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 and/or primary peritoneal cancer (Protocol)

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

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Diagnostic Test Accuracy Protocol]

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

Jacob P Hoogendam[1], Joline F Roze[1], Fleur T van de Wetering[2], René Spijker[2], Leen Verleye[3], Joan Vlayen[3], Wouter B Veldhuis[4], Rob JPM Scholten[2], Ronald P Zweemer[1]

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

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

A B S T R A C T

This is a protocol for a Cochrane Review (Diagnostic test accuracy). The objectives are as follows:

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

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.

B A C K G R O U N D

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). 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 non-specific including abdominal pain or discomfort and fatigue (Olson 2001). The extent of ovarian cancer is categorised using the International Federation of Gynaecology and Obstetrics (FIGO) staging criteria. 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

cavity or towards regional lymph nodes (stage III), or towards extra-abdominal lymph nodes and/or with haematogenous spread resulting in distant metastasis (e.g. liver parenchyma, stage IV) (Mutch 2014; Prat 2014). This late presentation makes ovarian cancer the leading cause of death from gynaecologic malignancies in developed countries worldwide with an absolute 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 containing veins, arteries, lymphatics), and the attempted resection of all visible tumour deposits (NCI 2015). The actual feasibility of the latter is in reality 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, divided in debulking with 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 patients with advanced stage epithelial ovarian cancer (Bristow 2002; Elattar 2011; NCI 2015; Vergote 2010). Unfortunately, despite chemotherapy and macroscopic debulking surgery, the majority of patients 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 radical in the sense of radical 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 seems feasible based on imaging, primary debulking surgery is attempted. If imaging indicates that the chance of macroscopic debulking is small, patients 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 wherein primary surgery is attempted but deemed unfeasible, which causes unnecessary morbidity and negatively influences prognosis. In contrast, macroscopic debulking is the strongest independent predictor of patient outcome and should be attempted whenever deemed possible (Vergote 2010). Therefore, it is unde-

sirable to select woman for neoadjuvant chemotherapy followed by interval debulking surgery when the tumour load would have been primary resectable. In conclusion, it is important to conscientiously select patients for either primary debulking surgery with adjuvant chemotherapy or neoadjuvant chemotherapy followed by interval debulking.

Bristow 2002 demonstrated the extensive heterogeneity between centres on 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 patients with residual tumour after primary debulking surgery still range up to 31% (Rutten 2014). 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).

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 patients in whom no treatment, surgical or chemotherapy, has been given prior to this surgery. Three target condition categories will be considered.

● Macroscopic debulking, which is 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) metastasis 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 mesenterial resections). Consequently, this leads to an incomplete debulking with residual deposits of ovarian cancer.

● Incomplete debulking with residual disease, divided into two subcategories, depending on whether macroscopically visible tumour deposits:
    ○ ≤ 1 cm in diameter remain at the end of surgery; or
    ○ >1 cm in diameter remain at the end of surgery.

Index test(s)

In this systematic review, we will consider the following three non-invasive and commonly available index tests.

- Whole body fluorodeoxyglucose-18 (FDG) positron emission tomography (PET), with or without a parallel conventional 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 (DWI), in addition to conventional MRI, an imaging method that uses the diffusion of water molecules to generate contrast.

Laparoscopy, performed either as ambulatory surgery or directly before the laparotomy, can also be considered as an alternative test. A recent Cochrane systematic review on laparoscopy for the assessment of tumour resectability in ovarian cancer remained inconclusive (Rutten 2014). However, a randomised controlled trial which assessed whether the proportion of incomplete debulking surgeries with residual disease ≤ 1 cm or > 1 cm in diameter can be reduced by laparoscopic assessment has been performed and we are now awaiting results (Rutten 2012).

Clinical pathway

With (subtle) symptoms, or based on accidental discovery of an abdominal mass, patients suspected of ovarian cancer preferably present to a gynaecological oncologist. Here, a standard diagnostic work-up is performed starting with obtaining information about i.e. medical history, symptoms, family history, known allergies, use of medication and the 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.

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 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 CT, in an attempt to filter out false-negatives (i.e. resectable based on 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 patients with residual disease outcome after primary debulking surgery. If PET(-CT) and/or MRI show superior accuracy, more adequate selection of patients 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.

Alternative test(s)

When studies do not follow an add-on design (i.e. substitute CT by one of the index tests), contrast enhanced CT of the abdomen can be considered as an alternative test (to the index tests) in patients with stage III - IV epithelial ovarian, fallopian tube and/ or primary peritoneal cancer.

OBJECTIVES

To assess the diagnostic test accuracy of PET(-CT), conventional and diffusion-weighted MRI as an replacement or an add-on to abdominal CT, for predicting tumour resectability at primary debulking surgery in patients with stage III - IV epithelial ovarian, fallopian tube and/or 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.

M E T H O D S

Criteria for considering studies for this review

Types of studies

We will include randomised comparisons of diagnostic tests, cross-sectional, retro- and prospective cohort studies, that address the DTA of preoperative PET(-CT), conventional or (additional) diffusion-weighted MRI on predicting tumour resectability in women who are planned to undergo primary debulking surgery. Studies which add-on to CT or when the index test replaces CT, will be included. To evaluate the add-on effect, the alternative imaging test should be performed within four weeks before or after the CT scan. Studies following a case-control design, which carry an inherent high risk of bias in a DTA research objective, will be excluded.

Participants

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

Index tests

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

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

Target conditions

The target condition is defined as the resectability of all deposits from epithelial ovarian, fallopian tube and/or primary peritoneal cancer at primary debulking surgery. This target condition has 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 will be primary debulking surgery. This is most commonly performed via a laparotomy, although in recent years laparoscopy has also 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 is determined by the surgeon at the end of this surgery.

Search methods for identification of studies

Our search for relevant literature will involve electronic databases (see Electronic searches) as well as from additional sources (see Searching other resources).

Electronic searches

We will search MEDLINE (Ovid), Embase (Ovid), Science citation index expanded (web of knowledge), social sciences citation index (web of knowledge), and arts & humanities citation index (web of knowledge). See Appendix 1 for a proposed draft search strategy to be run in MEDLINE (OVID). Similarly, structured search strategies will be designed using search terms appropriate for each database. Controlled vocabulary such as MeSH terms and EMTREE will be used where appropriate. Search filters (collections of terms aimed at reducing the number needed to screen) will not be used as an overall limiter because those published have not proved sensitive enough (Beynon 2013). No language restriction will be applied to the electronic searches, and translation services will be used as necessary.

Initial searches will be performed by one review author with extensive experience in systematic reviews. Screening of abstracts and titles will be conducted independently by two review authors.

Searching other resources

We will search both clinicaltrials.gov and WHO-ICTRP to identify prospectively registered trials. The reference lists of all relevant studies will be searched for additional relevant studies. These studies will also be used to search the electronic databases to identify additional studies through the use of the related article feature and to identify newly citing articles by way of cited reference search.

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

4

We will also contact research groups authoring studies used in the analysis for unpublished data.

Data collection and analysis

The data collection and analysis will adhere 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 will be downloaded to a reference management database (*Endnote*) and duplicates will be removed. The remaining references will be independently examined by two review authors (JFR and JPH) on title and abstract using the pre-set inclusion and exclusion criteria as stated above. Afterwards, discrepancies in judgement between both review authors will be discussed in an attempt to reach consensus. When the possible inclusion or exclusion of an individual study remains unclear, full-text assessment will be independently performed by the same two review authors for a final decision. Articles considered directly eligible based on title and abstract screening will also be read in full text to definitively confirm adherence to the inclusion and exclusion criteria. When discrepancies in judgement between both investigators persist following full-text assessment, RJPMS will act as a referee to reach a final verdict. Excluded studies will be documented and the reasons for exclusion 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) will independently perform data extraction. Missing data will be sought, if possible, by contacting the study authors concerned. Data will be checked and entered into RevMan 5 by one review author and checked by another review author.

For included studies, data on characteristics of patients (inclusion criteria, age, stage, co-morbidity, previous treatment, number enrolled in each arm) and interventions (type, formulation, dose, duration, regimen), risk of bias, duration of follow-up, outcomes and deviations from protocol will be abstracted independently by two review authors (JFR and JPH) onto a data abstraction form specially designed for the review.

Assessment of methodological quality

The QUADAS-2 assessment tool for diagnostic accuracy studies in the context of systematic reviews will be completed for all included studies (Whiting 2011). This assessment will be performed independently by two review authors, one of whom has a clinical background (JFR), and the other both a clinical and methodological background (JPH). Final results will be based on consensus discussion. When consensus cannot be reached or uncertainty persists, RJPMS will act as a referee to reach a final decision. Operational definitions of QUADAS items are derived from Rutten 2014 and described in Appendix 2.

Statistical analysis and data synthesis

We will perform separate analyses for two different definitions of resectability versus non-resectability, based on the three common categories of the target condition (see; Target condition being diagnosed):

- Macroscopic debulking versus incomplete debulking with residual disease of any size (i.e. a dichotomisation of tumour deposits of 0 cm in diameter).
- Macroscopic debulking or incomplete debulking with residual disease ≤ 1 cm in diameter versus incomplete debulking with residual disease > 1 cm in diameter (i.e. a dichotomisation of tumour deposits at 1 cm).

Figure 1 and Figure 2 outline the definitions of the two by two table for these analyses.

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 $\leq$ 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 correctly judged as unfeasible on MRI or PET(-CT) whereas surgery resulted in an incomplete resection with residual disease of any size	FP: Macroscopically complete resection was incorrectly judged as unfeasible on MRI or PET(-CT) whereas surgery resulted in a macroscopically complete resection
	Negative Macroscopically complete resection is feasible	FN: Macroscopically complete resection was incorrectly judged as feasible on MRI or PET(-CT) whereas surgery resulted in an incomplete resection with residual disease of any size	TN: Macroscopically complete resection was correctly judged as feasible 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 $\leq$ 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 correctly judged as unfeasible 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 incorrectly judged as unfeasible 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 incorrectly 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 correctly 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

We will perform separate analyses for the two different roles of the index tests: a) as a replacement of CT, and b) as an add-on test in women who are considered resectable by CT.

We will perform meta-analyses of DTA studies according to the guidelines described in the Chapter 10 of the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (Macaskill 2010). Review Manager software (Review Manager 2014) will be used to prepare forest plots of paired sensitivity and specificity of the included studies, the distribution of studies in the receiver operating characteristics (ROC)-space and the graphical presentation of the results of the meta-analyses, whereas the actual meta-analyses will be done by the use of SAS (module METADAS). This module includes random-effects methods for meta-analysis of DTA studies in which overall sensitivity and specificity (or a summary ROC-curve) are jointly estimated, whilst taking account of the existing covariance of those two parameters and the existing heterogeneity between studies, which is the rule rather than the exception in meta-analyses of DTA studies (Macaskill 2010).

We assume, that implicit uniform thresholds will have been applied for each index test. Therefore, we will use pairs of sensitivity and specificity to calculate a summary estimate of sensitivity and specificity using a generalised linear mixed model (binomial family) (Reitsma 2005). If thresholds vary across studies, we will estimate the underlying ROC curve by the use of a hierarchical summary receiver operating characteristics (HROC) model (Rutter 1995; Rutter 2001).

To compare the accuracy of different index tests, we will extend the model with covariates indicating the type of index test. This allows for formal comparison of differences in mean sensitivity and specificity between index tests. In addition, we will restrict the analysis to studies that applied more than one index test in the same patients.

We will perform all statistical analyses with the macro METADAS using the statistical software SAS (Macaskill 2010).

We will assign 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 prepare 'Summary of findings' tables (Schünemann 2008). Labelling the tumour status erroneously as resectable ('false negatives') is considered worse than labelling the tumour status erroneously as non-resectable ('false positives'), For GRADE, therefore, the DTA outcome 'false negative' is considered 'critical' (9) and the DTA outcome 'false positive' as less critical (8). The other outcomes (TP and TN) are considered 'important'.

The methods for GRADEing the level of evidence for DTA studies are still under development. We will apply the methods described in 2008 (Schünemann 2008) and in a more recent draft paper by Schünemann and colleagues (Schünemann personal communication). To create the GRADE-profiles and 'Summary of findIngs' tables, we will use the Guideline Development Tool (http://gdt.guidelinedevelopment.org/central˙prod/ ˙design/client/index.html).

Investigations of heterogeneity

We will explore heterogeneity by adding covariates to the statistical model for the following characteristics. The following covariates will be 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 will 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; Schrag 2006).
- Whether primary debulking surgery is performed by a subspecialised gynaecological oncologist. Quality of care and associated outcomes (including the probability to undergo 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 patients with stage IIIC/IV ovarian cancer. It could be more difficult to achieve macroscopic debulking in these patients compared to stage IIIA/IIIB patients.

Sensitivity analyses

Sensitivity analyses will be performed by excluding studies at high risk of bias for each of the QUADAS domains.

Assessment of reporting bias

No assessment of reporting bias will be 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).

ACKNOWLEDGEMENTS

REFERENCES

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]

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. Available from: http://srdta.cochrane.org/*. The Cochrane Collaboration, 2013.

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

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. International Agency for Research on Cancer; Lyon, France: Available from: http://globocan.iarc.fr, accessed on 27/5/2015 2013.

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 editor(s). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. 1.0. The Cochrane Collaboration, 2010.

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]

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 (Bethesda, MD, USA). PDQ® Ovarian Epithelial, Fallopian Tube, and Primary Peritoneal Cancer Treatment. Available at: http://www.cancer.gov/types/ovarian/hp/ovarian-epithelial-treatment-pdq: Date last modified March 27, 2015. Accessed June 8th 2015.

NICE 2011
National Institute for Health and Clinical Excellence (NICE). Ovarian cancer: The recognition and initial management of ovarian cancer. https://www.nice.org.uk/guidance/CG122/ 2011.

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.

Reitsma 2005

Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology* 2005; **58**(10):982–90. [PUBMED: 16168343]

Review Manager 2014 [Computer program]

The Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager (RevMan) [Computer program]. Version 5.3. The 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 2012

Rutten MJ, Gaarenstroom KN, Van Gorp T, van Meurs HS, Arts HJ, Bossuyt PM, et al. Laparoscopy to predict the result of primary cytoreductive surgery in advanced ovarian cancer patients (LapOvCa-trial): a multicentre randomized controlled study. *BMC Cancer* 2012;**12**:31.

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]

Rutter 1995

Rutter CM, Gatsonis CA. Regression methods for meta-analysis of diagnostic test data. *Academic Radiology* 1995;**2** (S1):S48–S56.

Rutter 2001

Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistical Methodology* 2001;**20**:2865–84.

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; 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;**18**(155):529–36.

* *Indicates the major publication for the study*

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

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

9

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 Present	Epub Ahead of tions, Ovid MEI Present
#	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.

(Continued)

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. Operational definitions of QUADAS-2 items

	Risk of bias			Applicability		
	Quality indicator		Notes	Quality indicator	Notes
Domain 1 Patient Selection	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)	
	1. Was a consecutive or random sample of patients enrolled?	"Yes" if a consecutive or random sample of patients was enrolled. "No" if a selected group of patients was enrolled "Unclear" if there is insufficient information on enrolment		1. Were the patients diagnosed by conventional diagnostic work-up for advanced stage ovarian cancer?	"Yes" if patients were diagnosed by conventional diagnostic work-up with advanced stage ovarian cancer. "No" if patients included in the trial are diagnosed with low-stage disease (FIGO I or II) only. No high stage disease patients in the trial "Unclear" if there is insufficient information on recruitment method, criteria for diagnosis of ovarian cancer
	2. Did the study avoid inappropriate exclusions?	"Yes" if there were no inappropriate exclusions. "No" if there were inappropriate exclusions. "Unclear" if there is insufficient information on exclusions		2. Were the patients planned for primary debulking surgery after conventional diagnostic work-up?	"Yes" if the patients were planned for primary debulking surgery after conventional diagnostic work-up. "No" if none of the patients were planned for primary debulking surgery "Unclear" if there is insufficient information.

(Continued)

Domain 2 Index Test	Could the interpretation of the Index test have introduced bias? (High/low/unclear)		Are there concerns that the index test, its conduct or the interpretation differ from the review question? (High/low/unclear)	
	1. Were the index test results interpreted without the knowledge of the results of the reference standard?	This will always be rated as yes, because the index test is performed before the reference standard		
	2.Was the threshold used pre-specified?	"Yes" if a clear description of the threshold is given which was specified before start of the study "No" if no clear description is given beforehand. "Unclear" if there is 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 is given about when the index test is positive or negative (e.g. what the cut-off for too extensive abdominal disease was) "No" if there is no clear description is given about what is classified as too extensive disease or not "Unclear" if there is insufficient information within the paper to determine whether or not a defined threshold was used to 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 is provided.		
	4. Did patients receive the same reference standard regardless of the index test result?	"Yes" if patients who underwent reference standard had laparotomy "No" if patients did not undergo laparotomy. "Unclear" if insufficient information is provided.		
Domain 3 Reference Standard	Could the interpretation of the reference standard have introduced bias? (High/low/unclear)		Are there concerns that the target condition as defined by the reference standard does not match the question? (High/low/unclear)	

(Continued)

	1. Is the reference standard likely to correctly classify the target condition?	"Yes" if the reference standard is laparotomy. "No" if the reference standard used is not the one defined in the protocol "Unclear" if the information is insufficient.	1. Did the study provide a clear definition of what was considered to be a 'positive' result for the reference standard?	"Yes" if a clear description is given about when the reference standard is positive or negative (e.g. if description is given about the size of the tumour deposits left after surgery) "No" if there is no clear description of tumour deposit size after surgery "Unclear" if there is insufficient information within the paper that describes 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 is performed by individuals who did not perform the index test. "No" if the reference test were done by the same person performing the index test "Unclear" if not reported.		
	3. Is the surgeon's expertise adequate to perform the reference standard?	"Yes" if the reference test is performed by a gynaecological oncologist "No" if the reference test is not performed by a gynaecological oncologist "Unclear" if not reported.		
Domain 4 Flow and Timing	Could the patient flow have introduced bias? (High/low/unclear)			
	1. Is 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 does not extend 6 weeks "No" if the time period is more than 6 weeks for an unacceptable high proportion of patients		

		"Unclear" if the information on the timing of tests is not provided		
	2. Did all patients receive the same reference standard?	"Yes" if all patients underwent the reference standard (laparotomy) "No" if not all patients underwent reference standard. "Unclear" if insufficient information is provided.		
	3. Were all patients included in the analysis?	"Yes" if for all patients entered in the study are included in the analysis "No" if not all the patients in the study are included in the analysis "Unclear" if it is not clear whether all patients were accounted for		
	4.Were withdrawals from the study reported?	"Yes" if for all patients entered in the study is reported what happened during the study, also those who withdraw or answer "yes" if no withdrawals where reported and all patients who entered in the study results were reported "No" if not all the patients in the study complete the study and these patients were not accounted for "Unclear" if it is not clear whether all patients were accounted for		

CONTRIBUTIONS OF AUTHORS

The study protocol was written by review authors JFR, JPH, RS, RJPMS and RPZ and subsequently revised by all authors.

DECLARATIONS OF INTEREST

All review authors declare they have 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.