevented this.

The following covariates would have been considered.

- Year of study initiation. Rapid advances have been made over the past decade(s) in the imaging sciences. Thus, heterogeneity caused by time-dependent qualitative differences in the index test would be explored by adding the year of study initiation to the model.
- Annual caseload at the study centre. Studies have suggested that better outcomes are achieved in hospitals with a high volume of debulking surgeries for advanced ovarian cancer ([Mercado 2010](#); [Schrug 2006](#)).
- Whether primary debulking surgery is performed by a subspecialised gynaecological oncologist. Quality of care and associated outcomes (including the probability of undergoing debulking surgery, survival, etc.) have been reported to be dependent on whether a general surgeon, general gynaecologist, or gynaecological oncologist performs the surgery ([Earle 2006](#); [Mercado 2010](#)).
- Percentage of women with stage IIIC/IV ovarian cancer. It could be more difficult to achieve macroscopic debulking in these women compared to women with stage IIIA/IIIB ovarian cancer.

INDEX TERMS

Medical Subject Headings (MeSH)

*Diffusion Magnetic Resonance Imaging; *Fluorodeoxyglucose F18; *Positron Emission Tomography Computed Tomography;

*Radiopharmaceuticals; Fallopian Tube Neoplasms [*diagnostic imaging] [pathology] [surgery]; Feasibility Studies; Neoplasm,

Residual [diagnostic imaging]; Ovarian Neoplasms [*diagnostic imaging] [pathology] [surgery]; Peritoneal Neoplasms [*diagnostic imaging] [pathology] [surgery]; Randomized Controlled Trials as Topic; Sensitivity and Specificity; Tomography, X-Ray Computed

MeSH check words

Female; Humans